Mathematical Modeling and Numerical Methods for Non-classical Transport in Correlated Media

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Dipl.-Math. Kai Krycki

aus Herten

Berichter: Univ.-Prof. Dr. rer. nat. Martin Frank

Univ.-Prof. Dr. rer. nat. Rodolphe Turpault

Tag der mündlichen Prüfung: 20.07.2015

To Tanja and Max.

Abstract

In this work, we investigate a new and non-classical linear transport equation for the transport of particles in correlated background media. We derive a time-dependent non-classical transport equation that is capable of reproducing arbitrary path length distributions, in contrast to the classical theory. This equation governs the distribution function of a microscopic particle game in the Boltzmann-Grad limit. This rigorous mathematical derivation is based on recent results on the periodic Lorentz gas, and it relies on an analytic expression for the distribution of free path lengths in the limit. The resulting equation has the distance s to the next collision as an additional independent variable, which makes it "non-classical". A Monte-Carlo study of path length distributions in homogeneous but correlated media suggests that this equation is valid for a wider class of spatially correlated media. We discuss generalizations of the resulting equation, as well as the connection to another recently proposed non-classical steady state transport equation.

In the following, we develop numerical methods for this non-classical steady state transport equation. Therefore, we investigate a variation of a standard finite volume HLL method for moment models of this equation. In a detailed numerical analysis we show that these schemes preserve the analytic asymptotic limit in a diffusive scaling. Furthermore, the schemes preserve the convex set of admissible and realizable sets. A coupling of the initial value and the full solution strongly suggests to use an iterative solution method. The naive source iteration method is shown to become arbitrarily slow in the scattering dominated regime, therefore we adapt a Diffusion Synthetic Acceleration method for the classical equation based on the diffusion approximation for the non-classical equation. We investigate the contraction rates via a von Neumann analysis of the full equation and the corresponding moment models, which shows that they are significantly decreased by the acceleration method.

Equipped with accurate and efficient numerical schemes, we present a method for the solution of inverse problems based on the non-classical steady state transport equation. We formulate the parameter estimation problem as an optimal control problem, and derive a first order optimality system using a Lagrange formalism. This formalism is based on adjoint calculus, therefore we formally derive the adjoint equation of the non-classical steady state transport equation. Then we adapt the numerical schemes to discretize this first order optimality system. We also show that these schemes are consistent with the underlying discrete optimality system. Numerical solutions of the optimality system are obtained using gradient based optimization methods and show that parameters can be reconstructed accurately at a reasonable computational cost.

Contents

1 Introduction

The transport of radiation in a background medium is an important physical phenomenon that appears in numerous applications. Examples of applications range from the transport of neutrons in a nuclear reactor and electrons in human tissue in radiotherapy, to the transport of photons in atmospheric clouds. In the corresponding mathematical models, radiation is treated as a gas of point particles which interact with the background medium. This medium is assumed to consist of static elements, and the point particles are scattered or absorbed when experiencing a collision with an element of the background medium. A classical situation would be elementary particles that interact with atoms. The steady state form of the linear Boltzmann equation (named after Ludwig Boltzmann, 1844-1906) describes the equilibrium distribution Ψ of particles in a homogeneous medium. The equation reads

$$\Omega \nabla_x \Psi(x,\Omega) + \Sigma_t \Psi(x,\Omega) = \int_{S^2} \sigma_s(\Omega,\Omega')\Psi(x,\Omega')\,\mathrm{d}\Omega' + Q(x,\Omega), \tag{1.1}$$

where x is the spatial variable, and Ω the direction of travel. This integro-differential equation is widely used to model the transport of radiation in a great number of applications. The interaction of radiation with the background medium is described by the total cross section Σ_t, and the scattering kernel σ_s. They specify the rate by which particles collide with the background medium, and the change of the direction of travel of the particles after a collision respectively.
To ensure the validity of the linear Boltzmann equation for a certain transport process, some simplifying assumptions have to be made. One of these assumptions, the Poissonian spatial distribution of the elements of the background medium, is closely related to the underlying path length distribution in the medium. Recently, path length distributions were observed in atmospheric clouds that cannot be explained by the classical theory. Therefore, in this work, we investigate a new and non-classical mathematical model for the transport of massless point particles through a random disordered background medium that is capable of reproducing arbitrary path length distributions. This model has first been suggested for the transport of photons in atmospheric clouds by Edward Larsen in 2007 [73]. It is therefore the motivational application for this work, and we refer to it at several places. The non-classical transport equation is set on an extended phase space and reads

$$\partial_s \Psi(s,x,\Omega) + \Omega \nabla_x \Psi(s,x,\Omega) + \Sigma_t(s)\Psi(s,x,\Omega) = 0, \tag{1.2a}$$

$$\Psi(0,x,\Omega) = c\int_0^\infty \int_{S^2} \sigma_s(\Omega,\Omega')\Sigma_t(s')\Psi(s',x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s' + Q(x,\Omega). \tag{1.2b}$$

The variable s is the distance to the last collision, and makes this equation non-classical. If the cross section $\Sigma_t(s)$ is constant, this equation reduces to the classical linear Boltzmann Equation (1.1). It is therefore a generalization of the classical model.

This work falls naturally into three parts. In the first part on **Mathematical Modeling** we study the rigorous mathematical derivation of non-classical transport equations from a random microscopic dynamics, and the connection to Equation (1.2). First, we review the classical linear Boltzmann equation in Chapter 2. We point out limitations of the classical theory, and present applications where observations underline the need of a new generalized theory in Chapter 3. In the following Chapter 4 we present a rigorous derivation of a time-dependent non-classical transport equation. There, we build upon recent results on the so-called periodic Lorentz gas [53, 81]. In Chapter 5, we investigate extensions of this equation via a Monte-Carlo study of spatially correlated media. Finally, in Chapter 6 we discuss the connection of this non-classical transport equation and Equation (1.2), and present certain extensions and simplifications.

In the second part of this work, entitled **Numerical Methods**, we consider the problem of numerically solving the non-classical steady state transport Equation (1.2). Solving the full equation is numerically costly due to its high dimensionality. Therefore we use different strategies to reduce the complexity of the full kinetic equation. These so-called moment models are presented in Chapter 7. Since a background medium like an atmospheric cloud may have optically thick regions, where transport is dominated by scattering, we are especially interested in the performance of numerical schemes in this regime. For approximate systems of angular moments we present a family of asymptotic preserving HLL schemes in Chapter 8. Coupling of the initial value and the full solution in Equation (1.2) requires an iterative solution method. We show that source iteration becomes arbitrarily slow in the diffusive regime, and propose initialization and acceleration schemes to overcome this problem in Chapter 9. Numerical results for these schemes are presented in Chapter 10.

In the third part, we investigate the problem of **Parameter Estimation**. The $s-$dependent total cross section $\Sigma_t(s)$ in the non-classical transport equation can be expressed in terms of the path length distribution $p(s)$ and vice versa. That means, Equation (1.2) can be used to identify path length distributions in practical applications. First, we formulate parameter estimation as an optimal control problem in Chapter 11. In the following Chapter 12 we present a formal analysis of the mathematical operators that are involved in this problem. This is used in Chapter 13, where we derive the first order optimality system of the optimal control problem by following a Lagrange formalism. Numerical methods for this optimality system are the topic of Chapter 14. We apply the numerical schemes introduced in the previous part for the non-classical equation and its adjoint in order to solve the optimality system using gradient

based methods, and present a number of numerical tests. Finally, in Chapter 15 we investigate the discrete form of the first order optimality system for the numerical schemes presented in the preceding chapter.

We close with a discussion and conclusion of the results of this work, and give an outlook on future research in Chapter 16.

Statement regarding the good scientific practice

This piece of work was written by Kai Krycki (in the following: the author of this work) during his employment at the Chair for Mathematics (CCES) at RWTH Aachen University, between June 2010 and March 2015 under the supervision of Prof. Martin Frank. Whenever we make use of existing works by different authors, this is explicitly indicated and we carefully cite the corresponding sources. All computer codes that were used to produce numerical results in this work were written by the author of this work. These codes are available upon request at the Chair for Mathematics (CCES) at RWTH Aachen University. Some parts of this work are based on publications of the author together with different coauthors, or involve essential contributions by others. In the following we specify exactly the way and to what extent others have contributed to this work.

The rigorous derivation of a non-classical transport equation in Chapter 4 is based on a sketch given by François Golse, originated in a discussion with the author. This sketch was then worked out by the author of this work.
The use of a Metropolis Monte-Carlo method in Chapter 5 to create spatial correlated media, based on radial distribution functions, was suggested by Ahmed Ismail. The development of the corresponding codes was assisted by Richard Vasques and Akash Pakanati.
The full Chapter 8, Section 9.1, and parts of Section 9.4 are based on the paper

K. Krycki, C. Berthon, M. Frank, R. Turpault: *Asymptotic preserving numerical schemes for a non-classical radiation transport model for atmospheric clouds*, Math. Meth. Appl. Sci., Vol. 36 (16), pp. 2101-2116 (2013).

The idea of the modification of the numerical scheme to ensure the asymptotic preserving property is due to Rodolphe Turpault and Christophe Berthon. The coauthors also carried out the analysis of the stability, realizability and admissibility of the numerical schemes. The contribution of the author of this work were the analysis of the continuous moment models, certain corrections and remarks in the analysis done by the coauthors, and the full investigation of the convergence of the iterative solution method.

One example in Chapter 10 is taken from the paper

R. Vasques, K.Krycki: *Solving the 1D non-classical transport equation in random periodic media*, Proceedings of the Joint International Conference on Mathematics and Computation, Nashville, TN (2015).

The derivation of the analytic path length distribution in the homogenized medium was mainly carried out by Richard Vasques.

Acknowledgements

First of all, I would like to thank my advisor Martin Frank for bringing me in contact with the very interesting topic of non-classical transport, and for the good and patient supervision of this work.
I thank Rodolphe Turpault and Christophe Berthon for the cooperation in the past years, and for hosting me at the University of Nantes during my stay in 2013.

Many thanks go to the following people for providing new and important perspectives and stimuli for this work: Ahmed Ismael, Edward Larsen, and Kevin Vynck. I am especially thankful to Francois Golse and Jens Marklof for pointing out the relation of non-classical transport and the periodic Lorentz gas, and for introducing me to the underlying theory.

I would like to thank all of the MathCCES staff for a great time and a pleasent working atmosphere. I want to express special gratitude to Richard Vasques for helping out with his profound physical intuition regarding non-classical transport, Richard Barnard for introducing me to the subtleties of adjoint-based numerical optimization, and Edgar Olbrant for the many fruitful discussions on transport theory and his practical advises regarding the use of Macs and MATLAB. I also thank Akash Pakanati for his work as a student assistant, where he worked on the development of the codes for the path length simulations in correlated background media.

Lastly, I thank Tanja Klietzing for carefully proofreading this work, and for so much more.

Part I

Mathematical Modeling

Introduction: Mathematical Modeling

The classical linear transport equation describes the transport of particles through a random and disordered background medium. It is a special form of the linear Boltzmann equation, first proposed by Ludwig Boltzmann in the late 19th century. Its rigorous derivation from first principles only holds if the positions of the scattering elements in the background medium are uncorrelated [49, 100]. A major consequence of this assumption on the spatial structure of the background medium is that the distribution of free path lengths in the random background medium has to be an exponentially decaying function. But current observations, for example for the transport of photons in atmospheric clouds, point towards a non-exponentially decaying path length distribution [66, 89, 67, 48]. Therefore, in this part, we study the rigorous derivation of non-classical transport equations for simplified transport processes through random media with an underlying general, possibly non-exponential path length distribution. Luckily, we can build upon recent results on the so-called periodic Lorentz gas [53, 81]. There, the scatterers are fixed on a grid and in the Boltzmann-Grad limit a kinetic equation set on an extended phase space has been derived. This was done in a two dimensional setting for the case of specular reflected particles. We extend this result to isotropic, random scattering. This is a more realistic model when it comes to the transport of radiation, for example photons in atmospheric clouds. Note that a periodic setting of scatterers is in a way the strongest possible spatial correlation.

In the first Chapter 2, we shortly review the classical linear transport equation, introduce the notation, and sketch the mathematically rigorous derivation. We point out limitations of the classical theory in Chapter 3. Therefore, we present applications where it comes to spatial correlations in the background medium, and accordingly non-exponential path length distributions. In the following Chapter 4 we present a rigorous derivation of a non-classical transport equation for the periodic setting of scatterers. We first give an overview on the recent results on the periodic Lorentz gas, and then present a similar derivation of a non-classical transport equation for the process of radiative transfer with isotropic scattering. The derivation suggests that it also holds for more general correlated background media. This is the subject of Chapter 5, where we investigate path length distributions in spatially correlated media in the Boltzmann-Grad limit via a Monte-Carlo study. Finally, in Chapter 6 we discuss the resulting non-classical transport equations and present certain extensions as well as simplifications and approximations thereof.

2 Classical radiative transfer

In this chapter we review the classical linear Boltzmann equation for radiative transfer, its rigorous mathematical derivation and especially point out its limitations when it comes to spatial correlations in the background medium. The introduction of the classical radiative transfer equation not only serves to point out the differences to the non-classical theory, which is the main subject of this thesis. It is also used to introduce important concepts and notations that are used in the remainder of this work.
The classical radiative transfer equation is a type of a linear Boltzmann equation. Also known as the Lorentz equation, it was first introduced by Drude and Lorentz at the beginning of the 20th century to describe the motion of electrons in metals. With the advent of nuclear technologies in the middle of the 20th century it was also used to describe the transport of neutrons in a nuclear reactor [33]. As a kinetic model for the transport of photons it was first used in the context of stellar atmospheres [27, 28]. When it comes to neutron transport it is usually simply called the "transport equation", while in the context of photons it is widely known as the "radiative transfer equation".

In all what follows $t \geq 0$ denotes the time variable, $x \in \mathbb{R}^N$ is the spatial variable and $\Omega \in S^{N-1}$ is the direction of travel, hence $|\Omega| = 1$. All figures and also most of the theory are restricted to the case of 2 space dimensions, but the arguments mostly hold for an arbitrary space dimension. We indicate whenever this is not the case. In all cases we assume an infinite spatial domain.

2.1 The classical linear Boltzmann equation

The classical radiative transfer equation for the transport of photons through an (in general) inhomogeneous background medium is a linear Boltzmann equation for the angular photon density Ψ. Its time-dependent version is given by the initial value problem

$$\partial_t \Psi(t,x,\Omega) + \Omega \nabla_x \Psi(t,x,\Omega) + \Sigma_t(x)\Psi(t,x,\Omega) = S\Psi(t,x,\Omega) + Q(t,x,\Omega), \tag{2.1a}$$

$$\Psi(0,x,\Omega) = \Psi_0(x,\Omega), \tag{2.1b}$$

where Ψ_0 is an initial photon density, Q is an external source term and S is the scattering operator defined by

$$S\Psi(t,x,\Omega) = \int_{S^2} \sigma_s(x,\Omega,\Omega')\Psi(t,x,\Omega')\,\mathrm{d}\Omega'.$$

The unknown Ψ is usually interpreted as

$$\Psi(t,x,\Omega) = \text{"particles at time } t \text{ at position } x \text{ that travel in direction } \Omega \text{ "}.$$

If the transport of particles other than photons is considered, which may travel by different speeds $|v| \in \mathbb{R}_+$, the variable Ω is usually replaced by the velocity variable v, where $v = |v|\Omega$. The total cross section Σ_t denotes the rate by which particles interact with the background medium. The scattering kernel σ_s represents the probability, that a particle that has travelled in direction Ω' before a collision changes direction of flight to Ω afterwards. Therefore, the so-called "detailed balance"

$$\int_{S^{N-1}} \sigma_s(x,\Omega,\Omega')\,\mathrm{d}\Omega = \int_{S^{N-1}} \sigma_s(x,\Omega,\Omega')\,\mathrm{d}\Omega' = 1$$

holds for all $\Omega, \Omega' \in S^{N-1}$ and all $x \in \mathbb{R}^N$. If the scattering kernel only depends on the angle between the incoming and outgoing direction, this is typically written as

$$\sigma_s(\Omega,\Omega') = \sigma_s(\Omega \cdot \Omega'),$$

and then only depends on a scalar. The total cross section and the scattering kernel are linked in the following way. Let the scattering cross section be defined as

$$\Sigma_s(x) = \iint_{S^{N-1}\times S^{N-1}} \sigma_s(x,\Omega,\Omega')\,\mathrm{d}\Omega\,\mathrm{d}\Omega'.$$

Then we have

$$\Sigma_t(x) = \Sigma_s(x) + \Sigma_a(x), \tag{2.2}$$

where Σ_a is the so-called absorption cross section. It denotes the rate by which particles are absorbed when colliding with an element of the background medium. The sum (2.2) implies, that a particle undergoing a collision is either scattered or absorbed. Therefore, the scattering ratio

$$c := \frac{\Sigma_s}{\Sigma_s + \Sigma_a} = \frac{\Sigma_s}{\Sigma_t}$$

is the probability that a particle is scattered, while

$$1 - c = \frac{\Sigma_a}{\Sigma_s + \Sigma_a} = \frac{\Sigma_a}{\Sigma_t}$$

is the probability that a particle is absorbed when undergoing a collision.
For the transport of photons it is often desirable to also take into account frequency dependence. The equation can be extended to this by introducing a frequency variable $\nu \in \mathbb{R}_+$ and by writing $\Psi = \Psi(t,x,\Omega,\nu)$. The parameters then become $\Sigma_t = \Sigma_t(x,\nu)$ and $\sigma_s = \sigma_s(x,\Omega,\Omega',\nu,\nu')$. The latter denotes the

probability that a particle that collides when traveling into direction Ω' having frequency ν', changes its direction of travel to Ω and and its frequency to ν afterwards. The scattering source term then becomes

$$(S\Psi)(t,x,\Omega,\nu) = \int_{S^2}\int_0^\infty \sigma_s(x,\Omega,\Omega',\nu,\nu')\Psi(t,x,\Omega',\nu')\,\mathrm{d}\nu'\,\mathrm{d}\Omega'.$$

In certain situations, Equation (2.1) can be written in a simpler form. If the background medium is homogeneous, the parameters Σ_t, Σ_s, and σ_s are independent of x. And, in the case of isotropic scattering, the scattering kernel is constant and reduces to $\sigma_s \equiv 1/4\pi$.
When the source term Q does not depend on t, and the medium is absorbing, that means $\Sigma_a > 0$, the solution of equation (2.1) converges to a steady state $\Psi = \Psi(x,\Omega)$ as $t \to \infty$, which is the solution of

$$\Omega\nabla_x\Psi(x,\Omega) + \Sigma_t(x)\Psi(x,\Omega) = S\Psi(x,\Omega) + Q(x,\Omega), \tag{2.3}$$

with $\Psi \in L^1(\mathbb{R}^N \times S^{N-1})$, and appropriate boundary conditions. We refer to this equation as the (classical) steady state transport equation.

2.2 Derivation of the classical equation

Although the linear Boltzmann equation was already used in various fields, a mathematically rigorous derivation of that equation from first principles was not present before Gallavotti [49] in the 1970s. There, the trajectories of particles through a background medium are calculated explicitly, then averaged over all possible configurations of the stochastic medium. Finally, this is homogenized in the so-called Boltzmann-Grad limit. We shortly sketch the derivation given in [49], based also on the review [53], since the ideas and concepts play an important role in the derivation of a non-classical linear Boltzmann equation.

We distinguish two types of particles. Massless point particles (an idealized model for photons, electrons, etc.) move through a background medium made up of infinitely heavy spherical particles (here called scatterers). In the original paper, Gallavotti uses the terminology W-particles and T-particles, since this type of model is often referred to as wind-tree-model, for obvious reasons. We use a notation that is better suited to the particular process of photon transport that we are interested in. The scatterers are at rest and, hence, do not interact with each other. We also neglect interactions between the point particles. By

$$\{c\} := \{c_1, c_2, \ldots, c_n, \ldots\}$$

we denote a configuration of infinitely many scatterers with radius r, represented by their center coordinates. We allow for overlaps to avoid the introduction of correlations between the center coordinates at this point. The scatterers are assumed to be randomly distributed in space under a Poisson

distribution. That means that

$$N(\tilde{V}) = \text{number of scatterers in the volume } \tilde{V}$$

is Poisson distributed, meaning that the probability of finding k scatterers in the (arbitrary, but measurable) set $\tilde{V}$ is given by

$$\mathbb{P}(N(\tilde{V}) = k) = \frac{(\overline{n}\,\mathrm{vol}(\tilde{V}))^k}{k!}\exp(-\overline{n}\,\mathrm{vol}(\tilde{V})). \tag{2.4}$$

The parameter $\overline{n} > 0$ has the physical interpretation

$$\overline{n} = \text{expected number of scatterers per unit surface in } \mathbb{R}^2.$$

This distribution of scatterers implies that there are no correlations between the spatial positions of the centers.
We assume that the point particles travel between the scatterers by unit speed into a direction denoted by $\Omega \in S^1$. When colliding with a scatterer at position $x = c_i + nr$, where n denotes the outward pointing unit vector on the sphere, the point particle undergoes a specular reflection and the new direction of travel after collision becomes

$$R(\Omega; n) := \Omega' = \Omega - 2(\Omega \cdot n)\Omega. \tag{2.5}$$

We also write the scattering map in terms of the deflection angle β as $R(\Omega;\beta)$. Hence, the trajectory of a single point particle is uniquely determined by its initial position and direction of travel and the configuration $\{c\}$. Choosing an initial position $\bar{x}$ and an initial direction $\bar{\Omega}$, the evolution in time is denoted by

$$(X_t, \Omega_t)^{\{c\}}(\bar{x}, \bar{\Omega}),$$

where X_t denotes the position and Ω_t the direction of flight at time t. Between scattering events a particle moves on linear trajectories in a constant direction. In case of a collision, this direction changes according to (2.5). Hence, (X_t, Ω_t) is a solution of the ordinary differential equation

$$\partial_t X_t = \Omega_t, \quad \partial_t \Omega_t = 0, \quad X_0 = \bar{x}, \quad \Omega_0 = \bar{\Omega},$$

with boundary conditions

$$X_{t^+} = X_{t^-}, \quad \Omega_{t^+} = R(\Omega_{t^-}; n),$$

if $X_t = c_i + nr \in \partial B(c_i, r)$. This evolution is clearly reversible by

$$(X_{-t}, \Omega_{-t})^{\{c\}}(x, \Omega) = (X_t, \Omega_t)^{\{c\}}(x, -\Omega).$$

Hence, choosing an initial one-particle distribution function for the point particles $f_0(x, \Omega)$ at $t = 0$, we can write the distribution of the particles a time

$t > 0$ as

$$f_r(t, x, \Omega; \{c\}) := f_0((X_{-t}, \Omega_{-t})^{\{c\}}(x, \Omega))$$

since no particles are lost. Denote by $\langle\cdot\rangle_{\{c\}}$ the average over all configurations $\{c\}$ such that the scatterers are Poisson distributed and no point particle is located inside a scatterer. Then the Boltzmann limit conjecture is that

$$\langle f_r(t, x, \Omega; \{c\})\rangle_{\{c\}} \longrightarrow f(t, x, \Omega)$$

converges to the solution of the linear Boltzmann equation

$$\partial_t f + \Omega\nabla_x f + \sigma f = \sigma \int_0^{2\pi} f(t, x, R(\Omega; \beta)) \sin\frac{\beta}{2}\frac{\mathrm{d}\beta}{4}, \tag{2.6a}$$

$$f(0, x, \Omega) = f_0(x, \Omega), \tag{2.6b}$$

as the radius of the scatterers shrinks to zero ($r \to 0^+$), while the scatterer density grows to infinity ($\overline{n} \to \infty$), keeping the ratio $2r\overline{n} \to \sigma \notin \{0, \infty\}$ constant. This is the so-called Boltzmann-Grad limit.
Therefore, we define the sequence of collision times $\{\tau_j,\ j = 1, 2, \ldots\}$, where

$$\tau_j = \tau_j(x, \Omega; \{c\})$$

is the time of the the j-th collision with a scatterer of a particle with initial position x and initial direction of travel Ω, which of course also depends on the specific configuration of scatterers $\{c\}$. Furthermore, we set $\tau_0 := 0$ and $\Delta\tau_j := \tau_j - \tau_{j-1}$. Then, the evolution of an initial single-particle density f^{in} can be obtained by simply following the trajectory, and it is given by

$$f_r(t, x, \Omega; \{c\}) = f^{\text{in}}(x - t\Omega, \Omega)\mathbb{1}_{t<\tau_1} \tag{2.7a}$$

$$+ \sum_{j\geq 1} f^{\text{in}}\left(x - \sum_{k=1}^{j} \Delta\tau_k \Omega_{-\tau_k^-} - (t - \tau_j)\Omega_{-\tau_j^+}, \Omega_{-\tau_j^+}\right)\mathbb{1}_{\tau_j<t<\tau_{j+1}}. \tag{2.7b}$$

Now we average over all possible scatterer configurations that are Poisson distributed. The average of the first term in the sum above is

$$f^{\text{in}}(x - t\Omega, \Omega)\langle\mathbb{1}_{t<\tau_1}\rangle_{\{c\}} = f^{\text{in}}(x - t\Omega, \Omega)\,\mathrm{e}^{-\overline{n}2rt},$$

since $t < \tau_1$ means that there is no collision ($k = 0$ in (2.4)) in the tube of width $2r$ and length t ($\mathrm{vol}(\tilde{V}) = 2rt$ in (2.4)). Then, in the Boltzmann-Grad limit, we have that

$$\langle f^{\text{in}}(x - t\Omega, \Omega)\mathbb{1}_{t<\tau_1}\rangle_{\{c\}} \longrightarrow f^{\text{in}}(x - t\Omega, \Omega)\,\mathrm{e}^{-\sigma t}$$

as $r \to 0^+, \overline{n} \to \infty$ and $2r\overline{n} \to \sigma$. This is the solution of the Lorentz kinetic

equation in the purely absorbing case

$$\partial_t f + \Omega \nabla_x f + \sigma f = 0, \qquad f_{|t=0} = f^{\text{in}}.$$

Hence, we expect that the averaged second term in the summation (2.7) corresponds to the remaining term by Duhamel's principle. The average of the j-th term in the sum (2.7b) over all scatterer configurations is

$$\int \cdots \int f^{\text{in}} \left(x - \sum_{k=1}^{j} \Delta\tau_k \Omega_{-\tau_k^-} - (t-\tau_j)\Omega_{-\tau_j^+}, \Omega_{-\tau_j^+} \right) \mathrm{e}^{-\overline{n}\,\mathrm{vol}(\mathrm{T}(\mathrm{t};\mathrm{c}_1,\ldots,\mathrm{c}_\mathrm{j}))} \frac{\overline{n}^j \,\mathrm{d}c_1 \ldots \mathrm{d}c_j}{j!},$$

according to (2.4) with $k = j$, and $T(t; c_1, \ldots, c_j)$ being the tube of width $2r$ around the particle trajectory. It holds that $\mathrm{vol}(T(t; c_1, \ldots, c_j)) = 2rt + O(r^2)$. We can change the variables by expressing the spatial positions of the first j scatterers in terms of the collision times $\tau_1, \ldots, \tau_j$ and the deflection angles at the j-th collision β_j. The volume element in the above integral then changes to

$$\frac{\overline{n}^j \,\mathrm{d}c_1 \ldots \mathrm{d}c_j}{j!} = r^j \sin\frac{\beta_1}{2} \cdots \sin\frac{\beta_j}{2} \frac{\mathrm{d}\beta_1 \cdots \mathrm{d}\beta_j}{2} \,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_j.$$

The mapping $(c_1, \ldots, c_j) \mapsto (\tau_1, \ldots, \tau_j, \beta_1, \ldots, \beta_j)$ is surjective if the particle does not collide twice with the same scatterer. Therefore, we split the distribution of particles into two parts

$$f_r(t, x, \Omega; \{c\}) = f_r^M(t, x, \Omega; \{c\}) + f_r^R(t, x, \Omega; \{c\}),$$

where the recollision part f^R is defined on the set of (x, Ω) such that a particle with initial position x and direction Ω collides twice with the same scatterer in the configuration $\{c\}$. Each term in the summation for $\langle f_r^M \rangle_{\{c\}}$ is bounded, and it now holds that

$$\begin{aligned}
\langle f_r^M(t, x, \Omega; \{c\}) \rangle_{\{c\}} \longrightarrow\ & \mathrm{e}^{-\sigma t} f^{\text{in}}(x - t\Omega, \Omega) \\
& + \sigma\, \mathrm{e}^{-\sigma t} \int_0^t \int_0^{2\pi} f^{\text{in}}(x - \tau_1\Omega - (t-\tau_1)R(\Omega;\beta_1), R(\Omega;\beta_1)) \sin\frac{\beta_1}{2} \frac{\mathrm{d}\beta_1}{4} \,\mathrm{d}\tau_1 \\
& + \sum_{j\geq 2} \sigma^j \,\mathrm{e}^{-\sigma t} \int_{0<\tau_1<\cdots<\tau_j<t} \int_{[0,2\pi]^j} \sin\frac{\beta_1}{2} \cdots \sin\frac{\beta_j}{2} \\
& \quad f^{\text{in}} \left(x - \sum_{k=1}^{j} \Delta\tau_k R(\Omega; \sum_{l=1}^{k-1} \beta_l) - (t-\tau_j) R(\Omega; \sum_{l=1}^{j} \beta_l), R(\Omega; \sum_{l=1}^{j} \beta_l) \right) \\
& \quad \frac{\mathrm{d}\beta_1 \cdots \mathrm{d}\beta_j}{4} \,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_j
\end{aligned}$$

in the Boltzmann-Grad limit, which is the Duhamel series that defines the

solution of the Lorentz kinetic equation (2.6). Furthermore, we have that

$$\iint f(t,x,\Omega)\,\mathrm{d}x\,\mathrm{d}\Omega = \iint f^{\mathrm{in}}(x,\Omega)\,\mathrm{d}x\,\mathrm{d}\Omega = \iint f_r(t,x,\Omega;\{c\})\,\mathrm{d}x\,\mathrm{d}\Omega$$

for all $t > 0$ and $r > 0$. That means that the recollision part f_r^R is only defined on a null set in the limit. Finally, by Fatous lemma it holds that

$$\langle f_r^R \rangle \to 0, \quad \langle f_r^M \rangle \to f \text{ in } L^1(\mathbb{R}^2 \times S^1) \text{ uniformly on bounded sets in } t.$$

That finishes the proof of the Boltzmann limit conjecture.

Remark 2.1. We conclude this derivation by pointing out some details of this derivation that are of special importance for the rest of this work.

(1) The change of coordinates implies, that the distribution of particles is determined, aside from the initial distribution f^{in}, by the sequence of free path lengths (given by the collision times) and the new direction of travel after scattering (given by the scattering map).

(2) The particle trajectories were constructed explicitly, and then, by averaging, an explicit solution formula was derived, from where the validity of the corresponding kinetic equation could be deduced.

(3) The scattering kernel for the Lorentz kinetic equation (2.6) takes the form

$$\sigma_s(\Omega,\Omega') = \delta(\Omega' = R(\Omega,\beta))\frac{\sin\frac{\beta}{2}}{4}.$$

For transport processes with different scattering maps, this scattering kernel can simply be replaced by a corresponding different $\tilde{\sigma}_s$ and the equation remains valid.

(4) Especially we see from this derivation that the linear Boltzmann equation and its parameters (here: the total cross section $\sigma > 0$) have to be understood in a homogenized sense.

We make use of these facts and ideas for the derivation of a non-classical transport equation in a following chapter.

3 Limitations of the classical theory

Since its first appearence in the early 20th century, the classical linear Boltzmann equation is used to model a large variety of different transport processes such as neutron transport, electron transport or the transport of photons. Its validity in a great number of applications is unquestionable. Nevertheless, the rigorous derivation in the previous chapter shows that several simplified assumptions have to be made in order to derive this equation. In particular the assumption that the positions of scatterers in the background medium are uncorrelated seems to be invalid in a number of applications. This assumption is closely related to the underlying path length distribution in the medium. Recently, path length distributions were observed in different areas of application that cannot be explained by the classical theory.

Therefore, in this chapter we first show in Section 3.1 that the classical transport equation can only reproduce exponential path length distributions in homogeneous media. Then we present two applications where current observations and theoretical considerations point towards non-exponential distributions. In Section 3.2 we summarize recent results for the transport of photons in atmospheric clouds. Section 3.3 deals with transport in the so-called Lévy glass.
This chapter serves to underline the need of a generalized and non-classical theory that is capable of reproducing these observations.

3.1 Path length distributions

One important quantity in the characterization of transport processes is the path length s and its distribution function $p(s)$. The path length s is defined as the distance that a particle travels between two interactions with the background medium. The path length distribution function is then defined by

$$p(s)\,\mathrm{d}s = \text{“Probability that a particle travels a distance } s' \in [s, s+\mathrm{d}s] \text{ without undergoing a collision.”}$$

This distribution can either be derived from statistical properties of the background medium, or it can be calculated from the steady state linear Boltzmann equation directly. Therefore, we assume one particle traveling through a purely absorbing medium, that means $\Sigma_t = \Sigma_a$ and $\Sigma_s = 0$. Starting at $x = 0$ into direction $\Omega = (1, 0, 0)^T$, the position x of the particle is the actual length s of its path. The probability that this particle travels a distance s without being

absorbed is in this case equivalent to the solution of the ordinary differential equation

$$\partial_s \Psi(s) + \Sigma_t \Psi(s) = 0, \quad \Psi(0) = 1,$$

that is just

$$\Psi(s) = \exp(-\Sigma_t s) = \mathbb{P}(\text{free path length } \geq s).$$

This fact is often called the Beer-Lambert or Beer-Lambert-Bouguer law [85]. The corresponding probability density is then obtained as the negative derivative of the distribution function $\Psi(s)$ (see Appendix A). This gives

$$p(s) = -\Psi'(s) = \Sigma_t \Psi(s) = \Sigma_t \exp(-\Sigma_t s). \tag{3.1}$$

Hence, we see that the classical equation can only reproduce an exponential path length distribution in a homogeneous medium. Note that this is as well the distribution of path lengths from an arbitrary position x to the next collision, as the distribution of path lengths between successive collisions. There are situations where these two quantities do not coincide (see Chapter 4).
In the following, we present two applications where it was shown that the transport of particles through a background medium is governed by a non-exponential path length distribution.

Remark 3.1. Davis and Marshak [32] showed, based on the classical transport equation, that the distribution of free paths in an inhomogeneous medium can never be an exponential. But non-exponential path length distributions are not necessarily caused by inhomogeneities in the background medium. It is important here to emphasize that there exist homogeneous media with an underlying non-exponential path length distribution. A new and non-classical equation is therefore needed to describe the transport therein.

3.2 Photon transport in atmospheric clouds

The interaction of solar radiation and atmospheric clouds is one of the major challenges in global climate modeling because of its influence on the radiation budget of the earth. It is still one of the main sources of uncertainty in current climate model predictions [85]. Current measurements of geometrical path lengths of photons in atmospheric clouds and their distributions provide strong evidence that the predictions of the classical theory of radiative transfer do not hold in many situations. There is still dissent in the scientific community on the underlying physical reasons.

In [89], the author investigates the scaling of the mean geometrical path length of transmitted light as a function of the vertical size of an atmospheric cloud. The classical theory of radiative transfer predicts a scaling law with an exponent, the so-called Lévy index, $\gamma = 2$. In contrast to the predictions of the

classical theory, measurements using oxygene A-band spectrometry yield Lévy indices in a range of $1 \leq \gamma \leq 2$. A Lévy index smaller than 2 indicates an underlying algebraic rather than exponential path length distribution. Using the same measurement techniques, the authors of [94] compare measured first and second moments of the total photon path length distribution to predictions of the classical theory, and anomalous diffusion theory, where the path lengths between individual scattering events are power-law distributed. They find that most of their measurements indicate a non-exponential distribution of path lengths, while a classical exponential distribution is only found for optically thick cloud covers.

In [48], the authors solve an inverse problem to estimate the actual distribution of path lengths. Therefore, measurements of the transmission of photons for different frequencies are used to reconstruct the underlying, frequency independent path length distribution by an inversion of a Laplace transformation. The measurements are again obtained using oxygen A-band spectroscopy. All resulting path length distributions for different cloud scenarios are gamma distributions, where the parameters vary with the cloud thickness.

It has been proposed [67, 68] that spatial correlations of the scatterers may be the reason for the non-exponential distribution of distances between scattering events. These spatial correlations are assumed to be caused by a clustering of the water droplets inside the atmospheric clouds [66], whereas the mechanism that causes the spatial correlations between scatterers is assumed to be the turbulent nature of the clouds [69]. A related explanation, not contradicting the aforementioned, is that the spatial structure of an atmospheric cloud is fractal. Measurements of the liquid water content [80, 84, 60] show scaling laws that indicate multifractal spatial structures. The transport in such media is then modeled by fractional diffusion equations [30, 31], which implicitly assume an underlying algebraically decaying path length distribution. These model calculations agree with the measurements in [89].

The Intercomparison of 3D Radiation Codes (I3RC) [24] is an ongoing project that publicly provides benchmark results and measurement data of the spatial structure, transmission and reflection of atmospheric clouds. It also serves as a platform for the exchange and comparison of different models and calculations regarding radiative transfer in this field. That makes the problem of transport in atmospheric clouds accessible for further research. Furthermore, none of the above discussed observations could be explained conclusively, and a transport model that is able to reproduce them is still missing. Therefore, this application is used as a motivational example throughout this work.

3.3 Lévy glass

A Lévy glass is an artificial disordered optical material. It is constructed by inserting small glass spheres into a host medium together with titanium dioxide

powder, that serves as scattering elements [7, 14]. An algebraically decaying path length distribution for ultraviolet light of the form

$$p(s) \sim s^{-(\alpha+1)}$$

for $0 < \alpha < 2$ was observed experimentally, and also justified theoretically in [14]. The path length distribution can be controlled by the sphere diameter distribution $p(\phi)$, where ϕ denotes the diameter of the glass sphere. For the relation between $p(s)$ and $p(\phi)$ an analytical expression is obtained in [14]. The distribution of sphere diameters is also an algebraically decaying function. Measurements of transmission profiles in [14] support these findings. Note that the spatial correlations of scatterers here are a result of the inhomogeneities in the medium on the microscale, since the glass spheres serve as void space when it comes to photon transport. Nevertheless, if the diameter of the glass spheres are small compared to the total size of the Lévy glass, it can be seen as a statistically homogeneous medium, although that was not the intenion of the authors in [7, 14].

In [56, 101] the transport of photons in a Lévy glass is investigated from a probabilistic random walk point of view. The authors of [56] consider the influence of correlations between scattering events, and find that the measured Lévy indices point towards a diverging second moment of the underlying path length distribution. On the other hand, in [101] the authors develop a probabilistic theory to estimate the first and second moment of the path length distribution in a Lévy glass, which depends on the sphere packing. For both monodisperse and fractal sphere packing they find that transport converges to a diffusion process in the long time limit, but with a non-classical diffusion constant, that indicates an underlying path length distribution different from the classical exponential.

To summarize: it is possible to construct background materials, where the decay of the path length distribution is non-exponential and the decay properties can be quantified exactly a priori. Transport in such materials cannot be described by the classical transport equation because of the aforementioned reasons. Furthermore, we note that non-exponential path length distribustions are not caused by a turbulent process, as it was proposed in atmospheric clouds, but by the fractal structure of the medium [101].

4 Rigorous derivation of a non-classical transport equation

The rigorous derivation of the linear Boltzmann equation presented in Chapter 2 heavily relies on the Poissonian distribution of scatterers, meaning that the spatial positions of the scatterers are uncorrelated. And we have also seen in Chapter 3 that this equation can only reproduce exponential path length distributions, whereas in several applications transport is obviously governed by an algebraically decaying distribution function. The spatial correlation of the position of scatterers in the background medium is assumed to be a reason for this kind of phenomena. Naturally, that brings up the question if it is possible to derive kinetic equations which are capable of reproducing non-exponential path length distributions. Especially the case of a periodic setting of scatterers has drawn a lot of attention in the past decades. This problem has been named periodic Lorentz gas. Originally, Lorentz considered a random configuration of scatterers to model the motion of electrons in metals [79]. The periodic model was first investigated by Sinai in [99], and a governing kinetic equation has been found very recently [22, 81]. The problem can be formulated in the following way:

- Given a background medium that consists of a periodic set of scatterers.
- Massless point particles travel through this medium.
- The point particles are scattered in case of collision.
- Is there a kinetic Boltzmann-like equation that governs this dynamics in the Boltzmann-Grad limit?

We proceed as follows. In the first Section 4.1 we review the results obtained by Golse et. al [53] in the case of specular reflecting scatterers. We sketch the derivation of the resulting kinetic equation, which is set on an extended phase-space, and especially introduce the notation. Then, in the following Section 4.2, we use these techniques to derive a non-classical transport equation for a periodic setting of scatterers. The difference to the theory in [22, 81] is that we now allow random, isotropic scattering, instead of a deterministic scattering process. This is a more realistic model when it comes to photon transport in atmospheric clouds or other radiation transport processes. We comment on possible extensions and generalizations of these results in Section 4.3.
The theory developed in this chapter relies on some deeper results and techniques in measure theory and the theory of stochastic processes. A short review of the theory that we use can be found in the Appendix A.

For reasons of clarity and simplicity the presentation of the theory in this chapter is restricted to the case of two space dimensions. But all results remain valid in arbitrary space dimensions.

4.1 The periodic Lorentz gas

One of the simplest cases where the positions of scatterers in a background medium are correlated, is the periodic Lorentz gas. In the periodic Lorentz gas we consider point particles that travel at speed 1 in the Euclidean plane. There are spherical obstacles with radius ε^2 placed on the grid $\varepsilon\mathbb{Z}^2$. When a point particle hits a scatterer it is specularly reflected, and therefore changes its direction of travel. A typical particle trajectory is sketched in Figure 4.1.

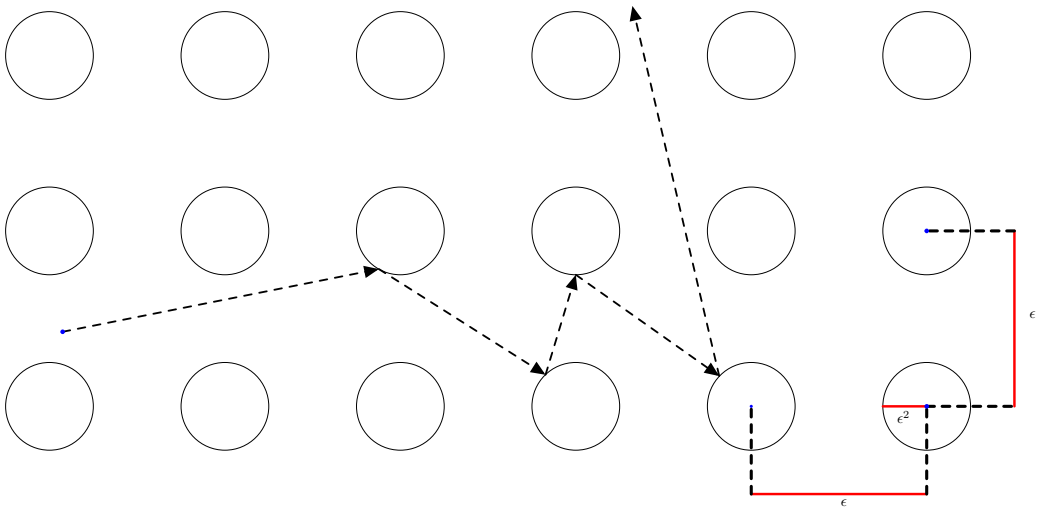


Figure 4.1: A trajectory of a particle in the periodic Lorentz gas. The distances between the scatterer centers is ε, their radius ε^2. In case of collision the travelling particle is specularly reflected.

For $\varepsilon \in (0, \frac{1}{2})$ we define by

$$Z_\varepsilon = \{x \in \mathbb{R}^2 \mid \operatorname{dist}(x, \varepsilon\mathbb{Z}^2) > \varepsilon^2\}$$

the domain in which the particles move. It can be shown [52], that the linear Boltzmann equation is not capable of describing the transport of particles in a background medium with a periodic setting of scatterers. Of course, as in the derivation of the classical linear Boltzmann equation in Section 2.2, this has to be understood in a homogenized sense. For $\varepsilon > 0$, the dynamics of the point particles that travel between the scatterers in Z_ε is governed by a classical linear Boltzmann equation with appropriate boundary conditions. We denote the solution of this equation by $f_\varepsilon \equiv f_\varepsilon(t, x, \Omega)$. In the limit $\varepsilon \to 0^+$, the following holds:

Theorem 4.1. *(Golse [52])*
There exist initial data $f^{in} \in C(\mathbb{R}^2)$ such that no subsequence of f_ε converges

in $L^\infty(\mathbb{R}_+ \times \mathbb{R}^D \times S^{D-1})$ *weak* to a solution* f *of a linear Boltzmann equation of the form*

$$(\partial_t + \Omega \cdot \nabla_x) f(t,x,\Omega) = \sigma_t\Big(\int_{S^1} \sigma_s(\Omega,\Omega') f(t,x,\Omega')\,\mathrm{d}\Omega' - f(t,x,\Omega)\Big),$$
$$f(0,x,\Omega) = f^{in}(x,\Omega),$$

where $\sigma_t > 0$ *is the total cross section and* $\sigma_s(\Omega,\Omega')$ *is a scattering kernel as introduced above.*

The reason for the failure of the classical theory here is again, that it was shown that the distribution of path lengths in the Periodic Lorentz gas decays algebraically, and not exponentially. Based on this observation, Caglioti and Golse [22], and independently Marklof and Strömbergsson [81], derived a non-classical kinetic equation for the periodic Lorentz gas with specular reflections. This is done via a homogenization technique in the Boltzmann-Grad limit similar to the classical derivation by Gallavotti [49]. The resulting kinetic equation lives in an extended phase space, and their derivations heavily rely on an analytic expression for the free path length distribution in the Boltzmann-Grad limit.

In the following, we sketch the derivation of this non-classical kinetic equation for the periodic Lorentz gas. Therefore, we first review the results for the path length distribution. At every point $x \in Z_\varepsilon$, and in any direction $\Omega \in S^1$ we define the free path as

$$\tau_\varepsilon(x,\Omega) := \inf\{s > 0 \mid x + s\Omega \in \partial Z_\varepsilon\} \qquad \text{for } (x,\Omega) \in Z_\varepsilon \times S^1.$$

Given (x,Ω), the function τ_ε denotes the distance to the next collision of a particle at x traveling in direction Ω. This function can be extended continuously to the boundary such that its domain of definition is

$$\big(Z_\varepsilon \times S^1\big) \cap \big(\Gamma_\varepsilon^+ \cup \Gamma_\varepsilon^-\big).$$

Now, by $\mathbb{1}_{\tau_\varepsilon(x,\Omega)>s}$ we define the free path distribution function at x in direction Ω. Let f be a function that is not everywhere defined in $\mathbb{R}^2$. In the following, we denote by $\{f\}$ the continuation of f by 0 to the full domain $\mathbb{R}^2$. It has been proven in [10] that

$$\{\mathbb{1}_{\tau_\varepsilon(x,\Omega)>s}\} \rightharpoonup q(s) \quad \text{in } L^\infty(\mathbb{R}_+ \times \mathbb{R}^2 \times S^1) \text{ weak}^* \tag{4.1}$$

as ε tends to 0. The limiting function has been computed explicitly [15] to be

$$q(s) = \int_s^\infty (\tau - s)\Theta(\tau)\,\mathrm{d}\tau$$

where Θ is the following piece wise defined function

$$\Theta(s) = \frac{24}{\pi^2} \begin{cases} 1 & \text{for } s \in (0, \frac{1}{2}], \\ \frac{1}{2s} + 2(1-\frac{1}{2s})^2 \log(1-\frac{1}{2s}) - \frac{1}{2}(1-\frac{1}{s})^2 \log|1-\frac{1}{s}| & \text{for } s \in (\frac{1}{2}, \infty). \end{cases}$$

This result was also found independently in [20], and is implied by the more general results in [81]. We note that $q(s)$ is the probability distribution function of the free path in the limit. That means, in the homogenized medium $q(s)$ has the interpretation

$$q(s) = \mathbb{P}\{\text{“distance to the next collision”} > s\}.$$

Hence, the probability density $p_0(s)$ for the path length to the first collision is given by

$$q(s) = \int_s^\infty p_0(s')\,\mathrm{d}s', \qquad \text{thus} \qquad p_0(s) = -q'(s),$$

and we directly calculate

$$p_0(s) = \int_s^\infty \Theta(\tau)\,\mathrm{d}\tau.$$

According to [36], the distribution of path lengths between collisions can be derived directly from these quantities. Let

$$Q(s) = \mathbb{P}\{\text{“distance between two collisions”} > s\}.$$

Then it holds that

$$Q(s) = -\frac{q'(s)}{q'(0)} = \frac{p_0(s)}{p_0(0)},$$

where the division by $q'(0)$ is the normalization of the distribution. Hence the probability density $p(s)$ for the path lengths between collisions is given by

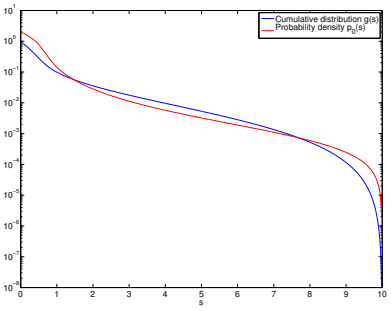
$$p(s) = -Q'(s) = -\frac{q''(s)}{q'(0)} = \frac{\Theta(s)}{\int_0^\infty \Theta(\tau)\,\mathrm{d}\tau}.$$

The probability densities p and p_0, and their corresponding cumulative distribution functions g and G are shown in Figure 4.2.

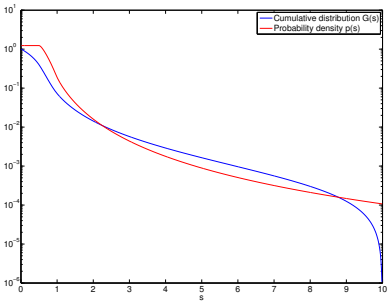
Remark 4.1. In classical transport theory, for an uncorrelated setting of scatterers, there is no difference between the distribution p_0 of free paths until the first collision, and the distribution p of path lengths between collisions. As we have seen in Chapter 2, in that situation there is $\sigma > 0$ such that

$$q(s) = \int_s^\infty p_0(s')\,\mathrm{d}s' = \exp(-\sigma s), \quad \text{hence} \quad p_0(s) = \sigma \exp(-\sigma s).$$

On the other hand, for the distance between collisions, according to [36], it



(a) distances to first collision



(b) distances between collisions

Figure 4.2: Cumulative distribution functions and probability densities for distances to the first and between collisions in the Boltzmann-Grad limit.

holds that

$$p(s) = \frac{q''(s)}{q'(0)} = \frac{\sigma^2 \exp(-\sigma s)}{\sigma} = \sigma \exp(-\sigma s) = p_0(s).$$

That explains why in classical transport theory one usually does not distinguish between these two quantities.

Caglioti and Golse [22] and Marklof and Strömbergsson [81] found a non-classical kinetic equation that describes the dynamics of the periodic Lorentz gas in the Boltzmann-Grad limit, hence, that is able to reproduce the underlying path length distribution. That kinetic equation is set on an extended phase space, that means in addition to the classical phase space variables x (space-variable) and Ω (direction-of-flight), two more independent variables are needed. The first is the distance s to the next collision, whose distribution has already been discussed. The second is the so-called impact factor h at the the next collision. For a particle that collides with a scatterer with radius r at the position x with outward pointing normal $n(x)$ and incoming direction Ω, the impact factor is defined as

$$h_r(x, \Omega) := \sin(\sphericalangle(n(x), \Omega)).$$

Defining

$F(t, x, \Omega, s, h) =$ "density of particles at time t and position x
that travel in direction Ω with distance s to the next collision,
where it will collide with impact factor h",

the resulting kinetic equation for F then reads

$$\partial_t F(t,x,\Omega,h,s) + \Omega\nabla_x F(t,x,\Omega,h,s) - \partial_s F(t,x,\Omega,h,s) = \tag{4.2a}$$

$$\int_{-1}^{1} P(s,h|h')F(t,x,R[\pi - 2\arcsin(h')]\Omega,0,h')\,\mathrm{d}h', \tag{4.2b}$$

$$F(0,x,\Omega,s,h) = f^{\mathrm{in}}(x,\Omega)\int_s^\infty \int_{-1}^{1} P(\tau,h|h')\,\mathrm{d}h'\,\mathrm{d}\tau. \tag{4.2c}$$

with $s \in [0,\infty)$ and $h \in [-1,1]$. Here, $P(s,h|h')$ is the transition kernel, denoting the probability that a particle that collides with impact factor h' has a new distance s to the next collision with impact factor h. The recently introduced path length distributions p_0 and p are related to this quantity by

$$p_0(s) = \int_s^\infty \int_{-1}^{1}\int_{-1}^{1} P(\tau,h|h')\,\mathrm{d}h'\,\mathrm{d}h\,\mathrm{d}\tau, \text{ hence } \; p(s) = \int_{-1}^{1}\int_{-1}^{1} P(s,h|h')\,\mathrm{d}h'\,\mathrm{d}h. \tag{4.3}$$

The rotation of an angle θ is again denoted by $R[\theta]$. Therefore, the in-scattering term (4.2b) can be interpreted as the particles that have recently scattered (since $s = 0$ in F) with impact parameter h', that are assigned a new direction of flight Ω and a new distance to collision s and impact factor h by the transition kernel P. The classical, physically relevant particle density $\Psi^* = \Psi^*(t,x,\Omega)$ can be obtained by integrating over the additional kinetic variables

$$\Psi^*(t,x,\Omega) = \int_0^\infty \int_{-1}^{1} F(t,x,\Omega,h,s)\,\mathrm{d}h\,\mathrm{d}s.$$

The main result can now be formulated as follows. It can be found in [22, 23, 53] or [81]. As in the derivation of the classical transport equation, we denote by $f_\varepsilon = f_\varepsilon(t,x,\Omega)$ the particle distribution in the periodic Lorentz gas with finite scatterer radius $r = \varepsilon^2$ and distance ε between lattice points.

Theorem 4.2. *(Golse, Caglioti; Marklof, Strömbergsson)*
Let $f^{in} \in C_C(\mathbb{R}^2 \times S^1)$ be a positive initial particle distribution. Then one has

$$f_\varepsilon(t,x,\Omega) \longrightarrow \int_0^\infty \int_{-1}^{1} F(t,x,\Omega,s,h)\,\mathrm{d}h\,\mathrm{d}s$$

in $L^\infty(\mathbb{R}_+ \times \mathbb{R}^2 \times S^1)$ weak as $\varepsilon \to 0^+$, where F is the unique solution of (4.2).*

We shortly sketch the proof of this theorem and follow again the presentation in [53]. For reasons of clarity and simplicity, we do not go into detail when it comes to some complicated theory and technical constructions that are not of importance for the remainder of this work. We refer to the already cited works for a complete and detailed proof.

In a finite size scatterer setting with radius $r = \varepsilon^2 > 0$, for a particle that

collides with a scatterer traveling in direction Ω with impact factor h', we define the transfer map

$$T_r(h', \Omega) = (s, h), \text{ where } \begin{cases} s = \text{ “distance to the next collision”}, \\ h = \text{ “impact factor at the next collision”}. \end{cases}$$

Particle trajectories in the periodic Lorentz gas are completely determined by this transfer map, since the new direction of travel can be calculated from Ω. The existence of the limit of T_r as $r \to 0^+$ can be shown for almost all $\Omega \in S^1$. The proof relies on a continued fraction expansion of the slope of Ω, therefore it only holds in dimension 2 [21]. To be more precise, T_r converges in distribution to the transition probability as

$$\frac{1}{|\log \eta|} \int_\eta^{1/4} f(T_r(h', \Omega)) \frac{\mathrm{d}r}{r} \longrightarrow \int_0^\infty \int_{-1}^1 f(s, h) P(s, h|h') \, \mathrm{d}h \, \mathrm{d}s,$$

for almost all $\Omega \in S^1$, and for every $f \in C(\mathbb{R}_+ \times [-1, 1])$ as $\eta \to 0^+$ [22, 23]. Note that this transition probability is independent of the direction Ω. The explicit calculation of $P(s, h|h')$ was done in [23], using another representation of the slope of Ω called Farey fractions, together with the results on the free path distribution in [15]. Using an additional independence assumption, simply formulated that the variables s and h become independent in the Boltzmann-Grad limit (this is called the assumption (H) in [22, 23, 53]), we can construct the particle trajectories explicitly. The trajectory of a particle with initial setting (x, Ω, s, h) in the limit $r \to 0^+$ is then fully determined by the sequence $(s_n, h_n)_{n \geq 1}$, whose distribution is known now (according to assumption (H)). The sequence of collision times is then

$$T_n = s_0 + \cdots + s_{n-1}, \ n \geq 1,$$

and the new direction of travel after the $n-$th collision can be computed from the impact factor h_n. The trajectory at time $t \geq 0$ is then given by the current position, direction of travel, distance to the next collision and impact factor at the next collision. We denote this by the stochastic process $(X_t, \Omega_t, S_t, H_t)(x, \Omega, s, h)$, and it holds that

$$\begin{aligned} X_t(x, \Omega, s, h) &= x + t\Omega, & S_t(x, \Omega, s, h) &= s - t, \\ \Omega_t(x, \Omega, s, h) &= \Omega, & H_t(x, \Omega, s, h) &= h, \end{aligned}$$

for $0 \leq t < s$, that means before the first collision. Between the n-th and $n+1-$th collision we have $T_n < t < T_{n+1}$ and

$$\begin{aligned} X_t(x, \Omega, s, h) &= x + (t - T_n)\Omega_n, & S_t(x, \Omega, s, h) &= s - T_{n+1}, \\ \Omega_t(x, \Omega, s, h) &= \Omega_n, & H_t(x, \Omega, s, h) &= h_n. \end{aligned}$$

In order to derive Equation (4.2) we denote by $F^{\text{in}}(x, \Omega, s, h)$ the initial distribution in the extended phase space, and let $\Phi = \Phi(x, \Omega, s, h)$ be an observable,

where $\Phi \in C_C^\infty(R^2 \times S^1 \times \mathbb{R}_+ \times [-1,1])$. We then define the density $F(t,\cdot,\cdot,\cdot,\cdot)$ by

$$\iiiint \Phi(x,\Omega,s,h)F(t,\,\mathrm{d}x,\,\mathrm{d}\Omega,\,\mathrm{d}s,\,\mathrm{d}h) \tag{4.4a}$$

$$= \iiiint \mathbf{E}[\Phi(x,\Omega,s,h)]F^{\mathrm{in}}(x,\Omega,s,h)\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s\,\mathrm{d}h, \tag{4.4b}$$

where $\mathbf{E}[\cdot]$ is the expectation on the random quantities involved. Furthermore, we define $g(t,x,\Omega,s,h) := \mathbf{E}[\Phi(x,\Omega,s,h)]$, the so-called Kolmogorov operator (compare Appendix A). We can split g by

$$g(t,x,\Omega,s,h) := \mathbf{E}[\mathbb{1}_{t<s}\Phi(x,\Omega,s,h)] + \mathbf{E}[\mathbb{1}_{s<t}\Phi(x,\Omega,s,h)],$$

since $\mathbf{E}$ is a linear operator. The first term can be easily calculated as

$$\mathbf{E}[\mathbb{1}_{t<s}\Phi(x,\Omega,s,h)] = \mathbb{1}_{t<s}\Phi(x+t\Omega,\Omega,s-t,h),$$

since there is no stochasticity involved until the first collision. The second term can be computed, using the Markov property of the stochastic process (X_t,Ω_t,S_t,H_t), to be

$$\mathbf{E}[\mathbb{1}_{s<t}\Phi(x,\Omega,s,h)] = \\ \mathbb{1}_{s<t}\iint P(s_1,h_1|h)g(t-s,x+t\Omega,R(2\arcsin(h)-\pi)\Omega,s_1,h_1)\,\mathrm{d}s_1\,\mathrm{d}h_1.$$

Hence, we have

$$g(t,x,\Omega,s,h) = \mathbb{1}_{t<s}\Phi(x+t\Omega,\Omega,s-t,h) \\ + \mathbb{1}_{s<t}\iint P(s_1,h_1|h)g(t-s,x+t\Omega,R(2\arcsin(h)-\pi)\Omega,s_1,h_1)\,\mathrm{d}s_1\,\mathrm{d}h_1,$$

and this is a solution of the backward Kolmogorov equation

$$\partial_t g + \Omega\cdot\nabla_x g + \partial_s g = 0, \qquad g(0,x,\Omega,s,h) = \Phi,$$

$$g(t,x,\Omega,0,h) = \iint_{\mathbb{R}_+\times(-1,1)} P(s_1,h_1|h)g(t,x,R(2\arcsin(h)-\pi)\Omega,s_1,h_1)\,\mathrm{d}h_1\,\mathrm{d}s_1,$$

for $t,s>0, x\in\mathbb{R}^2, \Omega\in S^1, h\in[-1,1]$. The corresponding forward Kolmogorov equation, obtained by taking the adjoints of all operators but the time derivative (see again Appendix A), is now Equation (4.2). That means, the evolution of an initial particle distribution under the stochastic process (X_t,Ω_t,S_t,H_t) is described by this equation. That finishes the proof.

Remark 4.2. In their paper [81], Marklof and Strömbergsson were able to prove this result in greater generality. They do not need an independence assumption as used in [22, 23, 53]. Furthermore, they show that Theorem 4.2 is valid for more general lattices in d dimensions and scattering maps corresponding

to smooth radial potentials. In a subsequent paper [82] they extend these results to quasicrystalline scatterer configurations, which are deterministic, non-periodic but strongly correlated. The difference to the proof presented above lies mainly in the techniques that were used to obtain the transition kernel P in the Boltzmann-Grad limit. Marklof and Strömbergsson consider measures on abstract space of lattices, and obtain the same explicit expressions in dimension 2. Another generalization of this theory was presented in [83]. There, Marklof and Strömbergsson consider transport in polycrystals, which consist of grains of the size of the order of a mean free path. The scatterers inside the grains are fixed on a lattice, but the lattices of different grains are shifted and turned. A non-classical transport equation of the form of (4.2) is obtained where the transition kernel can be derived from the one for the simple lattice, but becomes x- and Ω dependent. Furthermore, it is shown that the path length distribution in this case decays exponentially, but is different from the classical path length distribution (3.1).

Remark 4.3. Again, we conclude this derivation by pointing out details that are of special importance for the derivation of a non-classical transport equation in the next section.

(1) The impact factor h is a necessary additional variable here, since the new direction of travel and the distance to the next collision depend on the position where the scatterer is hit by the point particle. It was shown in [81] that the underlying stochastic process is a Markov process with memory 2, that means that it is not possible to describe the dynamics in the limit with less than two extra independent variables. If, on the other hand, the scattering map is independent of the position of collision on the scatterer surface, we can expect that h is not needed anymore as an independent variable.

(2) Equation (4.2) is valid in arbitrary space dimensions [81], although the result as presented above only holds for $d = 2$. For higher dimensions there are no explicit formulas for the transition kernel P, hence, for the path length distributions p_0 and p. But the existence of these quantities in the Boltzmann-Grad limit is guaranteed (see again [81]).

(3) The distributions of free-path length p_0 and p are both included in the Equation (4.2). Initial distances to collision are distributed as p_0, while distances to collisions of particles that are in-scattered are distributed as p (according to (4.3)).

(4) In contrast to the results for the Poissonian configuration of scatterers, the convergence in the Boltzmann-Grad limit here is only weak in L^∞. The main reason is that already the convergence of the transfer map, hence of the transition kernel, only holds in distribution.

4.2 Homogenization of radiative transfer

In this section, we apply the ideas and techniques used in the derivation of a kinetic equation for the periodic Lorentz gas to the process of radiative transfer through a periodic setting of scatterers. We also restrict ourselves to the 2D case such that all previous results can be directly applied. The resulting equation is a non-classical transport equation set on a phase space that is only extended by the path length variable s.

As a simplified model for radiative transfer we consider point particles that travel at speed 1 in the Euclidean plane. Exactly as in the case of the periodic Lorentz gas there are spherical obstacles (or scatterers) with radius ε^2 placed on the grid $\varepsilon\mathbb{Z}^2$ and particles travel within the domain Z_ε. When a point particle hits an obstacle it is thermalized and instantly reemitted into another (random) direction. Now we note that each obstacle is uniquely characterized by its center coordinates εk for $k \in \mathbb{Z}^2$ and the boundary ∂Z_ε of the obstacles can be parametrized as

$$x = \varepsilon k + \varepsilon^2 n, \qquad \text{for } n \in S^1,\ x \in \partial Z_\varepsilon.$$

At each point x on the boundary we can define the pairs (x, Ω) of incoming and outcoming directions at the boundary as

$$\Gamma_\varepsilon^+ = \{(x,\Omega) | x = \varepsilon k + \varepsilon^2 n,\ n, \Omega \in S^1,\ \Omega \cdot n > 0\},$$
$$\Gamma_\varepsilon^- = \{(x,\Omega) | x = \varepsilon k + \varepsilon^2 n,\ n, \Omega \in S^1,\ \Omega \cdot n < 0\},$$

where $\cdot$ denotes the scalar product in $\mathbb{R}^2$. Furthermore, at every point $x \in Z_\varepsilon$ in any direction $\Omega \in S^1$ we define the free path as

$$\tau_\varepsilon(x,\Omega) := \inf\{s > 0 | x + s\Omega \in \partial Z_\varepsilon\} \qquad \text{for } (x,\Omega) \in Z_\varepsilon \times S^1.$$

We proceed as follows. First, we describe the transport process in the domain Z_ε for $\varepsilon > 0$. We define a stochastic process that describes the evolution of a single particle. That process includes the current path length of the particle, but the evolution in the classical phase space can still be described by the the classical linear Boltzmann equation. The dynamics in the limit $\varepsilon \to 0^+$ can also be described by a stochastic process in the homogenized medium. There we make use of the explicit knowledge of the resulting path length distribution in the Boltzmann-Grad limit. We derive the corresponding backward and forward Kolmogorov equations for this stochastic process. In a last step, we prove the convergence in the limit as $\varepsilon \to 0^+$ to the solution of a non-classical transport equation.

4.2.1 Dynamics in Z_ε

The dynamics of particles for $\varepsilon > 0$ can be described naturally by a stochastic process. We describe the transport of particles in the domain Z_ε by a Markov process $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)$, where X_t^ε is the position, Ω_t^ε the direction of flight, and

S_t^ε the distance to the next collision at time $t \geq 0$. The evolution of the corresponding probability density of this process then includes the solution of the classical linear Boltzmann equation in the domain Z_ε with appropriate boundary conditions and a delta distribution for the path length variable.

The probabilistic model works as follows. We want to define a stochastic process that describes the evolution of a particle that has initial position x and initial direction of flight Ω. All notations and important definitions on probability and measure theory are collected in the Appendix A. The main point here is that we also include the distance to the next collision of this particle in our probabilistic description. In other words, we extend the phase space in that this dynamics happens by an additional dimension. Between scatterers, a particle travels along straight lines. Stochasticity in this description comes only into play if a particle hits a scatterer with center coordinate c_i. Then the random quantities are

- the point $x = c_i + \varepsilon^2 n$ on the boundary of the scatterer where the particle is reemitted, uniquely determined by the normal vector n, and
- the new direction of flight Ω, which has to satisfy $\Omega \cdot n > 0$ such that it is an outgoing direction.

Since we assume isotropic scattering, the distribution of the random variables (n, Ω) is given by the joint Borel probability measure

$$\mathbf{b}(\,\mathrm{d}n\,\mathrm{d}\Omega) := \frac{1}{2\pi^2}(\Omega \cdot n)_+ \,\mathrm{d}n\,\mathrm{d}\Omega.$$

So let

$$(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x, \Omega, s)$$

be the position, direction of flight and distance to the next collision at time $t > 0$ for initial values (x, Ω, s). Then we can construct the trajectory of a particle explicitly. This trajectory is uniquely determined by the sequence of random scattering events, that can be described by the sequence of points on the boundary of Z_ε. So pick a sequence

$$(n_j, \Omega_j)_{j \geq 1} \subset G = \{(n, \Omega) \in S^1 \times S^1 \mid \Omega \cdot n > 0\}$$

distributed under the Borel probability measure b. The particle will travel in straight lines between the scattering events. Let $[x] \in \varepsilon\mathbb{Z}$ be the the scatterer center that is closest to $x \in Z_\varepsilon$, and define $\tau_j^\varepsilon := \tau_\varepsilon(\varepsilon^2 n_j, \Omega_j)$ to be the path length between the $j-1-$th and $j-$th collision.
This is independent of the specific scatterer, since the setting of scatterers is periodic. Set also

$$T_j^\varepsilon := \tau_1^\varepsilon + \cdots + \tau_j^\varepsilon.$$

Since particles are assumed to travel by speed 1, T_j^ε is just the time that has

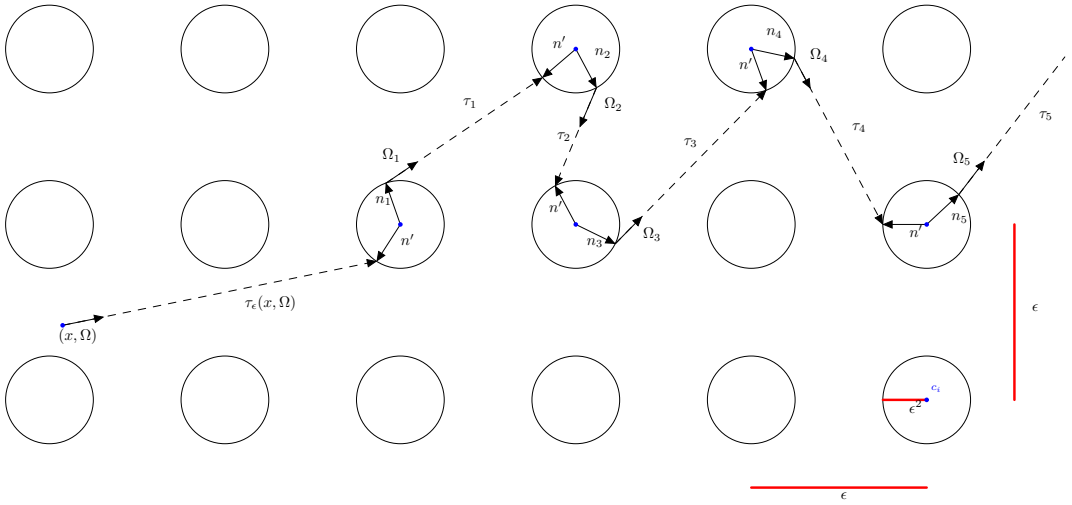


Figure 4.3: A path of $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)$, starting at (x, Ω), in the periodic domain Z_ε with grid size ε.

elapsed between the first and the $j-$th collision event. Fixing an initial setting (x, Ω, s) for the particle, the trajectory until the first collision is given by

$$(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x, \Omega, s) = (x + t\Omega, \Omega, s - t).$$

After the first collision, the trajectory is given by

$$(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x, \Omega, s) = ([X_{s-0}^\varepsilon] + \varepsilon^2 n_1 + (t - s)\Omega_1, \Omega_1, s + T_1^\varepsilon - t)$$

for $s \leq t < s + T_j^\varepsilon$. Between the $j-1-$th and $j-$th collision the trajectory is now given as

$$(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x, \Omega, s) = ([X_{s+T_j^\varepsilon-0}^\varepsilon] + \varepsilon^2 n_{j+1} + (t - s - T_j^\varepsilon)\Omega_{j+1}, \Omega_{j+1}, s + T_{j+1}^\varepsilon - t).$$

This is a complete description of the particle trajectory in the extended phase-space $Z^\varepsilon \times S^1 \times \mathbb{R}_+$. A sample path which is described by this stochastic process is shown in Figure 4.3. The additional variable S_t^ε can be seen as a simple counter-variable. It does not affect the evolution in $Z_\varepsilon \times S^1$. Therefore, we can state an equation for the evolution of the probability densities which is independent of s.

The evolution of a probability density function of $(X_t^\varepsilon, \Omega_t^\varepsilon)$ in $Z_\varepsilon \times S^1$ can be described by the solution $f_\varepsilon = f_\varepsilon(t, x, \Omega)$ of a classical linear Boltzmann equation with appropriate boundary conditions. Let $f_\varepsilon = f_\varepsilon(t, x, \Omega)$ be the density of photons at position x traveling into direction Ω at time $t \geq 0$. We assume isotropic scattering. Then between the scatterers, f_ε satisfies the following transport equation

$$\partial_t f_\varepsilon + \Omega \nabla_x f_\varepsilon = 0 \qquad x \in Z_\varepsilon,\ \Omega \in S^1. \tag{4.5}$$

Once a particle hits a scatterer at a position $x' \in \partial Z_\varepsilon$ having traveled into direction Ω', it is absorbed, and immediately reemitted at $x \in \partial Z_\varepsilon$ into direction Ω with probability $1/2\pi^2$, where $x = \varepsilon k + \varepsilon^2 n$ and $\Omega \cdot n > 0$ such that the new direction is outgoing. The scattering process is sketched in Figure 4.4.

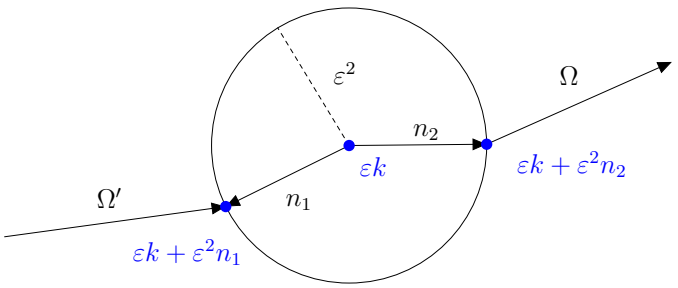


Figure 4.4: Scattering from direction Ω' to Ω.

That means, the value at the boundary in outgoing direction is given by the integral over all particles that have hit the scatterer, divided by the uniform probability. Therefore, the boundary condition for f_ε is given by

$$f_\varepsilon(t,x,\Omega) = \frac{1}{2\pi^2} \iint_{S^1\times S^1} f(t, \varepsilon k + \varepsilon^2 n, \Omega')(\Omega' \cdot n)_- \,\mathrm{d}\Omega' \,\mathrm{d}n$$
$$x \in \partial Z_\varepsilon,\ x = \varepsilon k + \varepsilon^2 n,\ \Omega \cdot n > 0.$$

Furthermore, an initial density at time 0 is assigned to f_ε by

$$f_\varepsilon(0,x,\Omega) = f^{\mathrm{in}}(x,\Omega), \qquad x \in Z_\varepsilon,\ \Omega \in S^1.$$

Since the evolution of the free path S_t^ε is deterministic and depends on x and Ω, its probability density is simply given by the Dirac measure $\delta(s - \tau_\varepsilon(x,\Omega))$. Hence, the evolution of the corresponding joint probability density of the stochastic process $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)$ is given by

$$f_\varepsilon(t,x,\Omega)\delta(s - \tau_\varepsilon(x,\Omega)).$$

Let $\Phi(x,\Omega,s) \in C_c(\mathbb{R}^2 \times S^1 \times \mathbb{R}_+)$ be a test function. We define the corresponding Kolmogorov operator of $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)$ by

$$g_\varepsilon(t,x,\Omega,s) := \mathbf{E}[\Phi((X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x,\Omega,s))].$$

Viewing Φ as an observable of the system, $g_\varepsilon(t,x,\Omega,s)$ describes its evolution in time. The densites $F_\varepsilon, f_\varepsilon$ and g_ε are related in the following way. We can make use of a fundamental connection between the solution of the Kolmogorov operator and the probability density of the stochastic process (see Appendix A). It is a probabilistic version of the classical characteristics formula for trans-

port equations and reads

$$\iiint \Phi(x,\Omega,s)F_\varepsilon(t,x,\Omega,s)\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s = \tag{4.6a}$$

$$\iiint \mathbf{E}[\Phi((X_t^\varepsilon,\Omega_t^\varepsilon,S_t^\varepsilon)(x,\Omega,s))]f^{\mathrm{in}}(x,\Omega)\delta(s-\tau_\varepsilon(x,\Omega))\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s. \tag{4.6b}$$

That means g_ε and F_ε are related by some form of duality. This is an important point later on in the convergence proof.

4.2.2 Dynamics in the limit

Now we describe the dynamics in the limit for $\varepsilon \to 0$. Since the scatterer size vanishes in the limit, we are now dealing with a homogenized background medium. The sources of stochasticity in the description of particle trajectories are now

- the path length s between two scattering events and
- the new direction of flight Ω after a scattering event.

In the case $\varepsilon > 0$, the probability that a path length in Z_ε is larger than s is given by

$$Q_\varepsilon(s) := \mathbf{p}(\{(n,\Omega) \in G \mid \tau_\varepsilon(\varepsilon^2 n,\Omega) \geq s\}).$$

Recall, that in the limit $\varepsilon \to 0^+$ this goes to

$$Q_\varepsilon(s) \longrightarrow Q(s) := -\frac{q'(s)}{q'(s)} = \frac{\int_s^\infty \Theta(\tau)\,\mathrm{d}\tau}{\int_0^\infty \Theta(\tau)\,\mathrm{d}\tau}.$$

Hence, the probability density for the path lengths between two successive collisions in the limit is given by

$$p(s) = -Q'(s) = \frac{\Theta(s)}{\int_0^\infty \Theta(\tau)\,\mathrm{d}\tau}. \tag{4.7}$$

In order to define a stochastic process in the extended phase space $\mathbb{R}^2 \times S^1 \times \mathbb{R}_+$ we pick a sequence of independent random variables $(\Omega_j, \tau_j)_{j\geq 1}$. Then Ω_j is the new direction of flight after the $j-$th scattering event and τ_j is the path length between the $j-$th and $(j+1)$-th collision. This sequence, together with an initial position and direction of flight, determines the trajectory of a single particle. Because of isotropic scattering and (4.7), this sequence is distributed under

$$\frac{1}{2\pi}p(s)\,\mathrm{d}\Omega\,\mathrm{d}s. \tag{4.8}$$

Let $s \in \mathbb{R}_+$, we define a random walk in the path length variable by

$$T_0 := s, \quad T_j := s + \tau_1 + \ldots + \tau_j, \quad j \geq 1.$$

Then, the dynamics in the extended phase space $\mathbb{R}^2 \times S^1 \times \mathbb{R}_+$ can be described by the jump process (X_t, Ω_t, S_t). This is defined as follows. Fix an initial setting (x, Ω, s) for the particle. The trajectory until the first collision is given by

$$(X_t, \Omega_t, S_t)(x, \Omega, s) := (x + t\Omega, \Omega, T_0 - s).$$

After the first collision, the trajectory is given by

$$(X_t, \Omega_t, S_t)(x, \Omega, s) = (X_{T_0} + (t - T_0)\Omega_1, \Omega_1, T_1 - t)$$

for $s \leq t < s + T_j$. Between the $(j-1)-$th and $j-$th collision the trajectory is now given as

$$(X_t, \Omega_t, S_t)(x, \Omega, s) = (X_{T_j} + (t - T_j)\Omega_{j+1}, \Omega_{j+1}, T_{j+1} - t).$$

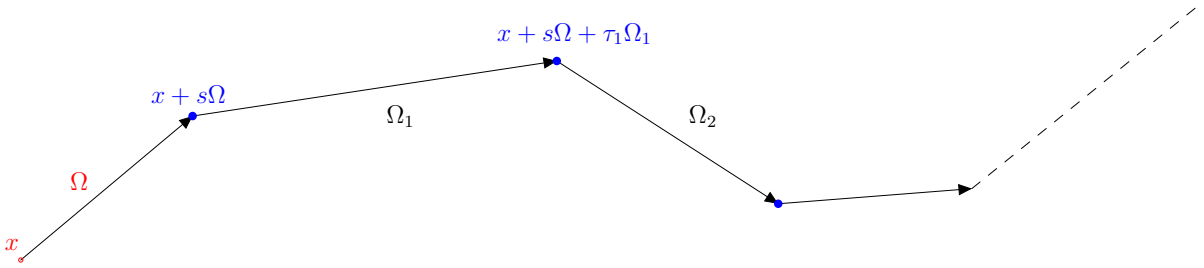


Figure 4.5: Trajectory defined by $(X_t, \Omega_t, S_t)(x, \Omega, s)$.

Choose an initial particle density $f^{\text{in}}(x, \Omega)p_0(s)$, and an observable $\Phi = \Phi(x, \Omega, s)$ in $C_c(\mathbb{R}^2 \times S^1 \times \mathbb{R}_+)$. The Kolmogorov operator (compare again Appendix A) is defined as

$$g(t, x, \Omega, s) := \mathbf{E}[\Phi((X_t, \Omega_t, S_t)(x, \Omega, s))],$$

and as in (4.4), the density $F(t, \cdot, \cdot, \cdot)$ of the stochastic process is defined by

$$\iiint \Phi(x, \Omega, s)F(t, \mathrm{d}x, \mathrm{d}\Omega, \mathrm{d}s) = \tag{4.9a}$$

$$\iiint \mathbf{E}[\Phi((X_t, \Omega_t, S_t)(x, \Omega, s))]f^{\text{in}}(x, \Omega)p_0(s)\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s. \tag{4.9b}$$

Because of the linearity of $\mathbf{E}$ we can write g as

$$\begin{aligned} g(t,x,\Omega,s) =& \mathbf{E}[\mathbb{1}_{t<s}\Phi((X_t,\Omega_t,S_t)(x,\Omega,s))] \\ &+ \mathbf{E}[\mathbb{1}_{s\le t<T_1}\Phi((X_t,\Omega_t,S_t)(x,\Omega,s))] \\ &+ \sum_{j=1}^{\infty}\mathbf{E}[\mathbb{1}_{T_j\le t<T_{j+1}}\Phi((X_t,\Omega_t,S_t)(x,\Omega,s))]. \end{aligned}$$

We now evaluate this expression term by term. First note that

$$\mathbf{E}[\mathbb{1}_{t<s}\Phi((X_t,\Omega_t,S_t)(x,\Omega,s))] = \mathbb{1}_{t<s}\Phi(x+t\Omega,\Omega,s-t)$$

since there is no stochasticity in the trajectory $(X_t,\Omega_t,S_t)(x,\Omega,s)$ until the first collision, meaning that $t<s$ (see Figure 4.5).
The second term is

$$\begin{aligned} &\mathbf{E}[\mathbb{1}_{s\le t<T_1}\Phi((X_t,\Omega_t,S_t)(x,\Omega,s))] \\ &\quad= \mathbf{E}[\mathbb{1}_{s\le t<T_1}\Phi((X_{t-s},\Omega_{t-s},S_{t-s})(x+s\Omega,\Omega_1,T_1-s))] \\ &\quad= \mathbf{E}[\mathbb{1}_{s\le t<T_1}\mathbf{E}[\Phi((X_{t-s},\Omega_{t-s},S_{t-s})(x+s\Omega,\Omega_1,\tau_1))\mid\Omega_1,\tau_1]], \end{aligned}$$

since the stochastic process is Markovian on the extended phase space. Because of this property, we have

$$\mathbf{E}[\Phi((X_{t-s},\Omega_{t-s},S_{t-s})(x+s\Omega,\Omega_1,\tau_1))\mid\Omega_1,\tau_1] = g(t-s,x+s\Omega,\Omega_1,\tau_1).$$

Hence,

$$\begin{aligned} &\mathbf{E}[\mathbb{1}_{s\le t<T_1}\mathbf{E}[\Phi((X_{t-s},\Omega_{t-s},S_{t-s})(x+s\Omega,\Omega_1,\tau_1))\mid\Omega_1,\tau_1]] \\ &\quad= \frac{1}{2\pi}\iint \mathbb{1}_{s\le t<T_1}g(t-s,x+s\Omega,\Omega_1,s_1)p(s_1)\,\mathrm{d}\Omega_1\,\mathrm{d}s_1. \end{aligned}$$

In the same way we can calculate the remaining terms. For $T_1\le t<T_2$ we have

$$\begin{aligned} &\mathbf{E}[\mathbb{1}_{T_1\le t<T_2}\Phi((X_t,\Omega_t,S_t)(x,\Omega,s))] \\ &\quad= \mathbf{E}[\mathbb{1}_{T_1\le t<T_2}\Phi((X_{t-T_1},\Omega_{t-T_1},S_{t-T_1})(x+s\Omega+\tau_1\Omega_1,\Omega_2,\tau_2))] \\ &\quad= \mathbf{E}[\mathbb{1}_{T_1\le t<T_2}\mathbf{E}[\Phi((X_{t-T_1},\Omega_{t-T_1},S_{t-T_1})(x+s\Omega+\tau_1\Omega_1,\Omega_2,\tau_2))\mid\Omega_1,\tau_1,\Omega_2,\tau_2]]. \end{aligned}$$

Again, we can evaluate the inner expectation explicitly as

$$\begin{aligned} &\mathbf{E}[\Phi((X_{t-T_1},\Omega_{t-T_1},S_{t-T_1})(x+s\Omega+\tau_1\Omega_1,\Omega_2,\tau_2))\mid\Omega_1,\tau_1,\Omega_2,\tau_2] \\ &\quad= g(t-T_1,x+s\Omega+\tau_1\Omega_1,\Omega_2,\tau_2), \end{aligned}$$

and obtain

$$\begin{aligned} &\mathbf{E}[\mathbb{1}_{T_1\le t<T_2}\mathbf{E}[\Phi((X_{t-T_1},\Omega_{t-T_1},S_{t-T_1})(x+s\Omega+\tau_1\Omega_1,\Omega_2,\tau_2))\mid\Omega_1,\tau_1,\Omega_2,\tau_2]] \\ &\quad= \frac{1}{4\pi^2}\iiiint \mathbb{1}_{T_1\le t<T_2}g(t-T_1,x+s\Omega+s_1\Omega_1,\Omega_2,\tau_2)p(s_1)p(s_2)\,\mathrm{d}s_1\,\mathrm{d}s_2\,\mathrm{d}\Omega_1\,\mathrm{d}\Omega_2. \end{aligned}$$

This calculation can be repeated for every $T_j \leq t < T_{j+1}$, and we obtain in an analogous manner

$$\begin{aligned}
&\mathbf{E}[\mathbb{1}_{T_j \leq t < T_{j+1}} \Phi((X_t, \Omega_t, S_t)(x, \Omega, s))] \\
&\quad = \frac{1}{(2\pi)^{j+1}} \int \cdots \int \mathbb{1}_{T_j \leq t < T_{j+1}} g\left(t - T_j, x + s\Omega + \sum_{k=1}^{j} (s_k \Omega_k), \Omega_{j+1}, \tau_{j+1}\right) \\
&\qquad \prod_{k=1}^{j} (p(s_k)\, \mathrm{d}s_k)\ \mathrm{d}\Omega_1 \cdots \mathrm{d}\Omega_j.
\end{aligned}$$

Hence, $g(t, x, \Omega, s)$ is a solution of the following integral equation:

$$\begin{aligned}
g(t, x, \Omega, s) &= \mathbb{1}_{t<s} \Phi(x + t\Omega, \Omega, s - t) \\
&\quad + \frac{1}{2\pi} \int_0^\infty \int_{S^1} \mathbb{1}_{s \leq t < T_1} g(t - s, x + s\Omega, \Omega_1, s_1) p(s_1)\, \mathrm{d}\Omega\, \mathrm{d}s_1 \\
&\quad + \sum_{j=1}^{\infty} \frac{1}{(2\pi)^{j+1}} \int_0^\infty \cdots \int_0^\infty \int_{S^1} \cdots \int_{S^1} \mathbb{1}_{T_j \leq t < T_{j+1}} \\
&\qquad g\left(t - T_j, x + s\Omega + \sum_{k=1}^{j} (s_k \Omega_k), \Omega_{j+1}, \tau_{j+1}\right) \prod_{k=1}^{j} (p(s_k)\, \mathrm{d}s_k)\ \mathrm{d}\Omega_1 \cdots \mathrm{d}\Omega_j.
\end{aligned}$$

Similar to the previous derivations of the classical equation and Equation (4.2), this is the solution formula of a partial differential equation. To be more precise, g is a solution of the backward Kolmogorov equation

$$\partial_t g(t, x, \Omega, s) - \Omega \nabla_x g(t, x, \Omega, s) + \partial_s g(t, x, \Omega, s) = 0, \tag{4.10a}$$

$$g(t, x, \Omega, 0) = \frac{1}{2\pi} \int_0^\infty \int_{S^1} p(s) g(t, x, \Omega, s)\, \mathrm{d}\Omega\, \mathrm{d}s, \tag{4.10b}$$

$$g(0, x, \Omega, s) = \Phi(x, \Omega, s). \tag{4.10c}$$

The corresponding Fokker-Planck equation for the probability density $F(t, x, \Omega, s)$ of the stochastic process (X_t, Ω_t, S_t) is now given by the adjoint of the semi-group of the above equation (see Appendix A). It then reads as

$$\partial_t F(t, x, \Omega, s) + \Omega \nabla_x F(t, x, \Omega, s) - \partial_s F(t, x, \Omega, s) = \tag{4.11a}$$

$$\frac{p(s)}{2\pi} \int_{S^1} F(t, x, \Omega, 0)\, \mathrm{d}\Omega, \tag{4.11b}$$

$$F(0, x, \Omega, s) = p_0(s) f^{\mathrm{in}}(x, \Omega). \tag{4.11c}$$

A detailed discussion of Equations (4.10) and (4.11) can be found in Chapter 6. For a formal derivation of the adjoint operators, we refer to Chapter 12, where similar calculations are carried out in detail.

4.2.3 Homogenization

For the homogenization we show that the sequence f_ε has a limit f as $\varepsilon \to 0$ and analyze this limit. It turns out that f does not satisfy a classical kinetic equation, but that it is the $s-$integral of a solution of a non-classical kinetic equation. We follow the reasoning of [10], where a similar homogenization problem is considered.

In order to be able to deal with all given objects in the same spaces, we recall the notation $\{y\}(x)$ for the extension of a function $y(x)$ to the full space by 0. Hence, $\{f_\varepsilon\}$ denotes the extension of f_ε to the area of the scatterers by 0. In order to homogenize the process of radiative transfer in the periodic background medium we prove the following theorem.

Theorem 4.3. *In the situation described above,*

$$\{f_\varepsilon\} \rightharpoonup^* f = \int_0^\infty \mathrm{d}F \qquad \text{as } \varepsilon \to 0^+,$$

where f_ε, f and F are measures, and F is the solution of Equation (4.11).

We recall that the path length variable s was introduced into the setting of Equation (4.5) by defining the following measure

$$F_\varepsilon(t, x, \Omega, \mathrm{d}s) := f_\varepsilon(t, x, \Omega)\delta(s - \tau_\varepsilon(x, \Omega))\,\mathrm{d}s. \tag{4.12}$$

Note that F_ε is only a generalized function, so we can only expect that the following theorem holds. We also recall, that $F_\varepsilon(t, x, \Omega, \mathrm{d}s)$ is the probability density of the stochastic process $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)$.

Then the proof falls naturally into the following five steps:

(1) For every $\varepsilon > 0$ we have $f_\varepsilon = \int_0^\infty F_\varepsilon \,\mathrm{d}s$.

(2) The set $\{f_\varepsilon\}, \varepsilon > 0$, is bounded, hence it has a convergent subsequence with respect to the weak*-topology by the theorem of Banach-Alaoglu[1], denoted by f.

(3) Furthermore, there is $\tilde{F}$, such that $\{F_\varepsilon\} \rightharpoonup^* \tilde{F}$, again by the theorem of Banach-Alaoglu.

(4) It holds, that $\tilde{F} = F$, where F is a solution of (4.11).

(5) Finally, we conclude that this F is unique and

$$\{f_\varepsilon\} \to f = \int_0^\infty \mathrm{d}F$$

as $\varepsilon \to 0^+$, again in the weak measure sense.

[1]For an underlying Banach space X the theorem of Banach-Alaoglu states that the unit ball $B^* \subset X^*$ is weakly* compact. If X is separable, it is also sequentially compact in the weak* topology. Hence, every bounded subset of X^* has a weak* congergent subsequence.

Proof. We prove the theorem above by showing, that (1)-(5) hold.

(1) Fix $\varepsilon > 0$. The probability density $f_\varepsilon(t, x, \Omega)$ is a measure on the phase space $Z_\varepsilon \times S^1$ for all $t \geq 0$. Hence, we choose an arbitrary test function $\phi \in C_b(Z_\varepsilon \times S^1)$ and calculate

$$\begin{aligned}
&\iint_{Z_\varepsilon \times S^1} \phi(x,\Omega) \int_0^\infty F_\varepsilon(t,x,\Omega,s)\,\mathrm{d}s\,\mathrm{d}\Omega\,\mathrm{d}x \\
&\quad = \iint_{Z_\varepsilon \times S^1} \phi(x,\Omega) \int_0^\infty \delta(s-\tau_\varepsilon(x,\Omega)) f_\varepsilon(t,x,\Omega)\,\mathrm{d}s\,\mathrm{d}\Omega\,\mathrm{d}x \\
&\quad = \iint_{Z_\varepsilon \times S^1} \phi(x,\Omega) f_\varepsilon(t,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x.
\end{aligned}$$

That shows, that $f_\varepsilon = \int F_\varepsilon \,\mathrm{d}s$ as a measure on $Z_\varepsilon \times S^1$. Obviously, replacing f_ε by $\{f_\varepsilon\}$ and choosing a test function $\phi \in C_b(\mathbb{R}^2 \times S^1)$ in the calculation above does not change the result. Consequently, $\{f_\varepsilon\} = \int F_\varepsilon \,\mathrm{d}s$ as a measure on $\mathbb{R}^2 \times S^1$.

(2) The classical theory of the linear Boltzmann equation guarantees the existence of a unique solution of (4.5) such that

$$\begin{aligned}
&\iint_{Z_\varepsilon \times S^1} f_\varepsilon(t,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x \leq \iint_{Z_\varepsilon \times S^1} f^{\mathrm{in}}(x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x \quad \text{for all } t \geq 0, \\
&0 \leq \sup_{(x,\Omega)\in\mathbb{R}^2\times S^1} f_\varepsilon(t,x,\Omega) \leq \|f^{\mathrm{in}}\|_{L^\infty(\mathbb{R}^2\times S^1)} \quad \text{a.e. on } \mathbb{R}_+ \times Z_\varepsilon \times S^1.
\end{aligned}$$

Therefore, the extension $\{f_\varepsilon\}$ also satisfies

$$\sup_\varepsilon \|\{f_\varepsilon\}\|_{L^\infty(\mathbb{R}_+\times\mathbb{R}^2\times S^1)} \leq \|f^{in}\|_{L^\infty(\mathbb{R}^2\times S^1)}.$$

Then, by the theorem of Banach-Alaoglu, there is a subsequence $\{f_{\varepsilon'}\} \subset \{f_\varepsilon\}$, and there is a measure f such that

$$\{f_{\varepsilon'}\} \rightharpoonup^* f.$$

(3) Note that

$$\begin{aligned}
\{F_\varepsilon\}(t,x,\Omega,s) &= \{\delta(s-\tau_\varepsilon(x,\Omega)) f_\varepsilon(t,x,\Omega)\} \\
&= \delta(s-\tau_\varepsilon(x,\Omega))\{f_\varepsilon(t,x,\Omega)\},
\end{aligned}$$

if we extend $\tau_\varepsilon(x,\Omega)$ by 0 to $x \notin Z_\varepsilon, \Omega \in S^1$. This extension is not continuous. The extended $\{F_\varepsilon\}$ is a measure on $\mathbb{R}^2 \times S^1 \times \mathbb{R}_+$ for all $t \geq 0$ and $\varepsilon > 0$.
Now let $\phi \in C_b(\mathbb{R}^2 \times S^1 \times \mathbb{R}_+)$ and let $\phi_n \in C_C(\mathbb{R}^2 \times S^1 \times \mathbb{R}_+)$, $n \in \mathbb{N}$ such that $\phi_n \longrightarrow \phi$ pointwise and $\|\phi_n\|_\infty \leq \|\phi\|$ for all $n \in \mathbb{N}$. Furthermore,

let $\operatorname{supp}(\phi_n) = G_n \times S^1 \times [0, T_n]$. We have that

$$\iiint_{\mathbb{R}^2\times S^1\times\mathbb{R}_+} \phi_n(x,\Omega,s)\{F_\varepsilon\}(t,x,\Omega,s)\,\mathrm{d}s\,\mathrm{d}\Omega\,\mathrm{d}x = \iint_{\mathbb{R}^2\times S^1} \phi_n(x,\Omega,\tau_\varepsilon(x,\Omega))f_\varepsilon(t,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x$$

for all $t \geq 0$ and $n \in \mathbb{N}$. Recall that $\tau_\varepsilon(x,\Omega)$ is neither bounded, nor continuous. Therefore, we write

$$\begin{aligned}\phi_n(x,\Omega,\tau_\varepsilon(x,\Omega)) =& \mathbb{1}_{\tau_\varepsilon(x,\Omega)>T_n}\phi_n(x,\Omega,\tau_\varepsilon(x,\Omega)) \\ &+ \mathbb{1}_{\tau_\varepsilon(x,\Omega)<T_n}\phi_n(x,\Omega,\tau_\varepsilon(x,\Omega)).\end{aligned}$$

Since

$$\mathbb{1}_{\tau_\varepsilon(x,\Omega)>T_n} + \mathbb{1}_{\tau_\varepsilon(x,\Omega)<T_n} = 1,$$

and $\{\mathbb{1}_{\tau_\varepsilon(x,\Omega)>s}\} \rightharpoonup^* q(s)$ in $L^\infty(\mathbb{R}^2\times S^1\times\mathbb{R}_+)$, we have

$$\{\mathbb{1}_{\tau_\varepsilon(x,\Omega)<s}\} \rightharpoonup^* 1-q(s) \text{ in } L^\infty(\mathbb{R}^2\times S^1\times\mathbb{R}_+), \tag{4.13}$$

with q as introduced in Section 4.1. The compact support of ϕ_n yields

$$\mathbb{1}_{\tau_\varepsilon(x,\Omega)>T_n}\phi_n(x,\Omega,\tau_\varepsilon(x,\Omega)) = 0,$$

and, taking into account (4.13) and that $q(s) \to 0$ as $s \to \infty$, we have

$$|\mathbb{1}_{\tau_\varepsilon(x,\Omega)<T_n}\phi_n(x,\Omega,\tau_\varepsilon(x,\Omega))| \leq \|\phi\|_\infty$$

for all $n \in \mathbb{N}$ and $\varepsilon > 0$. That means that

$$\left|\iiint_{\mathbb{R}^2\times S^1\times\mathbb{R}_+} \phi_n(x,\Omega,s)F_\varepsilon(t,x,\Omega,\,\mathrm{d}s)\,\mathrm{d}\Omega\,\mathrm{d}x\right| \leq \|\phi_n\|_\infty \iint_{\mathbb{R}^2\times S^1} |f_\varepsilon(t,x,\Omega)|\,\mathrm{d}\Omega\,\mathrm{d}x$$

holds for all $n \in \mathbb{N}$ and $\varepsilon > 0$. Finally, letting $n \to \infty$, we obtain

$$\left|\iiint_{\mathbb{R}^2\times S^1\times\mathbb{R}_+} \phi(x,\Omega,s)F_\varepsilon(t,x,\Omega,\,\mathrm{d}s)\,\mathrm{d}\Omega\,\mathrm{d}x\right| \leq \iint_{\mathbb{R}^2\times S^1} \|\phi\|_\infty f_\varepsilon(t,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x$$
$$\leq \|\phi\|_\infty \cdot \|f_\varepsilon(t,\cdot,\cdot)\|_{L^1(\mathbb{R}^2\times S^1)} \leq \|\phi\| \cdot \|f^{\mathrm{in}}\|_{L^1(\mathbb{R}^2\times S^1)}$$

for all test functions $\phi \in C_b(\mathbb{R}^2\times S^1\times\mathbb{R}_+)$ and every $\varepsilon > 0$. That means, that $\{F_\varepsilon\}$ is bounded in the weak* topology. Again, we can conclude by the theorem of Banach-Alaoglu, that there is a measure $\tilde{F}$ on $\mathbb{R}^2\times S^1\times\mathbb{R}_+$ such that

$$\{F_\varepsilon\} \longrightarrow \tilde{F} \quad \text{in the weak measure sense.}$$

(4) The stochastic process $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)$ in the finite scatterer setting is uniquely determined by the initial position and direction of flight, and the sequence

$$(n_j^\varepsilon, \Omega_j^\varepsilon)_{j\geq 1} \subset \{(n,\Omega) \in S^1 \times S^1 \mid \Omega \cdot n > 0\}$$

of direction of flights Ω_j^ε after the $j-$th collision, and the corresponding new position on the surface of the scatterer, given by n_j. The exact position on the surface of the scatterer with center coordinate c_j can be calculated as $c_j + \varepsilon^2 n_j^\varepsilon$. The distance to the next collision after the $j-th$ collision is then given by

$$\tau_j^\varepsilon(\Omega_j, n_j) := \tau_\varepsilon(c_j + \varepsilon^2 n_j^\varepsilon, \Omega_j^\varepsilon),$$

and each trajectory $(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x,\Omega,s)$ can be described by the sequence $(\Omega_j^\varepsilon, n_j^\varepsilon, \tau_j^\varepsilon(\Omega_j^\varepsilon, n_j^\varepsilon))$. A trajectory $(X_t, \Omega_t, S_t)(x,\Omega,s)$, on the other hand, is determined by the sequence

$$(\Omega_j, \tau_j)_{j\geq 1} \subset S^1 \times \mathbb{R}_+$$

of new directions of flight after the $j-$th collision Ω_j, and new distances to the next collision τ_j. Obviously, the random variable Ω_j^ε converges in distribution to Ω_j as $\varepsilon \to 0^+$, since they share exactly the same distribution function. The results on τ_ε, stated in (4.1), implies that

$$\tau_j^\varepsilon \longrightarrow \tau_j$$

in distribution as $\varepsilon \to 0^+$. The random variable n_j becomes obsolete in the Boltzmann-Grad limit, since the size of the scatterers shrinks to 0. Therefore, we conculde that

$$(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon) \longrightarrow (X_t, \Omega_t, S_t)$$

in distribution as $\varepsilon \to 0^+$. This implies (see Appendix A), that for every $\Phi \in C_b(\mathbb{R}^2 \times S^1 \times \mathbb{R}_+)$

$$\mathbf{E}[\Phi(X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)] \longrightarrow \mathbf{E}[\Phi(X_t, \Omega_t, S_t)].$$

Because of the relations (4.6) and (4.9) of the Kolmogorov operator and the distribution of the stochastic processes it holds that

$$\begin{aligned}
&\iiint \Phi(x,\Omega,s) F_\varepsilon(t,x,\Omega,s)\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s \\
&\quad = \iiint \mathbf{E}[\Phi((X_t^\varepsilon, \Omega_t^\varepsilon, S_t^\varepsilon)(x,\Omega,s))] f^{\mathrm{in}}(x,\Omega)\delta(s-\tau_\varepsilon(x,\Omega))\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s \\
&\quad \longrightarrow \iiint \mathbf{E}[\Phi((X_t, \Omega_t, S_t)(x,\Omega,s))] f^{\mathrm{in}}(x,\Omega) p_0(s)\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s \\
&\quad = \iiint \Phi(x,\Omega,s) F(t,x,\Omega,s)\,\mathrm{d}x\,\mathrm{d}\Omega\,\mathrm{d}s,
\end{aligned}$$

for every $\Phi \in C_b(\mathbb{R}^2 \times S^1 \times \mathbb{R}_+)$ and every $t \geq 0$. This shows that

$$F_\varepsilon \rightharpoonup^* F.$$

(5) Combining (1) and (3) we see that

$$f \overset{*}{\leftharpoonup} f_\varepsilon = \int_0^\infty \mathrm{d}F_\varepsilon \rightharpoonup^* \int_0^\infty \mathrm{d}\tilde{F}.$$

Since Equation (4.11) is linear, the solution is unique and therefore (4) yields

$$f = \int_0^\infty \mathrm{d}\tilde{F} = \int_0^\infty \mathrm{d}F.$$

That finishes the proof.

□

4.3 Discussion and consequences

We close this chapter on the rigorous derivation by commenting on certain aspects of the techniques that we have used, and the mathematical and physical nature of the involved quantities. Furthermore, we point out some consequences for the applicability and extendibility of this new mathematical model.

(A) We have seen in the previous sections how a non-classical transport equation can be derived from a stochastic process in a periodic setting of scatterers via homogenization in the Boltzmann-Grad limit. In order to derive the non-classical kinetic equation only the stochastic description of the already homogenized transport process (X_t, Ω_t, S_t) was necessary. Nevertheless, this full homogenization process yields additional insights into the nature of the new and non-classical parameters, and also establishes a strong connection between the classical and the non-classical equation.

(B) While classical transport is mainly driven by the total cross section Σ_t, the most important parameter in the non-classical equation is the path length distribution $p(s)$. The two parameters are of a very different nature. The total cross section Σ_t is usually interpreted as a (homogenized) rate, by that particles undergo collision. The function $p(s)$ is a probability density, that describes the distributions of free-path lengths between collisions. Nevertheless, the total cross section Σ_t and the free-path distribution are related by $p(s) = \Sigma_t \exp(-\Sigma_t s)$ in the case of classical transport. In Chapter 6, we show how a quantity related to the total cross section can be introduced into the non-classical setting.

(C) Equation (4.11) has both the probability density $p_0(s)$ of distances to the first collision, and the probability density $p(s)$ of distances between

collisions as a parameter. Nevertheless, we have pointed out in Section 4.1, that these two quantities are directly related. In order to calculate $p_0(s)$ and $p(s)$, only the cumulative distribution $q(s)$ of distances to the first collision needs to be known. This relation is of a very general nature and holds for arbitrary settings of scatterers [36]. Therefore, this is the quantity that we consider in the following chapter when we study more general random media in the Boltzmann-Grad limit.

(D) The scattering was assumed to be isotropic, but there is no problem in introducing an anisotropic scattering kernel into the equation. As long as the medium is homogeneous we can expect that the cumulative distribution of path lengths to the first collision (as the limit of $\mathbb{1}_{\tau_\varepsilon(x,\Omega)>s}$) exists and that it is independent of x and Ω, hence leads to a density $p_0 = p_0(s)$ and corresponding $p(s)$. Then, introducing a general scattering kernel σ_s, the probability density (4.8) can be replaced by

$$p(s)\sigma_s(\Omega,\Omega')\,\mathrm{d}s\,\mathrm{d}\Omega'.$$

The following calculations still holds, and we obtain again Equation (4.11), where the scattering kernel $\sigma_s(\Omega,\Omega')$ replaces the factor $1/2\pi$ in the scattering source term. Note that we cannot assume that this derivation holds in cases where the distribution of path lengths depends on x and/or Ω in the limit. Extensions of the non-classical equation are discussed in Chapter 6.

(E) The restriction to the two dimensional setting is mostly of a technical nature. The construction of the stochastic processes or the resulting equations are valid for arbitrary space dimensions. The existence of limiting path length distributions in the periodic Lorentz gas for arbitrary space dimensions has been proven in [15]. Furthermore, the corresponding results for the periodic Lorentz gas also holds (compare Remark 4.2). The main advantage of the two dimensional setting is that an explicit form of the limiting path length distribution is available. This strongly suggest that the above derivation is also valid for space dimensions $d \geq 2$, different only in the form of the distribution functions.

Consequences

Once an analytic expression for the distribution of path lengths is available, we were able to derive the non-classical equation. This also relied on the homogeneity of the background medium. Also, there is only one realization of the periodic background medium. We did not have to average over all possible realizations of the medium, in contrast to the derivation of the classical linear Boltzmann equation.
This suggests, that the equation is also valid for more general correlated background media as long as they are homogeneous. Of course, an additional averaging process over all realizations of a medium of that type is necessary. In the following chapter we investigate the validity of this assumption. Especially we

study the effect of random perturbations on the deterministic correlations in the periodic Lorentz gas, and how general spatial correlated background media can be constructed that also yield non-exponential path length distributions (in the Boltzmann-Grad limit, averaged over realizations).

5 Extension to general homogeneous correlated media: a Monte-Carlo study

Motivated by the rigorous derivation of a non-classical transport equation for the periodic scatterer setting in the previous chapter, we now discuss the extension to general homogeneous background media. Therefore, we note that the dynamics in the homogenized system is governed by the path length distribution in the Boltzmann-Grad limit. We are (at least at this point) not able to extend the rigorous derivation to general correlated media, but we can analyze the path length distributions that arise for certain spatial correlations in the Boltzmann-Grad limit by a Monte-Carlo method. The main goal of this chapter is the following: we show that there exists a continuum of correlated but homogeneous stochastic media besides the periodic configuration of scatterers that have a path length distribution in the Boltzmann-Grad limit that is different from the classical exponential distribution. Hence, the non-classical equation can be used to describe particle transport in these media.

Therefore, we study two different types of spatially correlated media. First, we calculate the distribution of path lengths for increasingly strong random perturbations of the periodic scatterer setting by a Monte-Carlo method. Then we express spatial correlations by a pair-correlation function, and use this to create random and spatially correlated background media by a Metropolis Monte-Carlo algorithm, and again calculate the path length distributions in the Boltzmann-Grad limit. As in the previous chapter, we restrict ourselves to the 2D case.

5.1 Spatially correlated background media

We make the following general assumptions for the background medium:

1. A scatterer X is a spherical object with radius $r > 0$. Its position in space is given by the coordinates of its center. We identify each scatterer with its center coordinates. These coordinates are assumed to be the outcome of a random variable.

2. Two scatterers are allowed to overlap. This is necessary since we want to be able to quantify the degree of spatial correlations that we introduce. In addition, in the Boltzmann-Grad limit the probability of overlaps becomes neglectable.

3. All scatterers are contained in a finite volume V, which is formally extended to infinity by assuming periodic boundaries.

4. The density of scatterers is the same as in the periodic Lorentz gas. That means, if $r = \varepsilon^2$, the number of scatterers in a box of volume 1 is $1/\varepsilon^2$. Of course, this only holds in a statistical sense, since we consider random media. Hence, the Boltzmann-Grad limit for all media that we consider is again $\varepsilon \to 0^+$.

5.1.1 Random perturbations of the periodic Lorentz gas

An obvious random but correlated medium to consider is a random perturbation of the periodic Lorentz gas. Therefore, we perturb the center coordinates (x, y) of each scatterer on the periodic lattice to $(\tilde{x}, \tilde{y})$, such that

$$\tilde{x} \sim \mathcal{N}(x, \sigma^2), \quad \tilde{y} \sim \mathcal{N}(y, \sigma^2).$$

That means, the perturbed center coordinates are Gaussian distributed around the original lattice points, where the variance $\sigma^2 = \kappa\varepsilon$ is proportional to the grid size ε, and the radius of the scatterers is $r = \varepsilon^2$. In Figure 5.1 we show a sector of the unperturbed periodic medium (Figure 5.1a), and perturbations whith variance 5% of ε (Figure 5.1b), 10% of ε (Figure 5.1c), and 20% of ε (Figure 5.1d). We see that the spatial correlations disappear for an increasing variance of the perturbation. Hence, we expect to see a transition in the path length distribution from the strong algebraic decay in the periodic medium to an exponential decay as the variance increases.

5.1.2 Quantification of spatial correlations

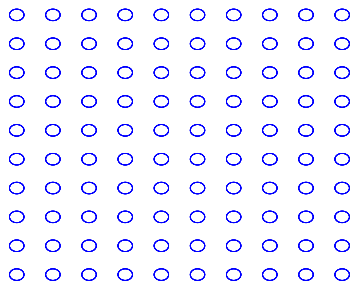
We first give the definition of a spatially uncorrelated medium, and then quantify the degree of correlation in a medium as its deviation from the uncorrelated case. This is done in terms of the so-called pair correlation function. In this section we follow [98], where mathematical measures of spatial correlations in atmospheric clouds are introduced, and we use the same terminology and notation.
Assume that we have N scatterers in a finite volume V. A medium is called **(spatially) uncorrelated** or **Poisson distributed** if the following counting process

$$N(\tilde{V}) = \text{number of scatterers in the volume } \tilde{V}$$

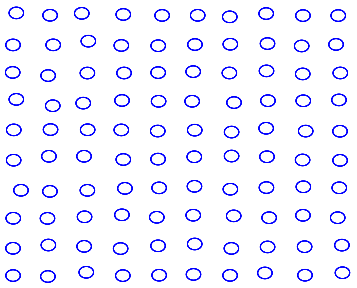
is Poisson distributed, i.e. its distribution function is given by

$$\mathbb{P}(N(\tilde{V}) = k) = \frac{(\lambda \operatorname{vol}(\tilde{V}))^k}{k!} \exp(-\lambda \operatorname{vol}(\tilde{V}))$$

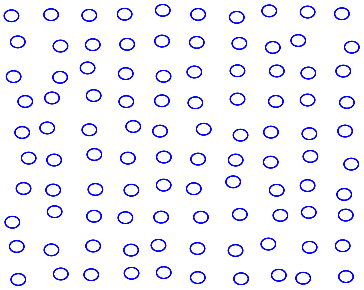
for some constant $\lambda > 0$ and $\tilde{V} \subset V$ arbitrary but measurable. This distribution implies that the positions of the scatterers, seen as an outcome of a random process, are **uncorrelated**. Let $V_1, V_2 \subset V$ be two volumes which



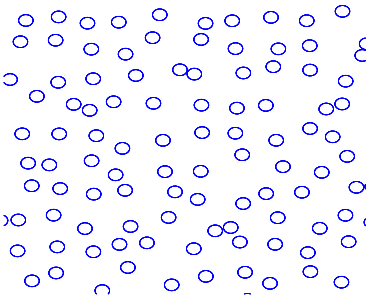
(a) Unpertubed periodic medium $\sigma^2 = 0$.



(b) Perturbed with variance $\sigma^2 = 0.05\varepsilon$.



(c) Perturbed with $\sigma^2 = 0.1\varepsilon$.



(d) Perturbed with $\sigma^2 = 0.2\varepsilon$.

Figure 5.1: Perturbed periodic Medium for different variances.

are sized such that they can contain only one or zero scatterers. Then the probability of finding scatterers in both volumes is given by the product

$$\mathbb{P}(1,2) = (\bar{n}\,\mathrm{vol}(V_1))(\bar{n}\,\mathrm{vol}(V_2)),$$

where $\bar{n} = N/\,\mathrm{vol}(V)$ is the mean scatterer number density. A **correlated medium** is now characterized by its deviation from the Poissonian distribution. The probability of finding two scatterers in these two given test volumes in the correlated case is given by

$$\mathbb{P}(1,2) = (\bar{n}\,\mathrm{vol}(V_1))(\bar{n}\,\mathrm{vol}(V_2))(1+\eta(r)),$$

where η is the so-called **pair-correlation function** and r is the distance between the two test volumes. If $\eta \equiv 0$ the medium is uncorrelated, otherwise it is spatially correlated. The pair-correlation function is related to the **radial distribution function** g by

$$\eta(r) = g(r) - 1.$$

The radial distribution function has the following interpretation: if we fix a scatterer X in V, then the local averaged density at a distance r from X is $\bar{n}g(r)$. The medium is called **homogeneous**, if the distribution of scatterers

and all its moments are invariant under translations, i.e. η is a global parameter of the system of scatterers. Otherwise, it is called **inhomogeneous**. We only consider homogeneous media here.

5.1.3 Generation of a correlated background medium

The generation of a spatially correlated medium is a very complex task. In [70], the authors use simplified heuristics to create either positively or negatively correlated media. In both cases, the correlation is introduced by inserting scatterers one after the other, where the position of an inserted scatterer depends on the position of the previously inserted ones. We follow a different strategy here, that is still of a heuristical nature, but allows a finer quantification of the correlation that we introduce.
First, the degree of correlation has to be defined by a pair-correlation function or radial distribution function. We always model the radial distribution function by taking a simple functional ansatz which can be parametrized by a small set of parameters. In order to be consistent with the spatial scales, this function should depend on the scatterer radius $r = \varepsilon^2$. For example, the following radial distribution function

$$g_\varepsilon(\tau) = \begin{cases} 10^{-20}, & \text{if} \quad 0 \leq \tau \leq 2\varepsilon^2, \\ \sin(A\pi(\tau - 2\varepsilon^2)) + 1 + 10^{-20}, & \text{if } \tau > 2\varepsilon^2, \end{cases} \tag{5.1}$$

describes a medium, where

- scatterers do not overlap (due to the repelling first condition),
- scatterers cluster on a small scale (due to the positivity of the sine near $\tau = 2\varepsilon$),
- neighbouring clusters have a distance determined by the parameter A, and between there are only few scatterers, because of the oscillations of the sine.

Then, we choose a number $N \gg 1$ of scatterers, a radius $r = \varepsilon^2 > 0$ and a finite volume $V \subset \mathbb{R}^2$. Now in a first step, scatterers are inserted by generating the center coordinates randomly, and the center coordinates are stored and the coordinates are binned in an additional coordinate index. In this way we generate a spatially uncorrelated medium and denote this setting of scatterers by $X = \{X_1, \ldots, X_N\}$.

In a second step, we perform the following procedure, which is known in the theory of molecular dynamics as "Metropolis Monte-Carlo Method" [86]. The applicability of this method is motivated by the following relation of the radial distribution function and a certain potential energy. We want to generate a scatterer setting that has spatial correlations which can be quantified by some pair-correlation function η or the corresponding radial distribution function g. Therefore, we assume that this setting is the static equilibrium state of a dynamic process driven by a pair potential $V(r)$ (this is an artificial assumption,

we do not want to say anything about the actual dynamics that is responsible for the spatial correlations!). Assuming that the distribution of a sample of scatterers, which are in equilibrium, is described by a radial distribution function $g(r)$, the corresponding pair potential $V(r)$ is given by

$$V(r) = -C\log(g(r)),$$

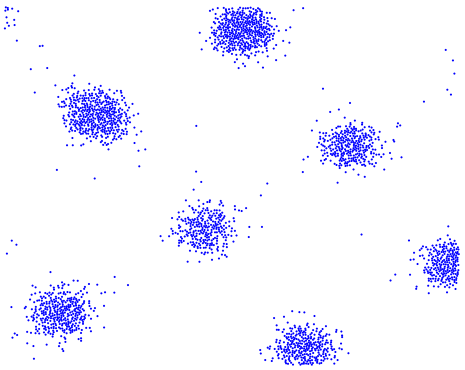
where $C > 0$ is a positive constant. See [26] for the theoretical background of this relation. Since we assume that the setting of scatterers is stationary, the kinetic energy of the system is 0 and hence, the total energy of the system can be defined as

$$E = \frac{1}{2}\sum_{i,j=1,i\neq j}^{N} V(r_{ij}), \qquad r_{ij} = \mathrm{dist}(X_i, X_j).$$

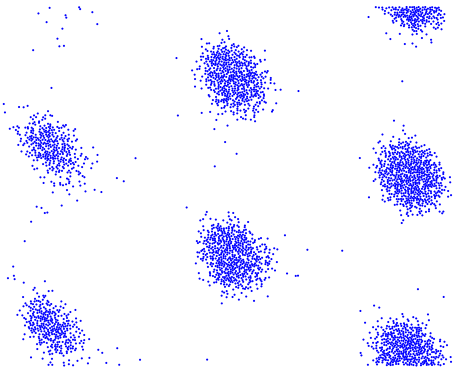
Therefore, starting with the uncorrelated setting generated in step one, we can transform it into a correlated system by the following procedure. We move scatterers in order to minimize the total energy of the system:

1. Calculate the energy E of the full system.
2. Pick a scatterer randomly (or a small number of scatterers).
3. Move the scatterer(s) randomly to another position (within a certain maximal distance from its original position).
4. Calculate the change of the total energy $\Delta E = E_{\mathrm{new}} - E$.
5. If $\Delta E < 0$ accept the new position(s). If $\Delta E > 0$ accept the new position(s) with a probability of $\exp(-E/C)$, reject otherwise.
6. Return to 1, if the stopping criterion is not yet satisfied.

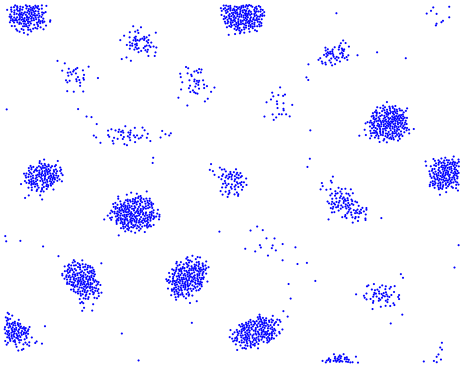
We stop this method if the change of the potential energy over a certain number of steps becomes small. Then the medium can be assumed to be converged. Figure 5.2 shows several sectors of correlated media that have been produced with this algorithm, with $r = 0.01$ and 10000 scatterers. Approximately 15 million Metropolis moves were needed to converge the algorithm. The blue dots are placed at the center coordinates of each scatterer. The first row (Figures 5.2a, 5.2b) shows the same sector of two different realizations of correlated media for the radial distribution function (5.1) with $A = 1$. We see that the scatterers form random, but spherical clusters, and that the cluster centers form a grid, where the distance between the cluster centers is 2, which coincides with the distance between the local maxima of the radial distribution function. Note that the orientation of this underlying grid is random. The same effects can be seen in the second row (Figures 5.2c, 5.2d), where (5.1) with $A = 2$ was used, and in the third row (Figures 5.2e, 5.2f), where (5.1) with $A = 4$ was used to create the media. Again, the scatterers form spherical clusters where the cluster centers form a grid with random orientation. The clusters become smaller and the distance between the cluster centers decreases, since the radial



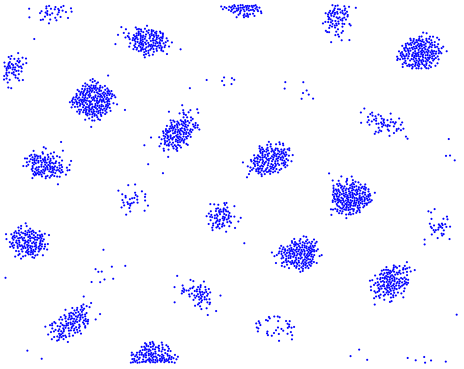
(a) Sector of a correlated medium with $A = 1$.



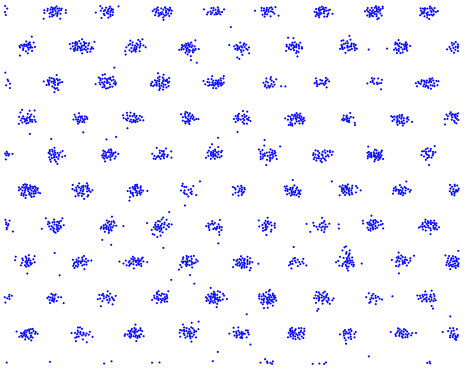
(b) Sector of a correlated medium with $A = 1$.



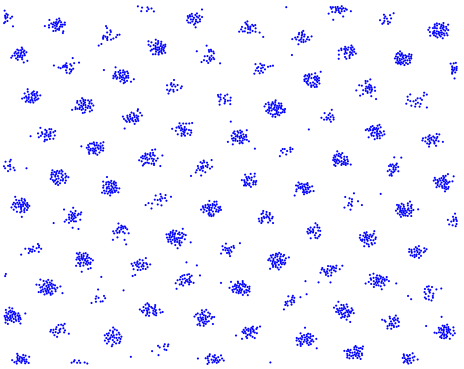
(c) Sector of a correlated medium with $A = 2$.



(d) Sector of a correlated medium with $A = 2$.



(e) Sector of a correlated medium with $A = 4$.



(f) Sector of a correlated medium with $A = 4$.

Figure 5.2: Correlated media for different radial distribution functions.

distribution function oscillates faster here. Distances between cluster centers are approximately 1 or 1/2, respectively, as expected.

5.1.4 Limiting path length distribution

We are now interested in the path length distribution in the Boltzmann-Grad limit. To be more precise, the quantity of interest here is the limit of the following cumulative distribution function

$$q(s) = \mathbb{P}(\text{“distance to collision } \geq s\text{”})$$

of the distance to the first collision s. As we have seen in the previous chapter, we can derive the probability density for the distance to the first collision, and also the density for the distance between collisions from this quantity. Note that analytic results for the convergence are only available for the periodic and the uncorrelated setting. But we expect that this limit also exists for more general correlated media, and that the limiting distribution becomes independent of space and direction as long as the medium is homogeneous. The radius of the spherical scatterers is of the order ε^2, hence we choose the number density of the medium to be $1/\varepsilon^2$. In the Boltzmann-Grad limit ε goes to 0.
Since we assume that our medium is infinite, we can also investigate this limiting behavior. Having a fixed setting of scatterers given by their center coordinates, we can make a series of path length calculations with successively smaller radii

$$\text{scatterer radius} = r, \frac{r}{2}, \frac{r}{4}, \ldots.$$

This series of distributions gives an idea of the convergence behavior. We can validate these calculations with a periodic setting, where an analytic result exists in dimension two [15], or with an uncorrelated medium, where we have

$$q(s) = \exp(-\sigma s)$$

with $\sigma = 2$ for the scatterer density that we consider [53].

5.2 Numerical results

For all numerical experiments we consider a box of size 10×10 that is periodically extended to infinity. That means the number density is $\bar{n} = N/100$ and the necessary number density $\bar{n}$ for the Boltzmann-Grad limit is $\bar{n} = 1/\varepsilon^2$. Hence, the number of scatterers that we choose for a given radius r are

$r = 0.1$	1000 scatterers,
$r = 0.05$	2000 scatterers,
$r = 0.02$	5000 scatterers,
$r = 0.01$	10000 scatterers,

and so on. We restrict our computations to 9 neighbouring boxes that form a square. As a consequence, our computations only yield reliable results for path lengths $0 \leq s \leq 10$, since this is the minimal distance of a scatterer to the boundary of the full domain. This is sufficient for our purposes.
First, we use the results for the uncorrelated and the periodic scatterer configuration as a verification and to test the parameters, especially the scatterer radius r. Then we show that the introduction of spatial correlations yields non-classical distributions of the free path lengths in the Boltzmann-Grad limit. Therefore, we study the path length distributions for the randomly perturbed periodic Lorentz gas, and for spatially correlated media created by the Metropolis Monte-Carlo method introduced in the previous section, with a sufficiently small scatterer radius, and averaged over several realizations.

Verification and parameter testing

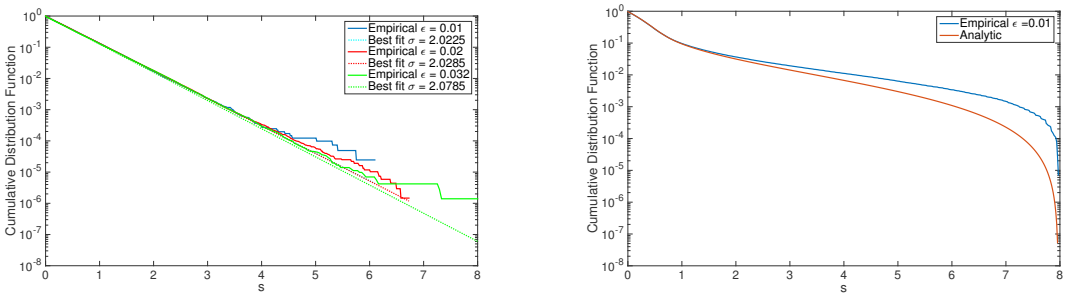
In the uncorrelated case we have $2\bar{n}r \to \sigma$ as $r \to 0$. Hence we expect that $\sigma \approx 2\bar{n}r = 2$ for $r, \bar{n}$ as above, and the cumulative distribution of free paths should converge to $\exp(-2s)$ in the limit. We simulate path lengths in the medium until the estimated statistical error in the second moment is below a predefined value. The statistical error in the second moment after the N-th calculated path length can be estimated [72] by

$$\text{statistical error in } \langle s^2 \rangle \approx \frac{\rho \sqrt{|\langle s^4 \rangle_N - \langle s^2 \rangle_N^2|}}{\langle s^2 \rangle_N \sqrt{N}},$$

where

$$\langle s^2 \rangle_N = \frac{1}{N} \sum_{i=1}^{N} s_i^2, \quad \langle s^4 \rangle_N = \frac{1}{N} \sum_{i=1}^{N} s_i^4,$$

and the additional parameter ρ determines the width of the confidence interval around the true value of $\langle s^2 \rangle$. In order to test the convergence of the cumulative distribution function in the uncorrelated case in the Boltzmann-Grad limit, we simulate path lengths in 30 realizations of the medium until the statistical error is below 0.03 for $\rho = 3$. We choose scatterer radii $r = 0.032, 0.02, 0.01$. In all cases the overall number of path lengths was around 700000. In Figure 5.3a we see that the convergence is qualitatively and quantitatively as expected. The resulting cumulative distribution functions are plotted together with a best fit exponential function, where we obtain the corresponding parameter σ by a non-linear regression. We see that $\sigma \to 2$ as r shrinks, and the change from $\varepsilon = 0.02$ to $\varepsilon = 0.01$ is small, while the relative error to the analytic value $\sigma = 2$ is below 2%. Hence, we conclude that a scatterer radius $r = 0.01$ and a number of 10000 scatterers are sufficient for our purposes.
In the periodic scatterer case we directly choose the scatterer radius $r = 0.01$. A statistical error of 0.02 and $\rho = 3$ were needed to obtain a relative error to the analytic solution (in the l^2 sense) of 2.5%, which results again in approximately 700000 simulated path lengths. The resulting distribution function



(a) Uncorrelated medium

(b) Periodic medium

Figure 5.3: Cumulative path length distribution for decreasing radii ε for the uncorrelated case (left) and the result for the periodic (right) case.

and the analytic solution are shown in Figure 5.3b. We note the very close correspondence for small path lengths and the clear non-exponential behavior of the distribution function.
In the following, we rely on the set of parameters that we have obtained by these calculations.

Random perturbations of the periodic Lorentz gas

We calculate path length distributions for increasingly strong perturbations of a periodic setting of scatterers as shown in Figure 5.1. The radius of the scatterers is fixed to $r = \varepsilon^2 = 0.01$, while the variance of the perturbations are chosen as $2\%, 5\%, 10\%, 15\%, 20\%$ and 100% of the distance between lattice points. For each variance we simulate path lengths in 20 realizations of the medium. The number of simulated paths varies from 204852 for the strongest perturbation up to 996841 in the 2% case. The resulting cumulative path length distributions are shown in Figure 5.4 together with the analytic results for the uncorrelated and the unperturbed periodic backround medium.
We observe a transformation between these two curve for increasingly strong perturbations. For a perturbation by 100% of the grid size, the result basically coincides with the analytic solution for the uncorrelated medium. On the other hand, the distribution functions for small perturbations are close to the analytic solution for the periodic setting, and we clearly see the non-exponential behavior. Furthermore, we note that the transformation is continuous as the strength of the perturbation grows. Hence, we conclude that there exists a continuum of random, but correlated, and homogeneous background media that have a non-exponential path length distribution in the Boltzmann-Grad limit, where the classical transport theory does not apply but the non-classical theory may be applicable.

Correlated media

Again, we we choose $r = \varepsilon^2 = 0.01$ with 10000 scatterers, and choose three ε-dependent sinusoidal radial distribution functions of the form (5.1), with parameters $A = 1$ (Testcase 1, as shown in the first row of Figure 5.2), $A = 2$

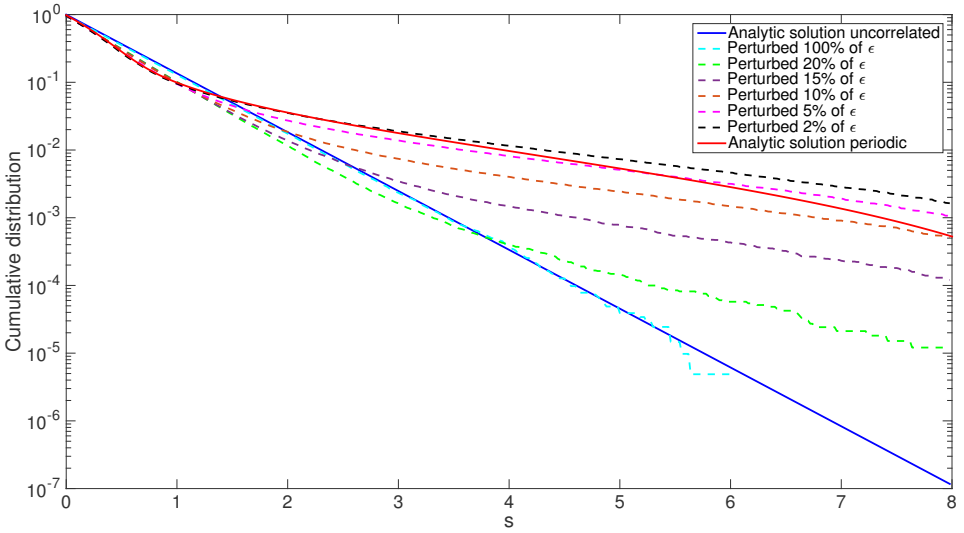
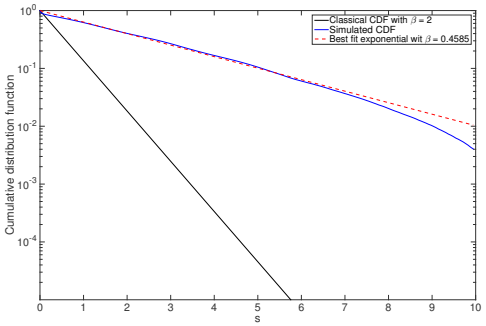


Figure 5.4: path length distributions for increasingly strong pertubations of a periodic background medium.

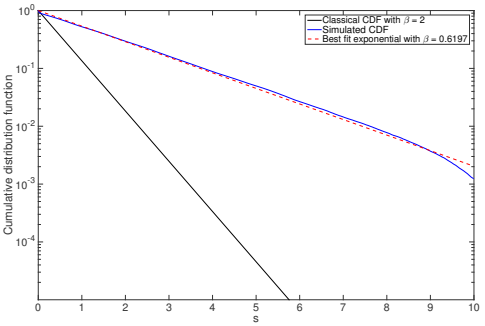
(Testcase 2, as shown in the second row of Figure 5.2), and $A = 4$ (Testcase 3, as shown in the third row of Figure 5.2). Then we simulate the path length distribution averaged over 10 realizations of the random medium and compare with the results for the classical, uncorrelated scatterer configuration. The media were created with approximately 15 million Metropolis moves each.
For the Monte-Carlo simulation of path lengths we choose $\rho = 3$, and run until the statistical error is below 0.03. Therefore, an overall of 353848 path length histories were needed for Testcase 1, 440222 path lengths for Testcase 2, and 888984 for Testcase 3. Figure 5.5 shows the simulated cumulative distribution function for Testcase 1 (Figure 5.5a), 2 (Figure 5.5b), and 3 (Figure 5.5c), together with the corresponding best fit exponential function and the analytic result for the uncorrelated medium. We see that the simulated distribution functions decrease nearly exponentially in the Testcases 1 and 2, but noticeably slower than the distribution in the uncorrelated case. The best fit exponential functions are of the form

$$q(s) = \exp(-\beta s),$$

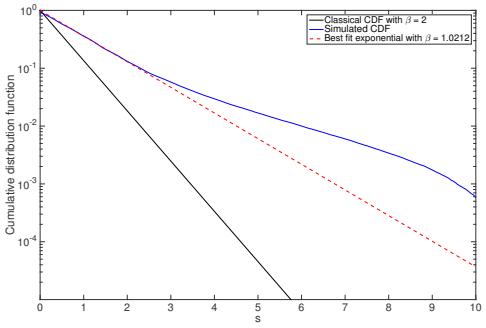
and we obtain parameters $\beta_1 = 0.4585$ for Testcase 1, and $\beta_2 = 0.6197$ for Testcase 2, in contrast to the analytic value $\beta = 2$ for the uncorrelated medium. For the Testcase 3 we observe exponential decay with parameter approximately $\beta_3 = 1.0212$ for small path length ($s < 3$). For larger path lengths the decrease of the cumulative distribution function becomes even slower, but is not strictly exponential anymore. Important to note is that in all three Testcases we observe path length distribution that deviate visibly from the result for the uncorrelated, and that the decrease is significantly slower. Hence, we conclude that we have found another class of correlated media with an underlying non-classical path length distribution. Although the decrease is partially nearly exponential, it can obviously not be described by the classical theory.



(a) CDF for $A = 1$.



(b) CDF for $A = 2$.



(c) CDF for $A = 4$.

Figure 5.5: Cumulative path length distributions for for different correlated media.

6 Non-classical transport equations

Motivated by the rigorous derivation in the periodic scatterer case, and the results on the analysis of more general correlated media in Section 5.2, we present several extended as well as simplified non-classical transport equations. First, we discuss extensions of the time-dependent non-classical equations, where we include additional features like absorption or anisotropic scattering. From these equations we derive the corresponding steady state equation, and discuss diffusion approximations and extensions thereof. The steady state equations is best suited for computer simulations. Therefore, we restrict the development of numerical methods to these equations in the following chapters.

6.1 Time-dependent equations

The time-dependent non-classical kinetic equation (4.11), which is the Fokker-Planck or forward Kolmogorov equation of the associated stochastic process in the homogenized background medium, is derived in Section 4.2 for a periodic medium only. But the discussion in Section 4.3, and our studies in Chapter 5 suggest, that we can state it also for more general media, again in 2 space dimensions, in the same form

$$\partial_t F(t,x,\Omega,s) + \Omega \nabla_x F(t,x,\Omega,s) - \partial_s F(t,x,\Omega,s) = \frac{p(s)}{2\pi} \int_{S^1} F(t,x,\Omega,0)\,\mathrm{d}\Omega,$$
$$F(0,x,\Omega,s) = f^{\mathrm{in}}(x,\Omega)p_0(s),$$

with general path length distributions $p_0(s)$ and $p(s)$, depending on the underlying (homogenized) background medium. The function F can be interpreted as

$$F(t,x,\Omega,s) = \text{"density of particles at time } t \text{ and position } x, \text{ travelling in direction } \Omega, \text{ that have a distance to the next collision } s\text{"}.$$

As the derivation in Section 4 has shown, it is connected to the classical particle density $f(t,x,\Omega)$ by

$$f(t,x,\Omega) = \int_0^\infty F(t,x,\Omega,s)\,\mathrm{d}s,$$

only that f does not satisfy a linear Boltzmann equation.
The left hand side of the equation can be interpreted as follows: with increas-

ing time, particles travel in direction Ω, while the distance to the next collision decreases (negative sign of the $s-$derivative). The right-hand side plays the role of a scattering source term. Particles scatter when their distance to collision is $s = 0$, they are scattered isotropically into every direction and they have a new distance to collision according to the path length distribution $p(s)$. The same holds for the initial distribution $F(0, x, \Omega, s)$. The classical distribution of particles in space and direction-of-flight is multiplied by $p_0(s)$ since the distances to collision is initially distributed by the same law. This interpretation gives a clear idea how this equation can be extended to external source terms and/or boundary sources. In both cases, we can choose a source term $Q(x, \Omega)$ and then multiply it by $p(s)$ to assign a natural distance to collision to the particles. Absorption can be modeled by introducing a scattering ratio $0 \leq c \leq 1$ that multiplies the scattering source term, as in the classical equation. We can also introduce a non-isotropic scattering kernel $\sigma(\Omega \cdot \Omega')$. The equation, now in N dimensions, then reads

$$\partial_t F(t, x, \Omega, s) + \Omega \nabla_x F(t, x, \Omega, s) - \partial_s F(t, x, \Omega, s) = \tag{6.1a}$$

$$p(s) \left(c \int_{S^{N-1}} \sigma(\Omega \cdot \Omega') F(t, x, \Omega, 0) \, \mathrm{d}\Omega + Q(x, \Omega) \right), \tag{6.1b}$$

$$F(0, x, \Omega, s) = f^{\mathrm{in}}(x, \Omega) p_0(s). \tag{6.1c}$$

Remark 6.1. This equation is obviously a generalization of the classical linear Boltzmann equation (2.1). By setting

$$p(s) = \sigma_t \exp(-\sigma_t s), \text{ and } \quad F(t, x, \Omega, s) := f(t, x, \Omega) \exp(-\sigma_t s),$$

we have that $f(t, x, \Omega)$ satisfies the classical transport equation (2.1), since the $s-$dependent terms cancel out. This is again evidence that the non-classical equation is valid for a wider class of background media.

Remark 6.2. The extension of this equation to a general scattering kernel formally requires, that the new distance to the next collision are independent of the new direction-of-flight. This is clearly not the case in a general stochastic medium, and it is an open question if it always holds when a statistically homogeneous medium is homogenized in the Boltzmann-Grad limit. We recall that the results of Chapter 5 point towards this being the case. Nevertheless, this is how the extra effects are usually modeled in the classical transport case (for example for non-homogeneous media). Potential errors are supposed to be small and are therefore neglected.

6.2 From time-dependent to steady state

In order to establish the connection to the non-classical transport equation introduced by Larsen in [73], we turn to the backward Kolmogorov equation

(4.10):

$$\begin{aligned}
&\partial_t g(t,x,\Omega,s) - \Omega\nabla_x g(t,x,\Omega,s) + \partial_s g(t,x,\Omega,s) = 0,\\
&g(t,x,\Omega,0) = \frac{1}{2\pi}\iint p(s) g(t,x,\Omega,s)\,\mathrm{d}\Omega\,\mathrm{d}s,\\
&g(0,x,\Omega,s) = \Phi(x,\Omega,s).
\end{aligned}$$

It describes the evolution of an observable of the transport process introduced in Section 4.2. Because of the different signs compared with the Fokker-Planck equation, this is usually interpreted to describe the same transport process, only in the backward direction $-\Omega$. In classical transport we can simply switch between the two equations by a transformation of the angular variable $\Omega \mapsto -\Omega$. We assume that the path length distribution $p(s)$ can be written as

$$p(s) = \Sigma_t(s)\exp\big(-\int_0^s \Sigma_t(\tau)\,\mathrm{d}\tau\big),$$

where $\Sigma_t(s)$ is, for now, a "smooth enough" function such that the following transformations are possible. We define $\Psi(t,x,\Omega,s)$ implicitly as

$$g(t,x,\Omega,s) =: \exp(\int_0^s \Sigma_t(\tau)\,\mathrm{d}\tau)\Psi(t,x,\Omega,s).$$

Hence, we obtain the following equation for Ψ:

$$\begin{aligned}
&\exp\big(-\int_0^s \Sigma_t(\tau)\,\mathrm{d}\tau\big)(\partial_t g - \Omega\nabla_x g + \partial_s g)\\
&\quad = \partial_t\Psi - \Omega\nabla_x\Psi + \partial_s\Psi + \Sigma_t(s)\Psi.
\end{aligned}$$

The initial condition for $s = 0$ becomes

$$\frac{1}{2\pi}\iint g(t,x,\Omega,s)p(s)\,\mathrm{d}\Omega\,\mathrm{d}s = \frac{1}{2\pi}\iint \Sigma_t(s)\Psi(t,x,\Omega,s)\,\mathrm{d}\Omega\,\mathrm{d}s. \tag{6.2}$$

If g is interpreted to describe transport in the backward direction, the left hand side describes the density of particles that undergo a collision at time t and position x, since $p(s)$ is the density of path length between collisions. Hence, the right-hand side has the same interpretation. Now inserting the quantity Ψ, it satisfies the following time-dependent non-classical transport equation

$$\begin{aligned}
&\partial_t\Psi - \Omega\nabla_x\Psi + \partial_s\Psi + \Sigma_t(s)\Psi = 0,\\
&\Psi(t,x,\Omega,0) = \frac{1}{2\pi}\iint \Sigma_t(s)\Psi(t,x,\Omega,s)\,\mathrm{d}\Omega\,\mathrm{d}s,
\end{aligned}$$

with a suitable initial condition. This equation is related to the steady-state equation introduced in [73]. Assuming that, for $t \to \infty$, the system reaches a

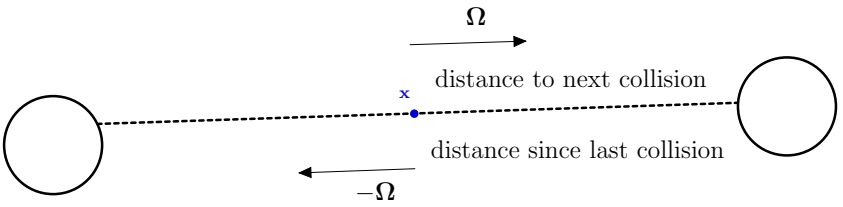


Figure 6.1: A particle that travels in direction Ω has a certain distance to the next collision. If we revert the direction of flight to $-\Omega$, this now becomes the distance since the last collision.

time-independent steady state $\Psi = \Psi(s, x, \Omega)$, it satisfies the equation

$$\partial_s \Psi(s, x, \Omega) - \Omega \nabla_x \Psi(s, x, \Omega) + \Sigma_t(s) \Psi(s, x, \Omega) = 0,$$
$$\Psi(0, x, \Omega) = \frac{1}{2\pi} \iint \Sigma_t(s) \Psi(s, x, \Omega) \, \mathrm{d}\Omega \, \mathrm{d}s.$$

In this equation, the variable s still has the interpretation as being the distance to the next collision. By reverting the direction of flight

$$\Omega \longmapsto -\Omega$$

this meaning is changed to

$$s = \text{ "distance SINCE last collison".}$$

This is illustrated in Figure (6.1). The change of the direction of flight is consistent with the usual interpretation of the Kolmogorow equation. That means, this now describes a process forward in time, only that it is formally independent of time as it is in steady state. This is possible because of the symmetry of the transport process in equilibrium. Quantities like particle density or the angular flux are invariant under this transformation.

Then we obtain the following equation, that has the same form as the non-classical transport equation in [73]:

$$\partial_s \Psi(s, x, \Omega) + \Omega \nabla_x \Psi(s, x, \Omega) + \Sigma_t(s) \Psi(s, x, \Omega) = 0, \tag{6.3a}$$
$$\Psi(0, x, \Omega) = \frac{1}{2\pi} \iint \Sigma_t(s) \Psi(s, x, \Omega) \, \mathrm{d}\Omega \, \mathrm{d}s. \tag{6.3b}$$

This equation is discussed in detail in the following section. That includes Larsen's physical interpretation of the quantities $\Sigma_t(s)$ and of Ψ.

Remark 6.3. We can also derive a steady state equation based on Equation (4.11). Assuming that the process is in equilibrium, we then obtain an equation

for $F = F(x, \Omega, s)$ as

$$-\partial_s F(x, \Omega, s) + \Omega \nabla_x F(x, \Omega, s) = \frac{p(s)}{2\pi} \int_{S^1} F(x, \Omega, 0) \, \mathrm{d}\Omega.$$

This equation needs to be equipped with a formal terminal condition $F(x, \Omega, \infty) = 0$, since there are no particles that have an infinite path length. Therefore, this equation has to be solved backwards in the pseudo-time variable s. An external source $p(s)Q(x, \Omega)$ can be introduced in the right-hand side as in the previous section. Nevertheless, in the remainder of this work we focus on Equation (6.3).

6.3 Steady state equation

When modeling radiation, the speed of the particles is usually large compared to the characteristic size of the medium, and the incoming radiation varies slowly in time. A good example is again the transport of photons in atmospheric clouds, where the speed of light needs to be compared to the size of a cloud, which is of the order of kilometres. Hence it is reasonable to view transport in these situations as a process in equilibrium. Independent of the rigorous derivation presented above, Larsen [73, 76] suggested the following non-classical transport equation to model the transfer of photons in atmospheric clouds:

$$\partial_s \Psi(s, x, \Omega) + \Omega \nabla_x \Psi(s, x, \Omega) + \Sigma_t(s) \Psi(s, x, \Omega) = 0 \tag{6.4a}$$

$$\Psi(0, x, \Omega) = c \int_0^\infty \int \sigma_s(\Omega, \Omega') \Sigma_t(s') \Psi(s', x, \Omega') \, \mathrm{d}\Omega' \, \mathrm{d}s' + Q(x, \Omega). \tag{6.4b}$$

This is a steady state transport equation whose connection to Equations (4.11) and (4.10) has been discussed in the previous section. The derivation in [73] relies on the idea to introduce a path length dependence to the total cross section. Then $\Sigma_t(s)$ is interpreted as a scattering rate, meaning that

$$\Sigma_t(s) \, \mathrm{d}s = \text{"probability, that a particle that has already traveled a distance } s \text{ will collide with path length } s_1 \in (s, s + \mathrm{d}s)\text{."}$$

Following the typical ad hoc derivation for the transport equation, Larsen obtains equation (6.4), where the unknown Ψ is interpreted as

$$\Psi(x, \Omega, s) = \text{"angular flux of particles at } x, \text{ travelling in direction } \Omega, \text{ with distance } s \text{ since its last collision"}.$$

The equation (6.4) carries two extensions compared to (6.3). First, an external source term $Q(x, \Omega)$ is introduced at $s = 0$. That means, we assume that all particles that enter the system have no (relevant) path length history. Second, the scattering ratio c multiplies the scattering source term. By doing so we also take into account absorption. Note that the scattering ratio does not depend

on s in this case since a collision is a purely local process. With other words, what happens with a particle in case of a collision is independent of the path length it has traveled before. Hence, the classical relation $\Sigma_t = \Sigma_s + \Sigma_a$ does not hold for this equation. Further extensions to spatial inhomogeneities and frequency dependence are discussed in Section 6.5.
This equation has a reduced complexity due to the absence of time as an independent variable. That makes it easier for a numerical treatment and further theoretical considerations. This is why we restrict ourselves to the study and analysis of this steady state equation in the remainder of this work. It was shown in [73] that in the case of a constant scattering frequency $\Sigma_t(s) \equiv \Sigma_t$, Equation (6.4) reduces to the classical steady-state Equation (2.3).

In contrast to the classical transport equation, this non-classical equation can reproduce any path length distribution $p(s)$ and is therefore suited for simulations of transport processes as described in Chapter 3. We shortly sketch the calculation of the path length distribution given in [73]. Again, assume one particle traveling through a purely absorbing medium. Starting at $x = 0$ into direction $\Omega = (1, 0, 0)^T$, the position x of the particle is equal to the actual length s of its path. Then, Equation (6.4) reduces to

$$\partial_s \Psi(s, x) + \partial_x \Psi(s, x) + \Sigma_t(s)\Psi(s, x) = 0.$$

In this situation we have $x(s) = s$, and therefore we define $F(s) := \Psi(s, x(s))$. Now $F(s)$ is the solution of the ordinary differential equation

$$\partial_s F(s) + \Sigma_t(s)F(s) = 0, \qquad F(0) = 1,$$

which is

$$F(s) = \exp(-\int_0^s \Sigma_t(\tau)\,\mathrm{d}\tau).$$

This gives the probability, that a particle travels a distance s without undergoing a collision. The corresponding probability density of the path length distribution is then obtained by taking the negative derivative of $F(s)$ (compare Chapter 3 and Appendix A). We then obtain

$$p(s) = -F'(s) = \Sigma_t(s)F(s) = \Sigma_t(s)\exp(-\int_0^s \Sigma_t(\tau)\,\mathrm{d}\tau),$$

the same function we have already introduced in the time-dependent case. This can be any probability density on $[0, \infty)$ depending on the choice of $\Sigma_t(s)$. Especially, it is exponential if and only if Σ_t is a constant. This calculation also shows, that $p(s)$ and $\Sigma_t(s)$ are related by

$$p(s) = \Sigma_t(s)\exp\left(-\int_0^s \Sigma_t(\tau)d\tau\right) \quad \text{if and only if} \quad \Sigma_t(s) = \frac{p(s)}{1 - \int_0^s p(\tau)d\tau}. \tag{6.5}$$

Remark 6.4. The non-classical cross section $\Sigma_t(s)$ arises naturally and is directly connected to the (homogenized) path length distribution. However, as in the classical setting, it can be interpreted as a collision rate: it is the rate by which particles that have traveled a certain distance s since their last interaction undergo collisions with the scatterers. From this point of view, the s variable is an unphysical memory variable. Therefore, we want to point out that $\Sigma_t(s)$ is inseparably linked to the distribution of path lengths between collisions $p(s)$, which is a classical physical quantity. It can be explicitly measured, and is therefore suitable for real world applications. Furthermore, the interpretation of $\Sigma_t(s)$ as a collision rate is consistent with the rigorous derivation of the non-classical transport equation given in this work. Since in the left-hand side of Equation (6.2) describes the density of particles that have experienced a collision, the quantity

$$\int_0^\infty \int \Sigma_t(s)\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}s$$

can be interpreted as the collision rate density, where $\Sigma_t(s)$ is the collision rate, and Ψ the angular flux.

A result on the existence, positivity and uniqueness of a solution of (6.4) was proven by Frank and Goudon [44]. It holds under the following mild assumptions:

(a) The source term $Q(x,\Omega) \geq 0$ satisfies $Q \in L^1(R^N, S^{N-1})$.

(b) The scattering kernel is positive $\sigma_s(\Omega,\Omega') \geq \bar{\sigma} > 0$ and normalized.

(c) The scattering ratio c satisfies $0 < c < 1$.

(d) The scattering frequency $\Sigma_t(s)$ is positive and satisfies

$$\int_0^\infty \Sigma_t(s)\,\mathrm{d}s = \infty.$$

We state this result here for reasons of completeness.

Theorem 6.1. *(Frank, Goudon)*
Under the assumptions above, there exists a unique non negative solution $\Psi(s,x,\Omega) \geq 0$ *of (6.4), satisfying* $\Psi \in L^\infty(0,\infty;L^1(\mathbb{R}^N \times S^{N-1}))$, $\Sigma_t(s)\Psi \in L^1((0,\infty)\times\mathbb{R}^N \times S^{N-1})$.

6.4 Diffusion approximations

A background medium may have optically thick regions where transport of particles is dominated by scattering, therefore we apply the following scaling. It relies on the assumption that the probability for absorption is small, and scattering is not forward-peaked, as it was introduced in [75]. A detailed analysis of the diffusion asymptotics can be found in [73, 76, 44]; we only give a sketch of it here.

We define the first and second moment of the path length distribution for the rest of this work as

$$\langle s\rangle := \int_0^\infty sp(s)\,\mathrm{d}s, \quad \langle s^2\rangle := \int_0^\infty s^2 p(s)\,\mathrm{d}s.$$

Now, let $0 < \varepsilon \ll 1$, then we scale the parameters Q, Σ_t and c as follows:

$$Q \to \varepsilon Q, \quad \Sigma_t(s) \to \frac{1}{\varepsilon}\Sigma_t(s/\varepsilon), \quad c \to 1 - \kappa\varepsilon^2,$$

where $\kappa > 0$. That means that the source is small, and the number of collisions is large, while the probability of absorption is small. By additionally scaling the pseudo time-scale by

$$\phi_\varepsilon(s, x, \Omega) := \phi(s\varepsilon, x, \Omega),$$

the new scaled unknown satisfies

$$\begin{aligned}
&\partial_s\phi_\varepsilon(s, x, \Omega) + \varepsilon\Omega\nabla_x\phi_\varepsilon(s, x, \Omega) + \Sigma_t(s)\phi_\varepsilon(s, x, \Omega) = 0;\\
&\phi_\varepsilon(0, x, \Omega) = (1 - \kappa\varepsilon^2)\int_0^\infty\int \sigma(\Omega, \Omega')\Sigma_t(s')\phi_\varepsilon(s', x, v')\,\mathrm{d}\Omega'\,\mathrm{d}s' + \varepsilon Q(x, \Omega).
\end{aligned}$$

This equation can be further simplified. By setting

$$\Phi_\varepsilon = \varepsilon\langle s\rangle \exp\left(\int_0^s \Sigma_t(\tau)d\tau\right)\phi_\varepsilon(s, x, \Omega)$$

we make use of (6.5), and the equation now reads

$$\partial_s\Phi_\varepsilon + \varepsilon\Omega\nabla_x\Phi_\varepsilon = 0; \tag{6.6a}$$

$$\Phi_\varepsilon(0, x, \Omega) = (1 - \kappa\varepsilon^2)\int_0^\infty\int \sigma(\Omega, \Omega')p(s)\Phi_\varepsilon(s', x, \Omega')\,\mathrm{d}s'\,\mathrm{d}\Omega' + \varepsilon^2\langle s\rangle Q(x, \Omega). \tag{6.6b}$$

Applying a formal Hilbert expansion in the parameter ε by

$$\Phi_\varepsilon = \Phi_0 + \varepsilon\Phi_1 + \varepsilon^2\Phi_2 + \cdots$$

and letting $\varepsilon \to 0$, we derive the following diffusion approximation (see again [73, 76] or [44] for the details):

$$-\frac{1}{3}\left(\frac{\langle s^2\rangle}{2} + \frac{\bar\mu_0}{1 - \bar\mu_0}\right)\nabla_x^2\Phi_0(x) + (1 - c)\Phi_0(x) = \langle s\rangle Q(x), \tag{6.7}$$

where $\bar\mu_0$ is the mean scattering cosine. If we consider an isotropic medium with constant probability for the outgoing angle, we simply have $\bar\mu_0 = 0$. All combined, Φ_0 is an approximation of the classical angular flux up to first order,

because of

$$\int_0^\infty \int \phi(s, x, \Omega) \, \mathrm{d}\Omega \, \mathrm{d}s = \Phi_0(x) + O(\varepsilon).$$

Note that this diffusion limit is not valid any more as soon as

$$p(s) \geq \frac{C}{s^3},$$

for constant $C > 0$, because in this case $\langle s^2 \rangle = \infty$. This case is discussed in the following remark.

Remark 6.5. The theory of anomalous transport is concerned with transport processes where, in general, the second moment of the path length distribution is infinite [63]. The resulting macroscopic equations that describe the transport in the diffusive regime are fractional diffusion equations. This strongly suggests that the diffusion limit for the non-classical equation in the case of path length distributions with infinite second moment is given by a fractional diffusion version of Equation (6.7). In an upcoming paper [43], Frank shows that, if the path length distribution has an algebraic tail and an infinite second moment, meaning $p(s) \sim C/s^{\alpha+1}$ for some constant $C > 0$, $s \gg 1$, and $1 < \alpha < 2$, the fractional diffusion limit of the non-classical transport equation is given by

$$-D(\alpha)\Delta^{\alpha/2}\Phi_0(x) + (1-c)\Phi_0(x) = \langle s \rangle Q(x),$$

where $D(\alpha)$ is a diffusion constant that depends on the exponent α, and $\Delta^{\alpha/2}$ is the so-called fractional Riesz-Feller derivative [63], usually defined by its representation in Fourier space

$$\Delta^{\alpha/2} u(x) = -|k|^\alpha \hat{u}(k),$$

where $\hat{u}$ is the Fourier transform of u. The operator $\Delta^{\alpha/2}$ can be interpreted as a generalization of the Laplace operator to orders less than 2.

Remark 6.6. Working on the application of the non-classical transport equation to neutron transport in pebble bed reactors, Larsen and Vasques [103, 76] extended Equation (6.4) to an angular dependent scattering frequency $\Sigma_t = \Sigma_t(s, \Omega)$. This leads to another non-classical diffusion approximation with anisotropic diffusion constants.

6.5 Frequency dependence and spatial inhomogeneities

A real world background medium, like an atmospheric cloud, is in general spatially inhomogeneous. In the classical radiative transfer equation this is usually expressed by a spatially dependent total cross section $\sigma_t(x)$ that describes the optical thickness locally around x. To take into account spatial inhomogeneities in the non-classical model, we introduce a space-dependent

quantity that is proportional to the classical cross section

$$\rho(x) \sim \sigma_t(x).$$

The total cross section is then the product of $\Sigma_t(s)$ and $\rho(x)$ since the optical thickness is independent of the path length a particle has travelled and therefore, $\Sigma_t(s)$ is a global parameter. Then the equation reads

$$\begin{aligned}
&\partial_s \Psi(s,x,\Omega) + \Omega \nabla_x \Psi(s,x,\Omega) + \Sigma_t(s)\rho(x)\Psi(s,x,\Omega) = 0,\\
&\Psi(0,x,\Omega) = c\rho(x) \int_0^\infty \int \sigma(\Omega,\Omega')\Sigma_t(s')\Psi(s',x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s'.
\end{aligned}$$

Especially, this yields that outside the medium we have $\rho(x) = 0$, hence the particles there are freely transported without undergoing further interactions with a background medium. In the same fashion frequency is included in the classical equation by introducing a frequency dependent total cross section $\sigma_t(x,\nu)$ and scattering kernel $\sigma_s(\Omega,\Omega',\nu,\nu')$. By the same argument this is written as

$$\sigma_t(x,\nu) = \gamma(\nu)\sigma_t(x), \quad \sigma_s(\Omega,\Omega',\nu,\nu') = \sigma(\Omega,\Omega')\gamma_s(\nu,\nu'),$$

with a frequency cross section $\gamma(\nu)$ that is the rate by which particles with frequency ν undergo a collision, and a frequency scattering kernel $\gamma_s(\nu,\nu')$ that gives the probability, that after a scattering event, a particle changes its frequency from ν' to ν. The inhomogeneous frequency-dependent non-classical equation then reads

$$\begin{aligned}
&\partial_s \Psi(s,x,\Omega,\nu) + \Omega \nabla_x \Psi(s,x,\Omega,\nu) + \Sigma_t(s)\rho(x)\gamma(\nu)\Psi(s,x,\Omega,\nu) = 0,\\
&\Psi(0,x,v,\nu) = c\rho(x) \int_0^\infty \int_0^\infty \int \sigma_s(\Omega,\Omega')\gamma_s(\nu',\nu)\Sigma_t(s')\Psi(s',x,\Omega',\nu')\,\mathrm{d}\Omega'\,\mathrm{d}\nu'\,\mathrm{d}s'.
\end{aligned}$$

Note again that this can only be an approximation, and it depends on the assumptions that these effects are independent.

Part II

Numerical Methods

Introduction: Numerical Methods

In this part we consider the problem of numerically solving the non-classical steady state transport equation. Solving the full equation is numerically costly due to its high dimensionality. Therefore we use different strategies to reduce the complexity of the full kinetic equation. The spherical harmonics or P_N method is a spectral Galerkin method for the angular variable, where the angular dependence is expanded in Legendre polynomials (in higher dimensions spherical harmonics). It can also be interpreted as a moment closure like the M_1-method. There, an overdetermined system for the first three angular moments is closed using an entropy minimization principle. In the S_N or discrete ordinates method the angular variable is directly discretized, typically at Gauss points.
Since a background medium like an atmospheric cloud may have optically thick regions where transport is dominated by scattering, we are especially interested in the performance of numerical schemes in this regime. For the system of angular moments we present a family of asymptotic preserving HLL schemes. The corresponding chapters of this part are based on the paper [71] on asymptotic preserving numerical schemes for the non-classical radiative transfer equation. Coupling of the initial value with respect of the pseudo-time and the full equation requires an iterative solution method. We show that source iteration becomes arbitrarily slow in the diffusive regime and propose initialization and acceleration schemes to overcome this problem.

For reasons of clarity and simplicity we present all numerical methods in 1D slab geometry form. Slab geometry means that we assume that the solution is rotationally invariant around the z-axis. Then the solution only varies in z-direction and the problem can be formulated as a pseudo 1D problem. The Equation (6.4) in slab geometry is

$$\partial_s \Psi(s,x,\mu) + \mu \partial_x \Psi(s,x,\mu) + \Sigma_t(s)\Psi(s,x,\mu) = 0 \tag{6.8a}$$

$$\Psi(0,x,\mu) = c\int_0^\infty \int_{-1}^1 \sigma(\mu)\Sigma_t(s')\Psi(s',x,\mu)\,\mathrm{d}\mu\,\mathrm{d}s' + Q(x,\mu), \tag{6.8b}$$

with $\mu \in [-1,1]$ being the cosine of the polar angle and $x \in (-\infty,\infty)$. As indicated in the previous chapter, in the diffusive scaling there is a small parameter ε which is basically the mean free path and the equation can be written as

$$\partial_s \Phi_\varepsilon(s,x,\mu) + \varepsilon\mu\nabla_x \Phi_\varepsilon(s,x,\mu) = 0; \tag{6.9a}$$

$$\Phi_\varepsilon(0,x,\mu) = (1-\kappa\varepsilon^2)\int_0^\infty \int_{-1}^1 \sigma(\mu,\mu')p(s)\Phi_\varepsilon(s',x,\mu')\,\mathrm{d}\mu'\,\mathrm{d}s' + \varepsilon^2\langle s\rangle Q(x,\mu), \tag{6.9b}$$

where

$$\Phi_\varepsilon = \varepsilon\langle s\rangle \exp\left(\int_0^s \Sigma_t(\tau)d\tau\right)\Psi_\varepsilon(s,x,v).$$

We only consider this form of the equation in this part.

7 Approximate models

The high dimensionality of the full linear Boltzmann equation makes it necessary to introduce approximate models. Hence, we derive systems of partial differential equations in which the angular dependence is eliminated either by a direct discretization, or by integration. The easiest approach is the so-called discrete ordinate method. Here, the angular variable is directly discretized at Gauss points on the unit sphere. We use this approximate method mainly in the next part for the parameter estimation. In this part on numerical methods, we are concerned with moment models of the non-classical steady state transport equation. Those yield an underdetermined system of equations. Hence, a closure for the system is needed. The P_N or spherical harmonics method is a spectra Galerkin method where the angular dependence of the flux is expanded in terms of the so-called spherical harmonics. This method can also be interpreted as a moment closure where the system is closed by truncating after the $N-$th moment. The minimum entropy system M_1 is derived by looking for a closure that minimizes a physical entropy while maintaining the moments up to order 1. Note that all of the resulting coupled systems of partial differential equations are hyperbolic.

7.1 Discrete ordinate method (S_N)

For the discrete ordinate approximation, the angular variable is discretized at N points $\mu_i,\ i = 1, .., N$ such that

$$\sum_{i=1}^{N} \omega_i \Psi(s, x, \mu_i) \approx \int_{S^2} \Psi(s, x, \Omega)\, \mathrm{d}\Omega,$$

where μ_i are usually chosen to be Gauss points with corresponding weights ω_i. This is called the S_N method. Then, in slab geometry, assuming isotropic scattering, we obtain the following system of equations:

$$\begin{aligned}
&\partial_s \Psi_i(s, x) + \varepsilon \mu_i \partial_x \Psi_i(s, x) = 0 \\
&\Psi_i(0, x) = \frac{(1 - \kappa\varepsilon^2)}{2} \int_0^\infty \sum_{i=1}^{N} \omega_i \Psi_i(s', x) p(s')\, \mathrm{d}s' + \varepsilon \langle s \rangle Q_i(x),
\end{aligned}$$

where $\Psi_i(s, x) = \Psi(s, x, \mu_i)$ and $Q_i(x) = Q(x, \mu_i)$. In slab geometry it can be shown that the S_N-method is equivalent to the P_{N-1} method (see below), if the angular variable is discretized at Gauss-points.

7.2 Moment models

The philosophy of moment methods is to derive a system of coupled equations for the angular moments of the angular flux Ψ. Multiplying the non-classical transport equation by polynomials in μ and then integrating out the angular variable, we obtain an underdetermined system of coupled equations. Therefore, a closure relation is needed.
We consider two different moment methods in this work. The linear spherical harmonics or P_N-method is a spectral Galerkin method in the angular variable, and relies on an expansion of the angular variable in terms of Legendre polynomials. The minimum entropy or M_1-method is based on the minimum entropy principle, and yields a non-linear coupled system of equations.

7.2.1 Spherical harmonics (P_N)

Let $P_l(\mu)$ be the $l-$th Legendre polynomial. The linear P_N-model is based on a Legendre polynomial expansion of Ψ as

$$\Psi(s,x,\mu) = \sum_{l=0}^{\infty} P_l(\mu)\alpha^{(l)}(s,x). \tag{7.1}$$

Testing with the Legendre polynomials and then integrating, we find that the coefficients of this expansion are the angular moments up to a factor:

$$\alpha^{(l)}(s,x) = \frac{2l+1}{2}\Psi^{(l)}(s,x),$$

where we define

$$\Psi^{(l)}(s,x) := \int_{-1}^{1} P_l(\mu)\Psi(s,x,\mu)\,\mathrm{d}\mu.$$

Testing Equation (6.9) with Legendre polynomials and integrating yields the corresponding P_N system. Using the recursion relation for the Legendre polynomials

$$\mu P_l(\mu) = \frac{l}{2l+1}P_{l-1}(\mu) + \frac{l+1}{2l+1}P_{l+1}(\mu),$$

and assuming isotropic scattering, it reads

$$\begin{aligned}
&\partial_s\Psi^{(l)}(s,x) + \varepsilon\partial_x\left[\frac{l}{2l+1}\Psi^{(l-1)}(s,x) + \frac{l+1}{2l+1}\Psi^{(l+1)}(s,x)\right] = 0, \quad l = 0,\dots,N,\\
&\Psi^{(0)}(0,x) = (1-\kappa\varepsilon^2)\int_0^{\infty} p(s)\Psi^{(0)}(s,x)\,\mathrm{d}s + \varepsilon^2\langle s\rangle Q(x),\\
&\Psi^{(l)}(0,x) = 0, \qquad l = 1,\dots,N,
\end{aligned}$$

where we set $\Psi^{(-1)} = 0$, and truncation of the expansion (7.1) yields $\Psi^{(N+1)} = 0$.

In the P_1 case we use the notation $E(s,x) := \Psi^{(0)}(s,x)$ and $F(s,x) := \Psi^{(1)}(s,x)$. Then the system of equations is given by

$$\partial_s E(s,x) + \varepsilon \nabla_x F(s,x) = 0, \qquad \partial_s F(s,x) + \frac{\varepsilon}{3} \nabla_x E(s,x) = 0, \tag{7.2a}$$

$$E(0,x) = \varepsilon^2 \langle s \rangle Q(x) + (1 - \kappa\varepsilon^2) \int_0^\infty p(s) E(s,x) ds, \quad F(0,x) = 0. \tag{7.2b}$$

In the following we develop and analyze numerical schemes for the P_1 system. These schemes can be naturally extended to solve the P_N systems for arbitrary $N > 1$. This is done in Chapter 14 for the unscaled Equation (6.8).

7.2.2 Minimum entropy (M_1)

The second moment method we use is the M_1-model based on the minimum entropy principle. This is the first representative of a hierarchy of models, called M_N. Only for the M_1 case a closed analytic form of the moment closure is available. The minimum entropy principle is, that for the moment closure a functional of the moments is chosen, which minimizes the radiative entropy and reproduces the lower order moments. The philosophical foundations for this approach were introduced in the 1950's in [61], while the first application in transport theory was made 20 years later in [87]. A first application to the equation of radiative transfer can be found in [35].
We introduce the first three angular moments of the flux Ψ in slab geometry:

$$E(s,x) := \int_{-1}^{1} \Psi(s,x,\mu) \, \mathrm{d}\mu, \qquad F(s,x) := \int_{-1}^{1} \mu \Psi(s,x,\mu) \, \mathrm{d}\mu,$$
$$P(s,x) := \int_{-1}^{1} \mu^2 \Psi(s,x,\mu) \, \mathrm{d}\mu.$$

Then we multiply Equation (6.9) by 1 and μ, respectively, and integrate over $(-1,1)$. We obtain, assuming isotropic scattering, the following coupled system of linear partial differential equations:

$$\partial_s E(s,x) + \varepsilon \nabla_x F(s,x) = 0, \qquad \partial_s F(s,x) + \varepsilon \nabla_x P(s,x) = 0,$$
$$E(0,x) = (1 - \kappa\varepsilon^2) \int_0^\infty p(s) E(s,x) \, \mathrm{d}s + \varepsilon^2 \langle s \rangle Q(x), \quad F(0,x) = 0.$$

Note that we have only two equations for three unknown functions. We set $f := \|\frac{F}{E}\|_2$. The M_1 closure leads to a coupled system of non-linear partial differential equations with

$$P = \chi(f) \cdot E,$$

where the Eddington factor in this case is given by

$$\chi(f) := \frac{3+4f^2}{5+2\sqrt{4-3f^2}}.$$

Hence, the system becomes

$$\partial_s E(s,x) + \varepsilon \nabla_x F(s,x) = 0, \qquad \partial_s F(s,x) + \varepsilon \nabla_x \chi(F/E) E(s,x) = 0, \quad (7.3a)$$

$$E(0,x) = (1-\kappa\varepsilon^2) \int_0^\infty p(s) E(s,x)\,\mathrm{d}s + \varepsilon^2 \langle s \rangle Q(x), \quad F(0,x) = 0, \qquad (7.3b)$$

with strong nonlinearity in the second equation.

8 Asymptotic preserving HLL schemes

In this chapter, we derive numerical schemes for the moment models introduced above, and analyze important properties of these schemes. A background medium, an atmospheric cloud for example, may have optically thick as well as optically thin regions. Numerical schemes can only yield reliable results if they cover the correct diffusion limit in the optically thick regime. Therefore, it is necessary to derive schemes which are correct both in the optically thick, diffusive regime and in the transition regime. This is called the asymptotic preserving property. It is well known that numerical problems arise in the diffusive regime [74, 1] where length scales of different magnitudes come into play. Furthermore, the standard HLL approach to construct finite volume methods results in numerical schemes which are typically not asymptotic preserving [13]. When calculating numerical solutions of moment models, we also face the additional constraint, that we need to be able to guarantee that the solutions are actual moments of an underlying distribution function on the full phase space. And the moments have to be in the domain of definition of the particular moment closure. The first property is called realizability, the second admissibility. For the first two angular moments there exists a simple condition that has to be satisfied.

Therefore, we present an adapted finite volume scheme for moment models of the non-classical steady state transport equation, and show that it is asymptotic preserving, and satisfies the admissibility and realizability conditions. We first derive a standard HLL scheme in Section 8.1, which is then adapted by a modification of the numerical diffusion term. In Section 8.2 we then investigate the asymptotic preserving, realizability and admissibility property for both the explicit and implicit pseudo time discretization. Furthermore, we investigate stability of the schemes for the linear P_1 case.

8.1 HLL schemes for moment models

We derive numerical schemes for the moment models of (6.4), which are introduced in the previous section. First, follow the formalism of Harten, Lax, and van Leer [57] to derive finite volume schemes for the P_1 and M_1 model. In the presentation of the derivation of the schemes we rely on the brief outline in [13], Section 3.1. A light modification of these schemes then yields numerical schemes with the widely admired asymptotic preserving property.

We choose a uniform time-space grid (s_n, x_i) with $s_{n+1} := s_n + \Delta s$ and $x_{i+1} := x_i + \Delta x$ for $n \in \mathbb{N}_0$ and $i \in \mathbb{Z}$. We interpret x_i as the center of the cell $C_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, where for the cell interfaces we have $x_{i+\frac{1}{2}} = x_{i-\frac{1}{2}} + \Delta x$. Then we use the abbreviations $Q_i := Q(x_i)$, $E_i^n := E(s_n, x_i)$ and $P_i^n := \chi(E_i^n, F_i^n)E_i^n$. Following [57] and [13], we introduce the notation

$$U = \begin{pmatrix} E \\ F \end{pmatrix}, \quad G(U) = \begin{pmatrix} F \\ \chi(E,F)E \end{pmatrix}, \quad U_i^n = \begin{pmatrix} E_i^n \\ F_i^n \end{pmatrix},$$

and write the Systems (7.2) and (7.3) as

$$\partial_s U + \varepsilon \partial_x G(U) = 0.$$

The discretization of the initial value is treated later and separately. Assume that we know the numerical solution U_i^n for all cells $i \in \mathbb{Z}$ at pseudo-time step n. Then, the values in the next pseudo-time step are computed as

$$U_i^{n+1} = U_i^n - \frac{\Delta s}{\Delta x}\big(\mathcal{F}_{i+\frac{1}{2}} - \mathcal{F}_{i-\frac{1}{2}}\big), \tag{8.1}$$

where the numerical fluxes $\mathcal{F}_{i\pm\frac{1}{2}}$ are given by

$$\mathcal{F}_{i+\frac{1}{2}} = G(U_i^n) - \frac{\Delta x}{2\Delta s} U_{i+1}^n + \frac{1}{\Delta s}\int_{x_{i+\frac{1}{2}}}^{x_{i+1}} U^h(x, s+\Delta s)\,\mathrm{d}x,$$

and $\mathcal{F}_{i-\frac{1}{2}}$ with the appropriate different indices. The function $U^h(x,s)$ is an approximation to the solution that is constructed in the following way. First note that at pseudo time s^n we have a piecewise constant representation of the solution:

$$U^h(x, s^n) = U_i^n \quad \text{for } x \in (x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}).$$

Then, at each cell interface $x_{i+\frac{1}{2}}$ we have a Riemann problem with left state U_i^n and right state U_{i+1}^n. We consider the approximate Riemann solver

$$U_{\mathcal{R}}\left(\frac{x}{s}; U_i^n, U_{i+1}^n\right) = \begin{cases} U_i^n & \text{for } \frac{x}{s} < b^- \\ U^*(U_i^n, U_{i+1}^n) & \text{for } b^- < \frac{x}{s} < b^+ \\ U_{i+1}^n & \text{for } b^+ < \frac{x}{s} \end{cases}.$$

For appropriate wave velocities b^-, b^+, the intermediate state U^* is given by

$$U^*(U_i^n, U_{i+1}^n) = \frac{b^+ U_{i+1}^n - b^- U_i^n}{b^+ - b^-} - \frac{1}{b^+ - b^-}\big(G(U_{i+1}^n) - G(U_i^n)\big).$$

An approximate solution for all $s \in (s^n, s^{n+1})$ can then be defined as

$$U^h(x,s) = U_{\mathcal{R}}\left(\frac{x - x_{i+\frac{1}{2}}}{s}; U_i^n, U_{i+1}^n\right) \quad \text{for } x \in (x_i, x_{i+1}).$$

We choose $b^+ = -b^- = \varepsilon$ since 1 is an upper bound for the absolute values of the eigenvalues of the flux both in the M_1 case (see [35]) and in the P_1 case (the eigenvalues can be explicitly calculated as $\pm 1/\sqrt{3}$ in this case). Hence, we can write the numerical fluxes as

$$\mathcal{F}_{i+\frac{1}{2}} = \varepsilon \frac{G(U_i^n) + G(U_{i+1}^n)}{2} - \varepsilon \frac{U_{i+1}^n - U_i^n}{2},$$

with corresponding expression for $\mathcal{F}_{i-\frac{1}{2}}$. Inserting this into Equation (8.1), we obtain the pseudo-time update

$$U_i^{n+1} = U_i^n - \frac{\Delta s}{\Delta x}\left(\varepsilon \frac{G(U_{i+1}^n) - G(U_{i-1}^n)}{2} - \varepsilon \frac{U_{i+1}^n - 2U_i^n + U_{i-1}^n}{2}\right).$$

Note that the first fraction in the capital brackets is a centered finite difference approximation of $\partial_x G(U)$, while the second fraction is a typical numerical diffusion term. We now modify this scheme in the following way: the numerical diffusion term is multiplied by another factor ε. The scheme is still consistent, it only reduces the amount of numerical diffusion that is introduced. This modification is necessary in order to obtain an asymptotic preserving scheme (see next section).

Now we can state the final discretization scheme, where also the initial value is discretized:

$$\frac{E_i^{n+1} - E_i^n}{\Delta s} + \varepsilon \frac{F_{i+1}^n - F_{i-1}^n}{2\Delta x} - \varepsilon^2 \frac{E_{i+1}^n - 2E_i^n + E_{i-1}^n}{2\Delta x} = 0, \tag{8.2a}$$

$$\frac{F_i^{n+1} - F_i^n}{\Delta s} + \varepsilon \frac{P_{i+1}^n - P_{i-1}^n}{2\Delta x} - \varepsilon^2 \frac{F_{i+1}^n - 2F_i^n + F_{i-1}^n}{2\Delta x} = 0, \tag{8.2b}$$

$$E_i^0 = (1 - \kappa\varepsilon^2) \sum_{n=0}^{\infty} \omega_n p_n E_i^n \Delta s + \varepsilon^2 \langle s \rangle Q_i, \qquad F_i^0 = 0. \tag{8.2c}$$

This is the scheme in case of an explicit pseudo-time discretization and for some infinite quadrature rule given by the weights ω_n. In practice the integration is cut off at some $s_{\max}$. This can be justified by the fact that p is a fast decaying function, hence, the contribution to the integral is small for $s > s_{\max}$. In Subsection 8.2.4 we also consider this scheme with an implicit pseudo-time discretization.

Remark 8.1. It is important here to note that HLL schemes are typically not asymptotic preserving. For the numerical solution of hyperbolic systems with source terms it was shown [13] that a modification of the HLL scheme makes it asymptotic preserving. Therefore, a free parameter is introduced into the source term, which is then used to correct the expected diffusion equation at the discrete level. The modification that we apply here is of a different kind. By reducing the amount of numerical diffusion, stability of the scheme is not guaranteed anymore. Therefore, this also needs to be investigated.

8.2 Asymptotic preserving property

A numerical scheme is called **asymptotic preserving (AP)**, if for each fixed discretization the leading order of the numerical solution is governed by a discretization of the correct analytic diffusion limit as ε tends to 0. In order to show this property, we apply a Hilbert expansion argument. This is a systematic formal procedure to identify one solution in the ring of formal power series in ε with coefficients in a suitable function space. We do not give a rigorous proof for uniform convergence in ε here. When strong nonlinearities are concerned, more rigorous proofs are out of reach. Therefore, most related works use a similar formal ansatz [51, 18, 11]. Rigorous results exist only for linear [54] and quasi-linear kinetic equations [3].

8.2.1 AP property of the moment models

We first show that both the M_1 and the first representative of the P_N models have the relevant asymptotic limit when $\varepsilon \to 0$. To do so, we perform a Hilbert expansion of E and F:

$$E = E_0 + \varepsilon E_1 + \varepsilon^2 E_2 + \dots,$$
$$F = F_0 + \varepsilon F_1 + \varepsilon^2 F_2 + \dots$$

and we define

$$\Pi_i(x) = \int_0^\infty p(s) E_i(s,x) ds.$$

Inserting these expansions into the System (7.3) and identifying the powers of ε, we get :

$$\underline{\varepsilon^0 \text{ terms:}} \qquad E_0 = \Pi_0 = \int_0^\infty p(s) E_0(s,x) ds, \quad F_0 = 0.$$

Since $F_0 \equiv 0$, we note that F is of order ε and so is $f = \frac{F}{E}$ (assuming $E \neq 0$ everywhere). Then we have in the M_1 case

$$\chi(f) = \frac{1}{3} + \frac{1}{2} f^2 + \frac{3}{32} f^4 + \mathrm{O}(f^6)$$

near $f = 0$. Consequently, both in the P1- and the M1-case,

$$\chi(F/E)E = \frac{1}{3} E + \mathrm{O}(\varepsilon). \tag{8.3}$$

Hence, we can continue by

$$\underline{\varepsilon^1 \text{ terms:}} \qquad E_1 = \Pi_1, \quad F_1 = -\frac{s}{3} \partial_x \Pi_0.$$

Inserting this into the next set of equations yields:

$$\underline{\varepsilon^2 \text{ terms:}} \qquad E_2 = \frac{s^2}{2}\frac{1}{3}\partial_x^2\Pi_0,$$
$$-\frac{1}{3}\frac{\langle s^2\rangle}{2}\partial_x^2\Pi_0 + \kappa\Pi_0 = \langle s\rangle Q.$$

This last equation coincides with the diffusion approximation (6.7) if scattering is isotropic, i.e. when the mean scattering cosine $\mu_0 = 0$.

8.2.2 AP property of explicit schemes

In order to simplify calculations, we write the scheme in the following as:

$$E_i^{n+1} = E_i^n - \varepsilon\mu(F_{i+1}^n - F_{i-1}^n) + \varepsilon^2\mu(E_{i+1}^n - 2E_i^n + E_{i-1}^n), \tag{8.4a}$$
$$F_i^{n+1} = F_i^n - \varepsilon\mu(P_{i+1}^n - P_{i-1}^n) + \varepsilon^2\mu(F_{i+1}^n - 2F_i^n + F_{i-1}^n), \tag{8.4b}$$
$$E_i^0 = (1-\kappa\varepsilon^2)\sum_{n=0}^{\infty}\omega_n p_n E_i^n + \varepsilon^2\langle s\rangle Q_i,\ F_i^0 = 0, \tag{8.4c}$$

where $\mu = \frac{\Delta s}{2\Delta x}$. The problem we address in this section is the behavior of the numerical scheme in the diffusive limit. The following proposition holds:

Proposition 8.1. *Assume the quadrature satisfies the following relation:*

$$\sum_{n=0}^{\infty}\omega_n p_n = 1. \tag{8.5}$$

Then, the numerical scheme (8.4) is asymptotic-preserving, i.e. it degenerates into a scheme consistent with (6.7) when $\varepsilon \to 0$.
Furthermore, the limit scheme is second-order accurate if the following condition holds:

$$\sum_{n=0}^{\infty} n\omega_n p_n = 0. \tag{8.6}$$

Proof. Mimicking the continuous case, we perform a Hilbert expansion of E_i^n and F_i^n:

$$E_i^n = E_{0,i}^n + \varepsilon E_{1,i}^n + \varepsilon^2 E_{2,i}^n + \dots,$$
$$F_i^n = F_{0,i}^n + \varepsilon F_{1,i}^n + \varepsilon^2 F_{2,i}^n + \dots.$$

We also define $\Pi_i := \sum_{n=0}^{\infty}\omega_n p_n E_{0,i}^n$. Now, plugging these expansions into the scheme (8.4) and identifying the powers of ε, we get:

- order (ε^0):

$$E_{0,i}^n = \Pi_i = \sum_{n=0}^{\infty} \omega_n p_n E_{0,i}^n,$$
$$F_{0,i}^n = 0,$$

  which is valid as soon as the quadrature weights ω_n satisfy Asumption (8.5).

- order (ε^1):

$$F_{1,i}^{n+1} = F_{1,i}^n - \frac{1}{3}\mu(\Pi_{i+1} - \Pi_{i-1}) = -(n+1)\frac{1}{3}\mu(\Pi_{i+1} - \Pi_{i-1}), \qquad (8.7)$$

  by summing up both sites from 0 to n.

- order (ε^2):

$$E_{2,i}^{n+1} = E_{2,i}^n - \mu(F_{1,i+1}^n - F_{1,i-1}^n) + \mu(\Pi_{i+1} - 2\Pi_i + \Pi_{i-1}),$$
$$E_{2,i}^0 = \sum_{n=0}^{\infty} \omega_n p_n E_{2,i}^n - \kappa\Pi_i + \langle s \rangle Q_i.$$

Inserting (8.7) inside the first equation and summing over n, then inserting this into the the second equation gives:

$$-\frac{1}{3}\sum_{n=0}^{\infty} \frac{n(n-1)}{2}\Delta s^2 \omega_n p_n \frac{\Pi_{i+2} - 2\Pi_i + \Pi_{i-2}}{4\Delta x^2}$$
$$- \Delta x \sum_{n=0}^{\infty} \Delta s n \omega_n p_n \frac{\Pi_{i+1} - 2\Pi_i + \Pi_{i-1}}{\Delta x^2} + \kappa\Pi_i = \langle s \rangle Q_i,$$

which is consistent with (6.7). Furthermore, this is a second order approximation under the Condition (8.6). □

Remark 8.2. Note that p is a probability density and the left-hand side of Condition (8.6) is a discretization of the first moment of p. Consequently this requires that at least some of the weights ω_n are negative, which is non-standard and usually not desirable.

Remark 8.3. As we have seen in the previous subsection in Equation (8.3), the zeroth order term of $P(E, F)$ is $\frac{1}{3}E$ both in the P_1 and in the M_1- case. Consequently, this calculation is independent from the closure and the theorem holds for the P_1 and the M_1 case.

Remark 8.4. Typically, HLL schemes are not asymptotic preserving. Only the manipulation of the numerical diffusion term by multiplying an additional factor ε yields this property in this case. It is obvious that this derivation would break otherwise.

8.2.3 Admissibility, realizability, and stability

The M_1 closure is a function of the relative flux F/E and is not defined everywhere. We immediately see that χ is only defined if $F/E \leq 4/3$. This condition is called **admissibility**. Furthermore, we say that F and E are **realizable** if and only if they are the first and zeroth angular moment of an underlying distribution on the full phase space. This is mathematically equivalent to the condition

$$0 \leq |F| \leq E, \quad \text{or} \quad \left|\frac{F}{E}\right| < 1, \tag{8.8}$$

both for the M_1 and the P_1 case. We see that realizability implies admissibility in this case, and that the set of realizable states is convex. The following theorem on admissibility and realizability of our numerical scheme holds.

Theorem 8.2. *There is $\Delta s_0 > 0$ such that if $\Delta s < \Delta s_0$, the numerical scheme (8.4) preserves the convex set of admissible and realizable states $\Omega = \{(E,F) \mid E > 0, \ \left|\frac{F}{E}\right| < 1\}$.*

Proof. We define $U_i^n := (E_i^n, F_i^n)^\top$ and $G(U_i^n) := (\varepsilon F_i^n, \varepsilon P_i^n)^\top$ so that we can write the scheme (8.4) in the usual compact form:

$$U_i^{n+1} = U_i^n - \frac{\Delta s}{\Delta x}(\mathcal{F}_{i+\frac{1}{2}}^n - \mathcal{F}_{i-\frac{1}{2}}^n),$$

with numerical flux

$$\mathcal{F}_{i+\frac{1}{2}}^n = \mathcal{F}(U_i^n, U_{i+1}^n) = \frac{1}{2}(G(U_{i+1}) - G(U_i^n)) - \frac{\varepsilon^2}{2}(U_{i+1}^n - U_i^n).$$

The scheme may be interpreted as an approximate Riemann solver in the fashion of Harten, Lax and Van Leer (see Section 3 in [57] for the abstract framework). It consists of two constant intermediate states separated by a stationary wave. Similar computations were carried out for the M_1 model of classical radiative transfer in [11]. We assume that the two other wave speeds are respectively $-b$ and b, where b is to be made precise later on. That means at each discrete point and in each discrete time step we consider a Riemann problem with left state U_L and right state U_R. Thus U has the following approximate form:

$$U(s,x) = \begin{cases} U_R & \text{if } \frac{x}{s} > b, \\ U_R^\star & \text{if } 0 < \frac{x}{s} < b, \\ U_L^\star & \text{if } -b < \frac{x}{s} < 0, \\ U_L & \text{if } \frac{x}{s} < -b, \end{cases}$$

where $U_L := (E_L, F_L)^\top$, $U_R := (E_R, F_R)^\top$, $U_L^\star := (E_L^\star, F_L^\star)^\top$ and $U_R^\star := (E_R^\star, F_R^\star)^\top$. This is also illustrated in Figure 8.1.
The approximate states $U_L^\star$ and $U_R^\star$ can be obtained as follows. The numerical flux of such an approximate Riemann solver can be expressed as a function of

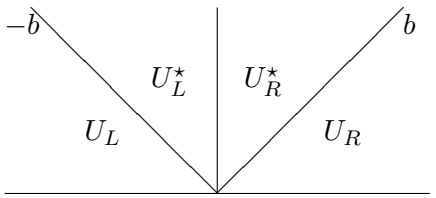


Figure 8.1: Illustration of the two-states approximate Riemann solver.

the intermediate states as in [57]:

$$\mathcal{F}(U_L, U_R) = G(U_R) - \frac{\Delta x}{2\Delta s} U_R + \frac{1}{\Delta s} \int_0^{\frac{\Delta x}{2}} U(s, x)\, \mathrm{d}x.$$

Using the form of the approximated solution $U(s, x)$, we have

$$\mathcal{F}(U_L, U_R) = G(U_R) - \frac{\Delta x}{2\Delta s} U_R + b(U_R^\star - U_R),$$

which leads to

$$\begin{aligned} U_R^\star &= U_R + \frac{1}{b}(\mathcal{F}(U_L, U_R) - G(U_R)) \\ &= U_R + \frac{1}{2b}\Big(G(U_L) - G(U_R) + \varepsilon^2(U_L - U_R)\Big). \end{aligned} \tag{8.9}$$

A similar computation yields

$$U_L^\star = U_L + \frac{1}{2b}\Big(G(U_R) - G(U_L) + \varepsilon^2(U_L - U_R)\Big).$$

We first consider the values of b which ensure the positivity of $E_L^\star$ and $E_R^\star$. We have

$$\begin{aligned} E_R^\star &= E_R + \frac{\varepsilon}{2b}(F_L - F_R) + \frac{\varepsilon^2}{2b}(E_L - E_R), \\ E_L^\star &= E_L + \frac{\varepsilon}{2b}(F_R - F_L) + \frac{\varepsilon^2}{2b}(E_L - E_R), \end{aligned}$$

therefore $E_L^\star > 0$ and $E_R^\star > 0$ if

$$b > b_1 := \frac{\varepsilon}{2} \max\Big(\frac{(F_R - F_L) + \varepsilon(E_R - E_L)}{E_R}, \frac{(F_L - F_R) + \varepsilon(E_R - E_L)}{E_L}\Big).$$

Then we consider the values of b which ensure the flux limitation i.e. $|F_L^\star| < E_L^\star$ and $|F_R^\star| < E_R^\star$. We write

$$F_R^\star = F_R + \frac{\varepsilon}{2b}(P_L - P_R) + \frac{\varepsilon^2}{2b}(F_L - F_R),$$

then $F_R^\star < E_R^\star$ and $-F_R^\star < E_R^\star$ if

$$b > b_2 := \frac{\varepsilon}{2}\max\Big(\beta_{2,1}, \beta_{2,2}\Big)$$

holds, where we have defined

$$\beta_{2,1} = \frac{P_L - P_R - F_L + F_R + \varepsilon(F_L - E_L - F_R + E_R)}{E_R - F_R},$$
$$\beta_{2,2} = \frac{-P_L + P_R - F_L + F_R + \varepsilon(-F_L - E_L + F_R + E_R)}{E_R + F_R},$$

and $P_L := P(E_L, F_L)$, $P_R := P(E_R, F_R)$. Similarly we have

$$F_L^\star = F_L + \frac{\varepsilon}{2b}(P_R - P_L) + \frac{\varepsilon^2}{2b}(F_L - F_R),$$

consequently $F_L^\star < E_L^\star$ and $-F_L^\star < E_L^\star$ holds if

$$b > b_3 := \frac{\varepsilon}{2}\max\Big(\beta_{3,1}, \beta_{3,2}\Big),$$

where we use the notation

$$\beta_{3,1} = \frac{P_R - P_L - F_L + F_R + \varepsilon(F_L - E_L - F_R + E_R)}{E_R - F_R},$$
$$\beta_{3,2} = \frac{-P_R + P_L - F_L + F_R + \varepsilon(-F_L - E_L + F_R + E_R)}{E_R + F_R}.$$

Finally, the CFL condition is given by the constraint of non-interacting waves:

$$\Delta s \leq \frac{\Delta x}{2b}.$$

Putting together the three former requirements on b, we obtain:

$$\Delta s \leq \Delta s_0 := \frac{\Delta x}{2\max(b_1, b_2, b_3)}. \tag{8.10}$$

□

Remark 8.5. When $\varepsilon \to 0$, the constants b_1, b_2 and b_3 defined in the former proof tend to zero. Therefore, we have $\Delta s_0 \to \infty$. As a consequence, the condition $\Delta s \leq \Delta s_0$ is not restrictive when asymptotic regimes are considered.

Finally, using the Fourier transform, we can show that the explicit scheme is L^2-stable under a mild CFL condition in the linear P_1 case. This is stated by the following lemma.

Lemma 8.3. *In the linear (P1) case i.e. when $P = \frac{1}{3}E$, the scheme (8.4) is L^2-stable under the CFL condition:*

$$\Delta s \leq \frac{2\Delta x}{\varepsilon^2 + \frac{1}{3} + \left|\frac{1}{3} - \varepsilon^2\right|}. \tag{8.11}$$

Proof. We introduce the auxiliary variables $U := F + E/\sqrt{3}$ and $V := F - E/\sqrt{3}$ to rewrite (8.4a)-(8.4b) as follows:

$$\begin{aligned} U_i^{n+1} &= (1 - 2\varepsilon^2\mu)U_i^n - \frac{\varepsilon\mu}{\sqrt{3}}(U_{i+1}^n - U_{i-1}^n) + \varepsilon^2\mu(U_{i+1}^n - U_{i-1}^n), \\ V_i^{n+1} &= (1 - 2\varepsilon^2\mu)V_i^n + \frac{\varepsilon\mu}{\sqrt{3}}(V_{i+1}^n - V_{i-1}^n) + \varepsilon^2\mu(V_{i+1}^n - V_{i-1}^n). \end{aligned}$$

Now, we perform a discrete Fourier transform in space ($\hat{U}$ and $\hat{V}$ respectively represent the discrete Fourier transforms of U and V) by setting

$$\hat{U}^{n+1} = A(\xi)\hat{U}^n, \quad \hat{V}^{n+1} = \overline{A(\xi)}\hat{V}^n,$$

with

$$A(\xi) = (1 - 2\varepsilon^2\mu + 2\varepsilon^2\mu\cos(\Delta x\xi)) - \frac{2i\varepsilon\mu}{\sqrt{3}}\sin(\Delta x\xi)).$$

Therefore, the scheme is L^2-stable if and only if for all $\xi \in \mathbb{R}$ we have $|A(\xi)| \leq 1$. Since

$$|A(\xi)|^2 - 1 = 4\varepsilon^2\mu(1 - \cos(\Delta x\xi))\Big(\varepsilon^2\big(1 - \cos(\Delta x\xi)\big) - 1 + \frac{1}{3}\mu\big(1 + \cos(\Delta x\xi)\big)\Big),$$

the scheme is stable if and only if the last term is positive for any $\xi \in \mathbb{R}$. Hence it is stable provided the following CFL condition

$$\mu \leq \frac{1}{\varepsilon^2 + \frac{1}{3} + \left|\frac{1}{3} - \varepsilon^2\right|},$$

holds, which is equivalent to (8.11). □

Even though the restrictions on Δs imposed by the former theorems are not extreme, our problem requires to iterate the computation of the approximate solutions up to the limit $s \to \infty$. Hence, it is more convenient to use an implicit scheme, especially for 2D and 3D simulations.

8.2.4 AP property of implicit schemes

We now consider the implicit version of the scheme (8.2):

$$E_i^{n+1} = E_i^n - \varepsilon\mu(F_{i+1}^{n+1} - F_{i-1}^{n+1}) + \varepsilon^2\mu(E_{i+1}^{n+1} - 2E_i^{n+1} + E_{i-1}^{n+1}), \tag{8.12a}$$
$$F_i^{n+1} = F_i^n - \varepsilon\mu(P_{i+1}^{n+1} - P_{i-1}^{n+1}) + \varepsilon^2\mu(F_{i+1}^{n+1} - 2F_i^{n+1} + F_{i-1}^{n+1}), \tag{8.12b}$$
$$E_i^0 = (1-\kappa\varepsilon^2)\sum_{n=0}^{\infty}\omega_n p_n E_i^n + \varepsilon^2\langle s\rangle Q_i,\ F_i^0 = 0, \tag{8.12c}$$

where $\mu = \frac{\Delta s}{2\Delta x}$ and $P_i^n = \chi(E_i^n, F_i^n)E_i^n$. Similar to the explicit case, the following proposition holds.

Proposition 8.4. *Assume the quadrature satisfies the Condition (8.5), then the numerical scheme (8.12) is asymptotic-preserving i.e. it degenerates into a scheme consistent with (6.7) when $\varepsilon \to 0$.*
Furthermore, the limit scheme is second-order accurate if the Condition (8.6) holds.

Proof. The proof is very similar to the explicit case. We perform an expansion of E_i^n and F_i^n by

$$E_i^n = E_{0,i}^n + \varepsilon E_{1,i}^n + \varepsilon^2 E_{2,i}^n + \dots,$$
$$F_i^n = F_{0,i}^n + \varepsilon F_{1,i}^n + \varepsilon^2 F_{2,i}^n + \dots.$$

We also define $\Pi_i := \sum_n \omega_n p_n E_{0,i}^n$. Now, plugging these expansions into the scheme (8.12) and identifying the powers of ε, we get, for the different orders of ε:

- order (ε^0):

$$E_{0,i}^n = \Pi_i = \sum_{n=0}^{\infty}\omega_n p_n E_{0,i}^n,$$
$$F_{0,i}^n = 0.$$

  which is valid as soon as the quadrature weights ω_n satisfy the assumption (8.5).

- order (ε^1):

$$F_{1,i}^{n+1} = F_{1,i}^n - \frac{1}{3}\mu(\Pi_{i+1} - \Pi_{i-1}) = -(n+1)\frac{1}{3}\mu(\Pi_{i+1} - \Pi_{i-1}), \tag{8.13}$$

  again by summation up to n.

- order (ε^2):

$$E_{2,i}^{n+1} = E_{2,i}^n - \mu(F_{1,i+1}^{n+1} - F_{1,i-1}^{n+1}) + \mu(\Pi_{i+1} - 2\Pi_i + \Pi_{i-1}),$$
$$E_{2,i}^0 = \sum_{n=0}^{\infty} \omega_n p_n E_{2,i}^n - \kappa\Pi_i + \langle s \rangle Q_i.$$

Inserting (8.13) inside the first equation and summing over n, then inserting into the second equation gives:

$$-\frac{1}{3}\sum_{n=0}^{\infty} \frac{n(n+1)}{2}\Delta s^2 \omega_n p_n \frac{\Pi_{i+2} - 2\Pi_i + \Pi_{i-2}}{4\Delta x^2}$$
$$-\Delta x \sum_{n=0}^{\infty} n\omega_n p_n \Delta s \frac{\Pi_{i+1} - 2\Pi_i + \Pi_{i-1}}{\Delta x^2} + \kappa\Pi_i = \langle s \rangle Q_i,$$

which is consistent with (6.7). Furthermore, this is once again a second order approximation under the condition (8.6). □

Remark 8.6. Comparing the diffusion limit of the implicit and explicit scheme we observe the following. The discretizations of the diffusion coefficient $\langle s^2 \rangle / 2$ are given by

$$\sum_{n=0}^{\infty} \frac{n(n+1)}{2}\Delta s^2 \omega_n p_n \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{n(n-1)}{2}\Delta s^2 \omega_n p_n,$$

respectively. Both discretizations yield an error of order Δs, which can be eliminated using a Crank-Nicholson scheme because of the alternating sign. The additional error term of order Δx is not eliminated in this way.

9 Iterative methods

We presented the numerical schemes in the previous chapter as a single linear system

$$A\Psi = q$$

that has to be solved at once. In practice, the size of A can be arbitrarily large, since the domain of both the pseudo-time variable s and the spatial variable x are not bounded (at least theoretically). That means that a direct solution can become arbitrarily inefficient only because of the sheer amount of memory used.
Therefore, we solve the system iteratively. It turns out that a simple source iteration method converges arbitrarily slow if the system is dominated by scattering, that means if the scattering coefficient $c \approx 1$. A suited initialization strategy can decrease the number of iterations significantly, but a more sophisticated method that relies on an approximation of the residual in every step can also decrease the contraction rate of the iteration.
We first present the source iteration method and initialization strategies, then introduce the Diffusion Synthetic Acceleration method as a natural extension of the source iteration method. A von Neumann analysis for the full equation as well as for the moment models shows, that source iteration converges slowly if $c \approx 1$, and that the contraction rate is significantly improved by the use of the Diffusion Synthetic Acceleration method.

9.1 Source iteration

The most simple iterative technique for transport equations is the so-called source iteration method. The idea of this iterative method can be presented best on an abstract operator level. Hence, we define the linear solution operator

$$L\Psi_0 = \Psi(s, x, \mu; \Psi_0),$$

that maps the initial function Ψ_0 to the solution Ψ, that means Ψ is the solution of the homogeneous Cauchy problem

$$\partial_s \Psi(s, x, \mu) + \varepsilon\mu\partial_x \Psi(s, x, \mu) = 0, \quad \Psi(0, x, \mu) = \Psi_0(x, \mu). \tag{9.1}$$

Furthermore, we define the linear integral operator

$$S\Psi = c \int_0^\infty \int_{-1}^1 \sigma(\mu, \mu') p(s) \Psi(s, x, \mu) \, \mathrm{d}\mu' \, \mathrm{d}s. \tag{9.2}$$

with $c = 1 - \kappa\varepsilon^2$. Setting $q := Q(x, \mu)$, a solution of Equation (6.4) satisfies the fixed point equation

$$\Psi = L(S\Psi + q).$$

Then the source iteration method works as follows: choose initial data $\Psi^{(0)}$, and in every step solve

$$\Psi^{(n+1)} = L(S\Psi^{(n)} + q). \tag{9.3}$$

This method is very simple and close to the physical interpretation. If we initialize by $\Psi^{(0)} \equiv 0$, the iterate $\Psi^{(n)}$ can be interpreted as the particles which have been scattered at most n times. But as we see in the following, this method has a severe drawback: it converges very slowly, namely with a contraction rate $c \approx 1$ if the system is dominated by scattering.

Initialization strategies

The source iteration Algorithm (9.3) requires a suitable initialization to overcome the problem of slow convergence. One way to choose it is to remark that in the case of a P_1 closure, the exact solution can be expressed in the following way. Assuming that (E, F) is the solution of the P_1 model, we define

$$\Pi(x) = \int_0^\infty p(s)E(s, x)ds,$$

and formally calculate its first derivative as

$$\begin{aligned}
\Pi'(x) &= \int_0^\infty p(s)\partial_x E(s, x)ds = -\frac{3}{\varepsilon}\int_0^\infty p(s)\partial_s F(s, x)ds \\
&= -\frac{3}{\varepsilon}\Big[\partial_s p(0)E(0, x)\Big]_0^\infty + \frac{3}{\varepsilon}\int_0^\infty \partial_s p(s)F(s, x)ds \\
&= \frac{3}{\varepsilon}\int_0^\infty \partial_s p(s)F(s, x)ds.
\end{aligned}$$

Then the second derivative of Π can be computed as

$$\begin{aligned}
\Pi''(x) &= \frac{3}{\varepsilon}\int_0^\infty \partial_s p(s)\partial_x F(s, x)ds = -\frac{3}{\varepsilon^2}\int_0^\infty \partial_s p(s)\partial_s E(s, x)ds \\
&= -3\partial_s p(0)\langle s\rangle Q(x) + 3\kappa\partial_s p(0)\Pi(x) \\
&\qquad + \underbrace{\left(-\frac{3}{\varepsilon^2}\partial_s p(0)\Pi(x) + \frac{3}{\varepsilon^2}\int_0^\infty \partial_s^2 p(s)E(s, x)ds\right)}_{(*)},
\end{aligned}$$

and, if $\Sigma_t(s) \equiv \Sigma_t$ is constant, the last term becomes 0 and this equation reads

$$\Pi''(x) = 3\kappa\Sigma_t^2\Pi(x) - 3\Sigma_t^2\langle s\rangle Q(x). \tag{9.4}$$

Since in this case $\langle s \rangle = 1/\Sigma_t$ and $\langle s^2 \rangle = 2/\Sigma_t^2$, Equation (9.4) equals the diffusion approximation (6.7) for $\mu_0 = 0$. The solution of this ODE is used as an initial guess in the Algorithm (9.3), equipped with the boundary conditions of the system, even though it is not valid anymore when the M_1 model is selected.
For nonconstant $\Sigma_t(s)$ the terms $(*)$ do not disappear, but we can neglect them and take the numerical solution of

$$\Pi''(x) = -3\kappa \partial_s p(0)\Pi(x) + 3\partial_s p(0)\langle s \rangle Q(x)$$

as an initial value for $\int_0^\infty p(s)E(s,x)ds$ in the iteration of (9.3).

Remark 9.1. This formal calculation gives us another interesting insight into the differences between the classical and the non-classical transport equation. It is known that for the classical linear Boltzmann equation, the P_1 model is equivalent to the diffusion approximation, at least for homogeneous total cross sections. We recover this fact for $\Sigma_t(s) \equiv \Sigma_t$ constant, that means in the case that is equivalent to the classical equation. However, this is not the case if $\Sigma_t(s)$ is not constant. We also see the reason for this in the equation for $\Pi''(x)$: the non-exponential form of p produces constants such that the terms $(*)$ do not cancel out any more.

9.2 Diffusion Synthetic Acceleration

Since the convergence of the above presented source iteration (SI) is very slow because the contraction rate is $c \approx 1$ for scattering dominated systems, we have to employ a suitable acceleration technique. In classical radiative transfer the so-called Diffusion Synthetic Acceleration (DSA) is widely used (see [1] for an comprehensive review). We extend this method to the non-classical, "pseudo-time"-dependent problem.
We follow the argumentation in [1]. Let L be again the linear solution operator of the Cauchy problem (9.1) and S the integral operator (9.2). Diffusion Synthetic Acceleration is a two-step-procedure: first we make one source iteration step, then calculate an approximation of the residual using the diffusion approximation and correct the current iterate. The residual in each iteration step is the difference between the source iteration step and the exact solution:

$$\begin{aligned} \Psi^{(n+\frac{1}{2})} &= L(S\Psi^{(n)} + q), && \text{(one source iteration)} \\ \Psi &= L(S\Psi + q), && \text{(exact solution)} \\ \Delta\Psi = \Psi - \Psi^{(n+\frac{1}{2})} &= L(S\Psi - S\Psi^{(n)}), && \text{(residual)} \end{aligned}$$

and we correct the iteration by

$$\Psi^{(n+1)} = \Psi^{(n+\frac{1}{2})} + \Delta\Psi.$$

The residual satisfies

$$\begin{aligned}\Delta\Psi = \Psi - \Psi^{(n+\frac{1}{2})} &= L(S\Psi - S\Psi^{(n)}) \\ &= L(S\Psi \pm S\Psi^{(n+\frac{1}{2})} - S\Psi^{(n)}) \\ &= L(S\Delta\Psi) + L(\underbrace{S(\Psi^{(n+\frac{1}{2})} - \Psi^{(n)})}_{=:\bar{q}}).\end{aligned}$$

This can be written as

$$\Delta\Psi - L(S\Delta\Psi) = L\bar{q},$$

which is Equation (6.9) for $\Delta\Psi$ with source term $\bar{q}$. We can therefore approximate the residual $\Delta\Psi$ using the diffusion approximation

$$-\frac{1}{3}\left(\frac{\langle s^2\rangle}{2} + \frac{\bar{\mu}_0}{1-\bar{\mu}_0}\right)\nabla_x^2\left(\Delta\Psi(x)\right) + \kappa\Delta\Psi(x) = \langle s\rangle\bar{q}(x). \tag{9.5}$$

We only need the s-integral of $p(s)\Delta\Psi$, so we update by

$$\int_0^\infty p(s)\Psi^{(n+1)}(s,x)\,\mathrm{d}s = \int_0^\infty p(s)\Psi^{(n+\frac{1}{2})}(s,x)\,\mathrm{d}s + \int_0^\infty p(s)\overline{\Delta\Psi}(x)\,\mathrm{d}s$$

for the initial value in the next step.

To sum up, the Diffusion Synthetic Acceleration algorithm for the non-classical transport equation reads:

Diffusion Synthetic Acceleration	
(0)	Choose an initial setting for $\Psi^{(0)}$ (or only $S\Psi^{(0)}$).
	Then in every step:
(1)	Make one source iteration: $\Psi^{(n+\frac{1}{2})} = L(S\Psi^{(n)} + q)$.
(2)	Diffusion approximation for the residual $\Delta\Psi^{(n+\frac{1}{2})} = \Psi - \Psi^{(n+\frac{1}{2})}$.
(3)	Calculate new initial value: $S\Psi^{(n+1)} = S\Psi^{(n+\frac{1}{2})} + S\Delta\Psi^{n+\frac{1}{2}}$.

DSA is a preconditioner

For the classical transport equation it is known that DSA is equivalent to a preconditioned Richardson iteration [1]. We show that the same holds for the non-classical equation. Using the operator notation from above, the non-

classical equation can be written as

$$(I - LS)\Psi = Lq,$$

hence, the source iteration method can be interpreted as a Richardson iteration:

$$\Psi^{(n+1)} = L(S\Psi^{(n)} + q) = \Psi^{(n)} - ((I - LS)\Psi^{(n)} - Lq).$$

Assuming that we approximate the residual exactly, we can write one step of DSA directly as

$$\Psi^{(n+1)} = \Psi^{(n)} - ((I - LS)\Psi^{(n)} - Lq) + (I - LS)^{-1}L(S(L(S\Psi^{(n)} + q) - \Psi^{(n)})).$$

Using only the diffusion approximation for the residual means that we use an approximation for the operator

$$M \approx (I - LS)^{-1},$$

where M is the solution operator of the diffusion Approximation (9.5), that maps a given source term to the corresponding solution. Then we can write DSA as

$$\begin{aligned}
\Psi^{(n+1)} &= \Psi^{(n)} - \big((I - LS)\Psi^{(n)} - Lq\big) + MLS\big(L(S\Psi^{(n)} + q) - \Psi^{(n)}\big)\big) \\
&= \Psi^{(n)} - \big((I - LS)\Psi^{(n)} - Lq - MLSLS\Psi^{(n)} - MLSLq + MLS\Psi^{(n)}\big) \\
&= \Psi^{(n)} - \big((I - LS - MLSLS + MLS)\Psi^{(n)} - (L + MLSL)q\big) \\
&= \Psi^{(n)} - (I + MLS)\big((I - LS)\Psi^{(n)} - Lq\big).
\end{aligned}$$

This has now the form of a preconditioned Richardson iteration with preconditioner $P^{-1} = I + MLS$. Note that the operator M acts on a lower dimensional space than $(I - LS)^{-1}$. Hence, the operator S needs to be applied to both sides of the equation above to compute the update $S\Psi^{(n+1)}$.

9.3 Von Neumann analysis for the non-classical equation

In this section we show that the contraction rate of source iteration is determined by the scattering ration c, and that DSA in fact improves the speed of convergence by significantly decreasing the contraction rate. We are interested in cases where the scattering ratio $c \approx 1$. Using the scaled version of (6.4), we set $\varepsilon = 1$ and we restrict our analysis to the simplest case of slab geometry and isotropic scattering. Then one source iteration step reads

$$\begin{aligned}
&\partial_s \Psi^{(l+\frac{1}{2})} + \mu \partial_x \Psi^{(l+\frac{1}{2})} = 0, \\
&\Psi^{(l+\frac{1}{2})}(x, \mu, 0) = c \int_0^\infty \int_{-1}^1 \frac{1}{2} p(s) \Psi^{(l)}(x, \mu, s) \, \mathrm{d}\mu \, \mathrm{d}s + \langle s \rangle Q(x),
\end{aligned}$$

and the diffusion approximation for this equation is given by

$$-\frac{1}{3}\frac{\langle s^2\rangle}{2\langle s\rangle}\nabla_x^2\Phi(x)+\frac{1-c}{\langle s\rangle}\Phi(x)=Q(x). \tag{9.6}$$

Now the error in every step is $f := \Psi - \Psi^{(l+\frac{1}{2})}$ and satisfies the equation

$$\partial_s f^{(l+\frac{1}{2})}(s,x,\mu)+\mu\partial_x f^{(l+\frac{1}{2})}(s,x,\mu)=0,$$

with initial condition

$$\begin{aligned} f^{(l+\frac{1}{2})}(0,x,\mu) &= c\int_0^\infty\int_{-1}^1 \frac{1}{2}p(s)(\Psi-\Psi^{(l)}(x,\mu,s))\,\mathrm{d}\mu\,\mathrm{d}s \\ &= c\int_0^\infty\int_{-1}^1 \frac{1}{2}p(s)f^{(l+\frac{1}{2})}(x,\mu,s)\,\mathrm{d}\mu\,\mathrm{d}s \\ &\qquad + c\int_0^\infty\int_{-1}^1 \frac{1}{2}p(s)(\Psi^{(l+\frac{1}{2})}-\Psi^{(l)})(x,\mu,s)\,\mathrm{d}\mu\,\mathrm{d}s. \end{aligned}$$

For the von Neumann analysis we define the quantities $\phi^{(l)}$ as

$$\phi^{(l)} := \int_0^\infty\int_{-1}^1 \frac{1}{2}p(s)f^{(l)}(x,\mu,s)\,\mathrm{d}\mu\,\mathrm{d}s,$$

and $F^{(l+\frac{1}{2})}$ as the solution of

$$-\frac{1}{3}\frac{\langle s^2\rangle}{2\langle s\rangle}\nabla_x^2 F^{(l+\frac{1}{2})}(x)+\frac{1-c}{\langle s\rangle}F^{(l+\frac{1}{2})}(x)=\frac{c}{\langle s\rangle}\left(\phi^{(l+\frac{1}{2})}-\phi^{(l)}\right).$$

We apply the following Fourier ansatz, where $\omega(\lambda)$ is the contraction rate of the source iteration with respect to a specific Fourier mode λ:

$$\begin{aligned} \phi^{(l)}(x) &= \omega^l(\lambda)\mathrm{e}^{i\lambda x}, & \phi^{(l+\frac{1}{2})}(x) &= \omega^l(\lambda)\beta(\lambda)\mathrm{e}^{i\lambda x}, \\ f^{(l+\frac{1}{2})}(x,\mu,s) &= \omega^l(\lambda)\alpha(\lambda,\mu,s)\mathrm{e}^{i\lambda x}, & F^{(l+\frac{1}{2})}(x) &= \omega^l(\lambda)\gamma(\lambda)\mathrm{e}^{i\lambda x}. \end{aligned}$$

Then the equation for the error, which is the original equation with source term $Q\equiv 0$, reads

$$\begin{aligned} &\omega^l(\lambda)\partial_s\alpha(\lambda,\mu,s)\mathrm{e}^{i\lambda x}+\mu\partial_x\omega^l(\lambda)\alpha(\lambda,\mu,s)\mathrm{e}^{i\lambda x}=0, \\ &\omega^l(\lambda)\mathrm{e}^{i\lambda x}\alpha(\lambda,\mu,0)=c\omega^l(\lambda)\mathrm{e}^{i\lambda x}, \end{aligned}$$

and this reduces to

$$\partial_s \alpha(\lambda,\mu,s) = -i\lambda\mu x\alpha(\lambda,\mu,s), \quad \alpha(\lambda,\mu,0) = c,$$

hence, $\alpha(\lambda,\mu,s) = c\mathrm{e}^{-i\lambda\mu s}$. So we calculate in a next step

$$\phi^{(l+\frac{1}{2})} = \omega^l(\lambda)\beta(\lambda)\mathrm{e}^{i\lambda x} = \int_0^\infty \int_{-1}^1 \frac{1}{2} p(s)\omega^l(\lambda)\alpha(\lambda,\mu,s)\mathrm{e}^{i\lambda x}\,\mathrm{d}\mu\,\mathrm{d}s,$$

hence

$$\beta(\lambda) = \int_0^\infty \int_{-1}^1 \frac{1}{2} p(s)\alpha(\lambda,\mu,s)\,\mathrm{d}\mu\,\mathrm{d}s.$$

If $\Sigma_t(s) \equiv \Sigma_t$, then we have $p(s) = \Sigma_t \exp(-s\Sigma_t)$, and we can calculate $\beta(\lambda)$ as

$$\begin{aligned}
\beta(\lambda) &= \int_0^\infty \int_{-1}^1 \frac{1}{2} p(s)\alpha(\lambda,\mu,s)\,\mathrm{d}\mu\,\mathrm{d}s = \int_0^\infty \int_{-1}^1 \frac{1}{2}\Sigma_t \exp(-s\Sigma_t)c\mathrm{e}^{-i\lambda\mu s}\,\mathrm{d}\mu\,\mathrm{d}s \\
&= \frac{c}{2}\Sigma_t \int_0^\infty \int_{-1}^1 \exp(-s\Sigma_t)(\cos(\lambda\mu s) - i\sin(\lambda\mu s))\,\mathrm{d}\mu\,\mathrm{d}s \\
&= \frac{c}{2}\Sigma_t \int_{-1}^1 \left(\frac{\Sigma_t}{\Sigma_t^2 + (\lambda\mu)^2} - \frac{i\lambda\mu}{\Sigma_t^2 + (\lambda\mu)^2}\right)\mathrm{d}\mu \\
&= \frac{c}{2}\int_{-1}^1 \frac{\Sigma_t^2}{\Sigma_t^2 + (\lambda\mu)^2}\,\mathrm{d}\mu = \frac{c}{2}\left[\left(\frac{\Sigma_t}{\lambda}\right)^2 \frac{\lambda}{\Sigma_t}\arctan\left(\frac{\mu}{\frac{\Sigma_t}{\lambda}}\right)\right]_{\mu=-1}^{\mu=1} \\
&= \frac{c}{2}\frac{\Sigma_t}{\lambda}\left[\arctan\left(\frac{\lambda}{\Sigma_t}\right) - \arctan\left(-\frac{\lambda}{\Sigma_t}\right)\right] = c\frac{\Sigma_t}{\lambda}\arctan\left(\frac{\lambda}{\Sigma_t}\right).
\end{aligned}$$

This is the same result as calculated for the classical transport equation in [1]. There it is shown that $\beta(\lambda) \approx c$ for small λ and that it tends to 0 as λ becomes large.
For an arbitrary path length distribution $p(s)$ we obtain

$$\beta(\lambda) = c\int_0^\infty \int_{-1}^1 \frac{1}{2}\frac{(\alpha-1)}{(s+1)^\alpha}\exp(-i\mu\lambda s)\,\mathrm{d}\mu\,\mathrm{d}s = c\int_0^\infty \frac{(\alpha-1)\sin(\lambda s)}{\lambda s(s+1)^\alpha}\,\mathrm{d}s.$$

This integral cannot be solved in a closed form, but we see that

$$\left|\frac{\sin(\lambda s)}{\lambda s}\right| \leq 1, \text{ for all } \lambda, s, \quad \text{and}$$
$$\frac{\sin(\lambda s)}{\lambda s} \longrightarrow 1 \quad \text{for } \lambda \to 0 \quad \text{and every } s > 0 \text{ fixed.}$$

Consequently,

$$\lim_{\lambda \to 0} \beta(\lambda) = c.$$

If the function $p(s)$ decays algebraically as

$$p(s) = \frac{\alpha - 1}{(s+1)^{\alpha}}, \quad \alpha > 3,$$

numerical calculations also show that

$$\beta(\lambda) \longrightarrow 0, \text{ as } \lambda \to +\infty.$$

For the calculations we choose $\alpha = 3.5, 4, 4.5, 5, 6, 7$, and $c = 1$. The results are shown in Figure 9.1.

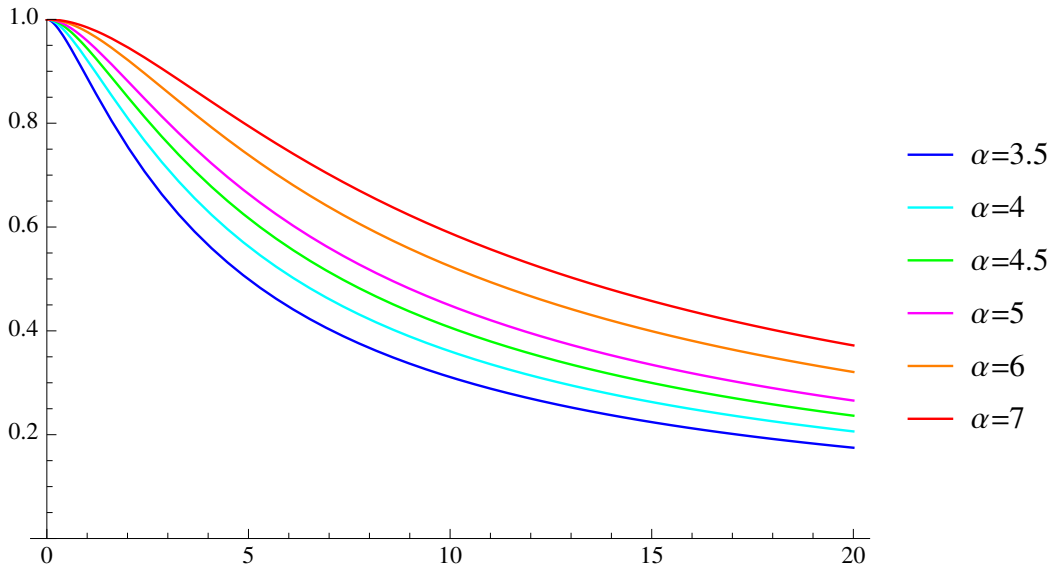


Figure 9.1: A comparison of β as a function of λ for different exponents α. For small modes β is close to 1 and decreases as $\lambda \to \infty$. The higher the exponent α, the slower the decrease of β.

We see here that the large modes are sufficiently suppressed while the small ones are not. Hence we are in the same situation as in the case above. Conse-

quently, for source iteration, the contraction rate is

$$\omega_{\mathrm{SI}} = \max_{-\infty<\lambda<+\infty}(\beta(\lambda)) = c.$$

For DSA we have to continue the calculation by a correction of the new iterate with an approximation of the residual

$$\phi^{(l+1)} = \phi^{(l+\frac{1}{2})} + F^{(l+\frac{1}{2})}.$$

We write the diffusion approximation for $F^{(l+\frac{1}{2})}$ as

$$DF^{(l+\frac{1}{2})} = \frac{c}{\langle s\rangle}(\phi^{l+\frac{1}{2}} - \phi^{(l)}).$$

Since $\phi^{(l+1)}(x) = \omega^{l+1}(\lambda)\mathrm{e}^{i\lambda x}$, we obtain

$$\begin{aligned}\gamma(\beta) &= D^{-1}(\beta(\lambda)-1) = \frac{6c}{\langle s^2\rangle\lambda^2 + 6(1-c)}(\beta(\lambda)-1),\\ \omega(\lambda) &= \beta(\lambda) + \gamma(\lambda).\end{aligned}$$

In detail this becomes

$$\begin{aligned}\omega(\lambda) &= \beta(\lambda) + \frac{6c}{\langle s^2\rangle\lambda^2 + 6(1-c)}(\beta(\lambda)-1)\\ &= \frac{6c}{\langle s^2\rangle\lambda^2 + 6(1-c)}\left(\left(\frac{\langle s^2\rangle\lambda^2}{6c} + \frac{1}{c}\right)\beta(\lambda) - 1\right),\end{aligned}$$

with $\beta(\lambda)$ depending on the choice of $p(s)$. For $\Sigma_t(s) \equiv \Sigma_t = 1$ we have

$$\omega_{\mathrm{DSA}}(\lambda) = \frac{3c}{\lambda^2 + 3(1-c)}\left(\left(\frac{\lambda^2}{3} + 1\right)\frac{\arctan(\lambda)}{\lambda} - 1\right),$$

just like in the case of classical transport, and this is bounded by $\omega_{\mathrm{DSA}}(\lambda) \leq 0.2247c$ (compare [1]). If the function $p(s)$ decays algebraically as

$$p(s) = \frac{\alpha-1}{(s+1)^\alpha}, \quad \alpha > 3,$$

we have

$$\omega_{\mathrm{DSA}}(\lambda) = \frac{6c}{\langle s^2\rangle\lambda^2 + 6(1-c)}\left(\left(\frac{\langle s^2\rangle\lambda^2}{6} + 1\right)\int_0^\infty \frac{\sin(\lambda s)}{\lambda s}p(s)\,\mathrm{d}s - 1\right).$$

Since the integral does not have a closed form solution, we compute the contraction rate again for the cases $\alpha = 3.5, 4, 4.5, 5, 6, 7$ with $c = 1$. The results are shown in Figure 9.2, and show a clear improvement of the contraction rates compared to the source iteration case. Small modes are now sufficiently suppressed.

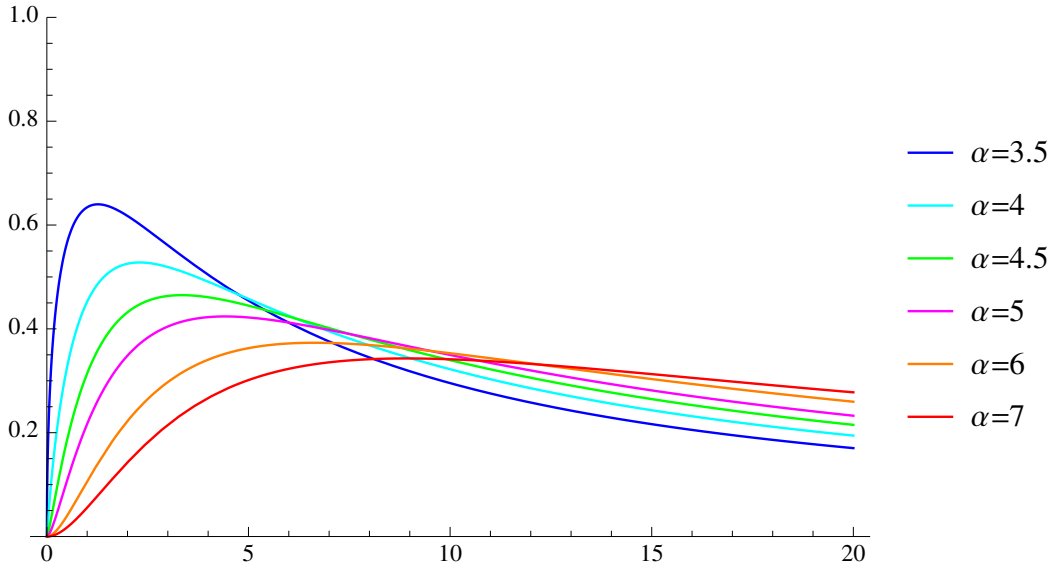


Figure 9.2: A comparison of ω as a function of λ for different exponents α and $c = 1$. The functions are bounded away from 1 in contrast to the source iteration case.

We conclude that the contraction rate is significantly decreased by DSA, especially for the small modes that cause slow convergence in the source iteration method. The higher the exponent α, the better is the improvement for the small modes, while the large modes are suppressed better if the exponent α is smaller.

9.4 Von Neumann analysis for moment models

In this section we analyze the contraction rate of the source iteration method and diffusion synthetic acceleration for moment models of the non-classical equation. We restrict our analysis here to the linear P_1 model, but we expect that the results remain true for the nonlinear M_1 closure because of Relation (8.3). As in the previous section we set $\varepsilon = 1$ and consider the scattering dominated case $c \approx 1$.
One source iteration step of the P_1 system reads

$$\begin{aligned}
&\partial_s E^{(l+\frac{1}{2})} + \nabla_x F^{(l+\frac{1}{2})} = 0, \qquad \partial_s F^{(l+\frac{1}{2})} + \frac{1}{3}\nabla_x E^{(l+\frac{1}{2})} = 0,\\
&E^{(l+\frac{1}{2})}(0,x) = c\int_0^\infty p(s)E^{(l)}(s,x)\,\mathrm{d}s + Q(x),\\
&F^{(l+\frac{1}{2})}(0,x) = 0,
\end{aligned}$$

and the diffusion approximation for E is given by Equation (9.6). The error between the iterate and the exact solution in the P_1 model is defined as

$$f^{(l+\frac{1}{2})} := \begin{pmatrix} E - E^{(l+\frac{1}{2})} \\ F - F^{(l+\frac{1}{2})} \end{pmatrix} = \begin{pmatrix} \tilde{E}^{(l+\frac{1}{2})} \\ \tilde{F}^{(l+1)} \end{pmatrix},$$

and it satisfies the equation

$$\partial_s \tilde{E}^{(l+\frac{1}{2})} + \nabla_x \tilde{F}^{(l+\frac{1}{2})} = 0, \qquad \partial_s \tilde{F}^{(l+\frac{1}{2})} + \frac{1}{3}\nabla_x \tilde{E}^{(l+\frac{1}{2})} = 0, \tag{9.7a}$$

$$\tilde{E}^{(l+\frac{1}{2})}(0,x) = c\int_0^\infty p(s)\tilde{E}^{(l)}(s,x)\,\mathrm{d}s, \tag{9.7b}$$

$$\tilde{F}^{(l+\frac{1}{2})}(0,x) = 0. \tag{9.7c}$$

which is simply Equation (7.2) without external source term. We now calculate the rate of convergence of the source iteration method and diffusion synthetic acceleration using a Fourier ansatz. Following again [1], we define the quantity $\phi^{(l)}(x)$ as

$$\phi^{(l)}(x) := \int_0^\infty p(s)\tilde{E}^{(l)}(x,s)\,\mathrm{d}s,$$

and $G^{(l+\frac{1}{2})}$ as the solution of the diffusion approximation

$$-\frac{1}{3}\frac{\langle s^2\rangle}{2\langle s\rangle}\nabla_x^2 G^{(l+\frac{1}{2})}(x) + \frac{1-c}{\langle s\rangle}G^{(l+\frac{1}{2})}(x) = \frac{c}{\langle s\rangle}\left(\phi^{(l+\frac{1}{2})} - \phi^{(l)}\right).$$

Analogous to the previous section, we apply the Fourier ansatz

$$\begin{aligned} \phi^{(l)}(x) &= \omega^l(\lambda)\mathrm{e}^{i\lambda x}, & \tilde{E}^{(l+\frac{1}{2})}(s,x) &= \omega^l(\lambda)\alpha(\lambda,s)\mathrm{e}^{i\lambda x}, \\ \phi^{(l+\frac{1}{2})}(x) &= \omega^l(\lambda)\gamma(\lambda)\mathrm{e}^{i\lambda x}, & \tilde{F}^{(l+\frac{1}{2})}(s,x) &= \omega^l(\lambda)\beta(\lambda,s)\mathrm{e}^{i\lambda x}, \\ G^{(l+\frac{1}{2})}(x) &= \omega^l(\lambda)\rho(\lambda)\mathrm{e}^{i\lambda x}, \end{aligned}$$

with $\lambda \in \mathbb{R}$ and α, β, ω and γ to be determined. The quantity $\gamma(\lambda)$ turns out to be the contraction rate of the source iteration method, and $\omega(\lambda)$ the contraction rate of the Diffusion Synthetic Acceleration method. Inserting this into Equations (9.7), and cancelling all common terms, we get

$$\begin{aligned} &\partial_s\alpha(\lambda,s) + i\lambda\beta(\lambda,s) = 0, && \partial_s\beta(\lambda,s) + i\lambda\frac{1}{3}\alpha(\lambda,s) = 0, \\ &\alpha(\lambda,0) = c, && \beta(\lambda,0) = 0. \end{aligned}$$

We now insert the left into the right equation and multiply by $i\lambda$, which gives us

$$-\partial_s^2\alpha(\lambda,s) - \frac{\lambda^2}{3}\alpha(\lambda,s) = 0,$$

with initial conditions (with respect to s):

$$\alpha(\lambda,0) = c, \qquad \partial_s\alpha(\lambda,0) = -i\lambda\beta(\lambda,0) = 0.$$

The solution of this ordinary differential equation is given by

$$\alpha(\lambda,s) = c\cos(\frac{1}{\sqrt{3}}\lambda s).$$

So in a next step, we insert this into the definition of $\phi^{(l+\frac{1}{2})}$, and also insert the Fourier ansatz to get an explicit form of the contraction rate

$$\gamma(\lambda) = \int_0^\infty p(s)\alpha(\lambda,s)ds.$$

If $\Sigma_t(s) \equiv \Sigma_t$, then we have $p(s) = \Sigma_t \exp(-s\Sigma_t)$, and we can calculate the contraction rate $\gamma(\lambda)$ as

$$\gamma(\lambda) = \int_0^\infty \Sigma_t \exp(-s\Sigma_t)c\cos(\frac{1}{\sqrt{3}}\lambda s)\mathrm{ds} = c\frac{\Sigma_t^2}{\Sigma_t^2 + \frac{1}{3}\lambda^2},$$

and we see that

$$\begin{aligned}\gamma(\lambda) &\longrightarrow 1, \quad \text{as } \lambda \to 0, \text{ and}\\ \gamma(\lambda) &\longrightarrow 0, \quad \text{as } \lambda \to \infty,\end{aligned}$$

in this case, independent of the choice of Σ_t. Hence, $\gamma(\lambda) \approx c$ for small Fourier modes λ, and it tends to 0 if λ becomes large. That means that large modes are sufficiently suppressed, while the small modes are responsible for slow convergence, which agrees with the results in [1], and our results in the previous section. For a general path length distribution $p(s)$ we obtain

$$\gamma(\lambda) = c\int_0^\infty p(s)\cos(\frac{1}{\sqrt{3}}\lambda s)\,\mathrm{d}s.$$

If the function $p(s)$ decays algebraically as

$$p(s) = \frac{\alpha - 1}{(s+1)^\alpha}, \quad \alpha > 3,$$

this reads

$$\gamma(\lambda) = c \int_0^\infty \frac{\alpha - 1}{(s+1)^\alpha} \cos(\frac{1}{\sqrt{3}}\lambda s)\, \mathrm{d}s.$$

This integral does not have a closed representation, numerical calculations recover the qualitative behavior from the previous section with

$$\gamma(\lambda) \approx c$$

for small values of λ. For the calculations we choose again $\alpha = 3.5, 4, 4.5, 5, 6, 7$ and $c = 1$. The results are shown in Figure 9.3.

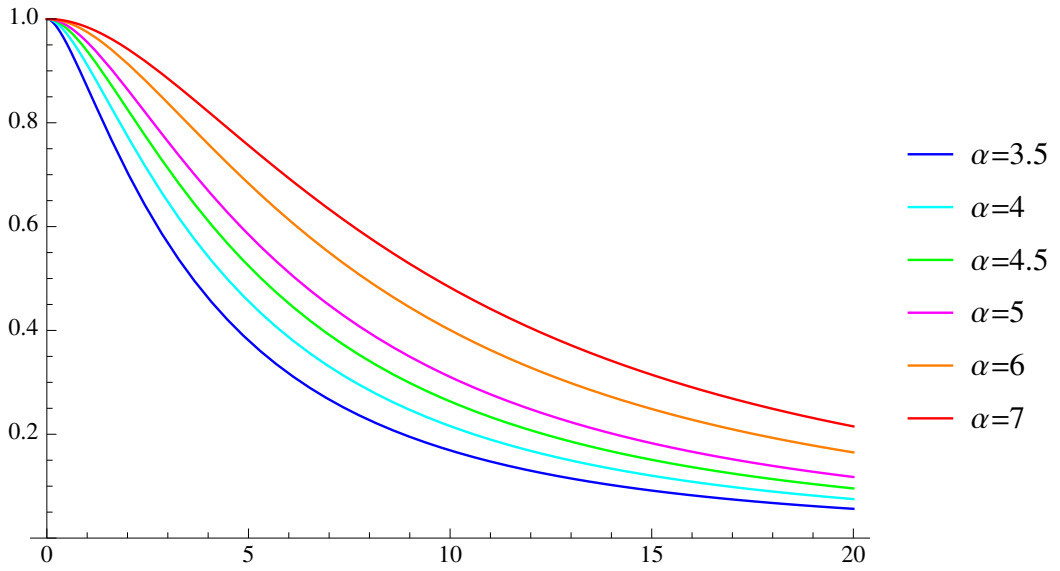


Figure 9.3: A comparison of γ as a function of λ for different exponents α and $c = 1$. We observe the same behavior as for the full equation.

We see again that the large modes are sufficiently suppressed, while the small ones are not. Hence, we are in the same situation as in the case above. Consequently, for source iteration, the contraction rate is

$$\omega_{\mathrm{SI}} = \max_{-\infty<\lambda<+\infty} \gamma(\lambda) = c,$$

as in the previous section.

For DSA we continue the calculation by a correction of the new iterate with an approximation of the residual

$$\phi^{(l+1)} = \phi^{(l+\frac{1}{2})} + G^{(l+\frac{1}{2})}.$$

We write the diffusion approximation for $G^{(l+\frac{1}{2})}$ as

$$DG^{(l+\frac{1}{2})} = \frac{c}{\langle s \rangle}(\phi^{l+\frac{1}{2}} - \phi^{(l)}).$$

Since $\phi^{(l+1)}(x) = \omega^{l+1}(\lambda)\mathrm{e}^{i\lambda x}$, we obtain

$$\begin{aligned}\rho(\beta) &= D^{-1}(\gamma(\lambda) - 1) = \frac{6c}{\langle s^2 \rangle \lambda^2 + 6(1-c)}(\gamma(\lambda) - 1),\\ \omega(\lambda) &= \gamma(\lambda) + \rho(\lambda).\end{aligned}$$

In detail this becomes

$$\begin{aligned}\omega(\lambda) &= \gamma(\lambda) + \frac{6c}{\langle s^2 \rangle \lambda^2 + 6(1-c)}(\gamma(\lambda) - 1)\\ &= \frac{6c}{\langle s^2 \rangle \lambda^2 + 6(1-c)}\left(\left(\frac{\langle s^2 \rangle \lambda^2}{6c} + \frac{1}{c}\right)\gamma(\lambda) - 1\right),\end{aligned}$$

with $\gamma(\lambda)$ depending on the choice of $p(s)$. For $\Sigma_t(s) \equiv \Sigma_t$ constant we have

$$\omega_{\mathrm{DSA}}(\lambda) = \frac{3\Sigma_t^2 c}{\lambda^2 + 3\Sigma_t^2(1-c)}\left(\left(\frac{\lambda^2}{3\Sigma_t^2} + 1\right)\frac{\Sigma_t^2}{\Sigma_t^2 + \frac{1}{3}\lambda^2} - 1\right) = 0,$$

for every $\lambda \in \mathbb{R}$ and every choice of $0 \leq c \leq 1$. This is consistent with the fact that in classical transport the P_1 model is equivalent to the diffusion approximation.

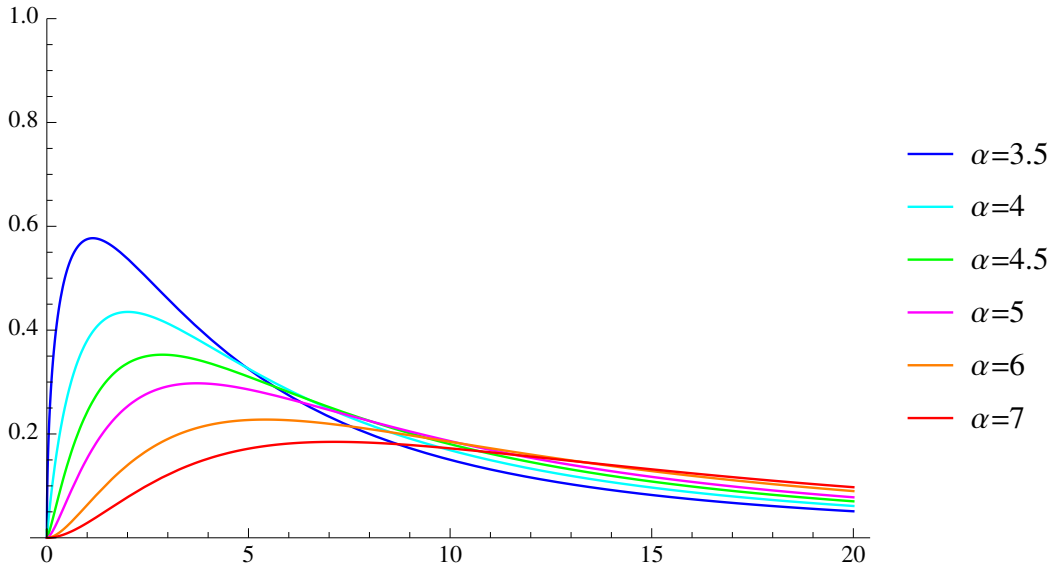


Figure 9.4: A comparison of ω as a function of λ for different exponents α and $c = 1$. Again we observe that the large modes are now sufficiently suppressed by DSA.

For a general path length distribution $p(s)$ we obtain

$$\omega_{\mathrm{DSA}}(\lambda) = \frac{6c}{\langle s^2\rangle\lambda^2 + 6(1-c)} \left(\left(\frac{\langle s^2\rangle\lambda^2}{6} + 1 \right) \int_0^\infty \cos\left(\frac{1}{\sqrt{3}}\lambda s \right) p(s)\,\mathrm{d}s - 1 \right).$$

Since the integral does not have a closed form solution, we compute the contraction rate again for path length distributions that decay algebraically. For our computations we choose again the exponents $\alpha = 3.5, 4, 4.5, 5, 6, 7$ and $c = 1$. The results are shown in Figure 9.4. Again we see that the small modes are now sufficiently suppressed. Furthermore, all values lie below the ones from Figure 9.2. Hence, we conclude that, also for moment models of the non-classical transport equation, the contraction rate is significantly decreased by DSA, especially for the small modes that cause slow convergence in the source iteration method.

10 | Numerical results

In this chapter we present numerical results for the non-classical steady state transport equation in slab geometry, obtained by the numerical methods developed in the previous Chapters 7-9. For the solutions of the non-classical transport equation we mostly present the scalar flux $\phi(x)$, which is calculated as

$$\phi(x) = \int_0^\infty \int_{-1}^1 \Psi(s, x, \mu) \, \mathrm{d}\mu \, \mathrm{d}s,$$

since the resulting 1D plots are best suited to visualize differences in the solutions and between different methods. For all numerical tests that we present isotropic scattering is assumed.

In the first Section 10.1 we present numerical results obtained by the HLL schemes for moment models of the non-classical transport equation. We compare P_1 and M_1 solutions for different non-classical cross sections, and investigate the performance of the schemes in the diffusive regime. Section 10.2 deals with qualitative properties of non-classical transport solutions. We show that the rate of decay of solutions towards the spatial boundaries is determined by the underlying path length distribution, and we present an example for non-classical transport in a special homogenized periodic medium. Finally, in Section 10.3 we present numerical results for the improvement of the contraction rate by the Diffusion Synthetic Acceleration method.

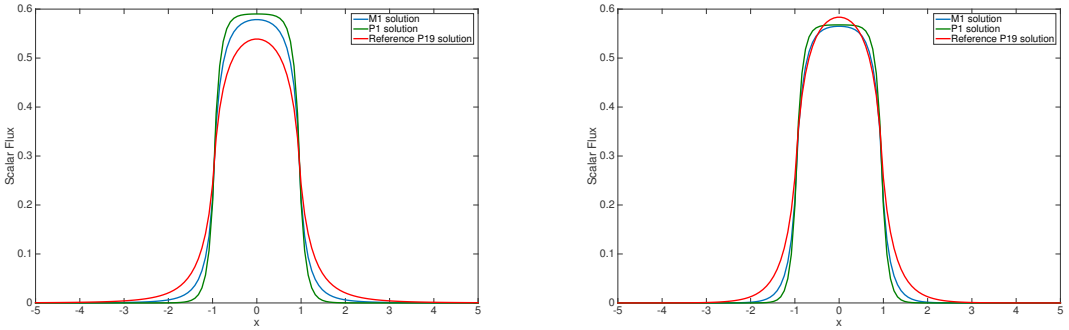
10.1 HLL schemes for moment models

In this section we present numerical results for the HLL finite volume schemes developed in Chapter 8. We compare the different moment models and present results that show the asymptotic preserving property in the diffusive regime.

Comparison of moment models

We present the results for the scalar flux $\phi(x)$ for the solution of the non-classical transport equation obtained by the M_1 and the P_1 moment model for two different choices of the non-classical cross section $\Sigma_t(s)$, and compare with a reference solution, which is a highly resolved P_{19} solution. Therefore, we choose a computational domain $x \in [-40, 40]$ and $s \in [0, 40]$. The spatial domain is discretized at 1000 points, and we choose a CFL number 1/3. To ensure a reasonably fast convergence of the source iteration, the scattering ratio is set to $c = 0.5$. The external source term is chosen to be $Q(x) = 1$

if $x \in [-1, 1]$, and $Q(x) = 0$ elsewhere. For the non-classical cross section $\Sigma_t(s)$ we choose an algebraically decaying function $\Sigma_t^1(s) = 4/(s+1)^5$, and a constant cross section $\Sigma_t^2(s) \equiv 3$, which corresponds to classical transport. This choice yields in both cases a mean free path of $\langle s \rangle = 1/3$.



(a) Algebraically decaying $\Sigma_t^1(s)$.

(b) Classical constant cross section $\Sigma_t^2(s)$.

Figure 10.1: Scalar fluxes for the different moment models and different choices for the non-classical cross section.

The numerical results for the scalar fluxes are shown in Figure 10.1. We observe that the M_1 and P_1 solution are relatively close in these scenarios. Compared with the reference solution, the scalar fluxes are overestimated for the algebraically decaying non-classical cross section (Figure 10.1a), and sligthly underestimated in the constant cross section case (Figure 10.1b). In both cases we see that the P_1 and the M_1 solution decay faster than the reference solution. Furthermore, the relative error (in the l^2-sense) to the reference solution is, in the M_1 case 0.1169 for Σ_t^1, and 0.0822 for Σ_t^2, while the P_1 solution yields 0.1870 for Σ_t^1, and 0.1244 for Σ_t^2. Hence, we conclude that the quality of the M_1 solution is better for both cases, and that both solutions are closer to the reference solution in the classical setting.

Different algebraically decaying cross sections

To study the qualitative effects of algebraically decaying non-classical cross section on both the M_1 and the P_1 moment model, we choose the same numerical setup as in the previous numerical test for the discretization of s, x, the scattering ratio c, and the external source term $Q(x)$. Then we perform calculations for a sequence of different algebraically decaying cross sections of the form

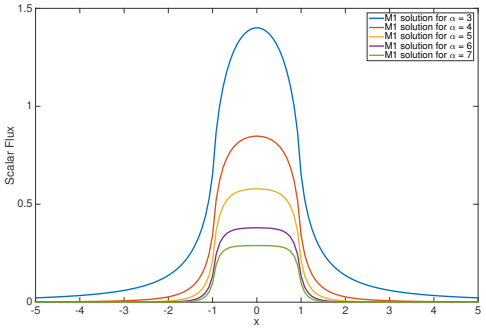
$$\Sigma_t(s) = \frac{\alpha - 1}{s + 1} \tag{10.1}$$

for $\alpha = 3, 4, 5, 6, 7$. These non-classical cross sections correspond to path length distributions of the form

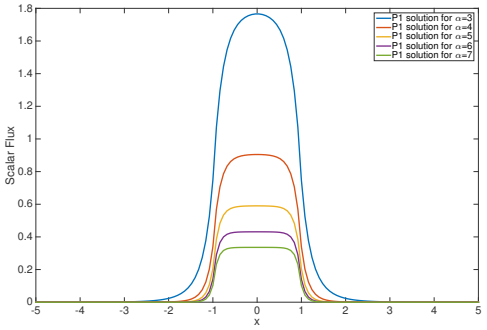
$$p(s) = \Sigma_t(s) \exp\left(- \int_0^s \Sigma_t(\tau)\, \mathrm{d}\tau \right) = \frac{\alpha - 1}{(s+1)^\alpha}.$$

Hence, α indicates the exponent of the algebraic decay of $p(s)$.

The resulting scalar fluxes are shown in Figure 10.2, the results for M_1 in Figure 10.2a, and the results for P_1 in Figure 10.2b.



(a) M_1 solutions for different algebraically decaying $\Sigma_t(s)$



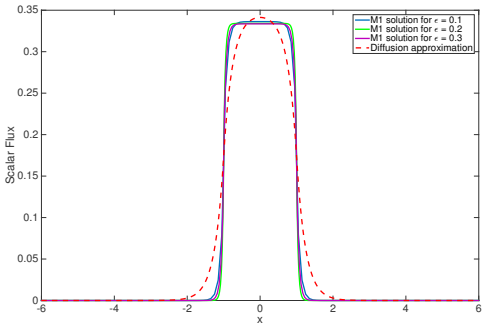
(b) P_1 solutions for different algebraically decaying $\Sigma_t(s)$.

Figure 10.2: Scalar fluxes for the M_1 and P_1 models for different choices of algebraically decaying non-classical cross sections.

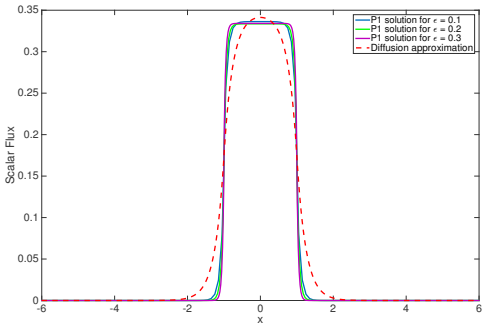
We clearly see that by increasing α the scalar flux is reduced, and becomes flatter around the origin in both cases. This is consistent with the interpretation of $\Sigma_t(s)$ as a collision rate. A higher value of α increases the probability of collisions for small path lengths, and since the source is centered around the origin the probability of absorption is increased in this region. Again we see that the decay of the scalar flux towards the boundary is faster in the P_1 case, compared to the M_1 case. And we also observe, that overall the P_1 solution exceeds the M_1 solution around the origin.

Asymptotic preserving schemes

We now present numerical results for the P_1 and M_1 moment model, and show that the HLL scheme recovers the correct diffusion limit in the diffusive regime. Therefore, we solve the scaled Equation (6.9), and compare results for $\varepsilon = 0.3, 0.2, 0.1$ with the corresponding diffusion approximation. We choose $\kappa = 1$, then the corresponding scattering ratio is $c = 1 - \kappa\varepsilon^2$, consequently we have $c = 0.91, 0.96, 0.99$. The non-classical cross section is chosen of the form (10.1) with $\alpha = 5$, and then scaled by ε. The diffusion approximation is calculated using a simple finite difference scheme. The computational domain depends on the scaling parameter ε: we have $x \in [-L, L]$ with $L = 30/\varepsilon$ and 2000 discrete points, and $s \in [0, T]$ with $T = 20\varepsilon$ and 3000 discrete points.
The results are shown in Figure 10.3. Comparison of the M_1 solutions together with the diffusion approximations are shown in Figure 10.3a, and a comparison of the P_1 solutions with diffusion approximation is shown in Figure 10.3b. We see that the diffusion approximation is approximated better and better for decreasing ε. The final error is much smaller than the expected $O(\varepsilon)$. As expected from Relation (8.3), there is only a very small difference between the P_1 and the M_1 solutions in the diffusive regime. Nevertheless, the deviation



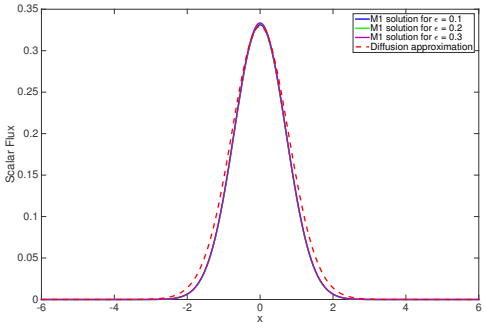
(a) M_1 solutions compared with diffusion approximation.



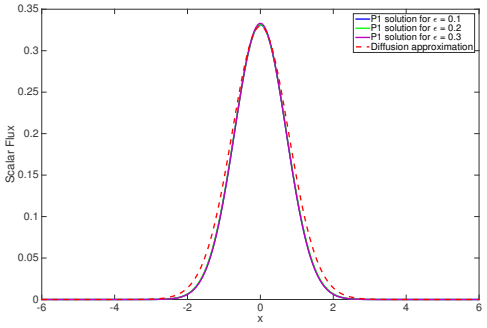
(b) P_1 solutions compared with diffusion approximation.

Figure 10.3: Scalar fluxes in the diffusive regime for the M_1 and P_1 moment model.

from the diffusion approximation is still essential, which is most likely caused by the discontinuity in the source term, that is strongly smoothed by the diffusion operator.
For a smooth source term $Q(x) = \exp(-x^2)$ we choose the same computational setting again, and consider $\varepsilon = 0.3, 0.2, 0.05$, where the computations for $\varepsilon = 0.05$ are done with 8000 discrete points in the x domain, and 5000 discrete points in the $s-$domain. The results are shown in Figure 10.4, the M_1 solutions in Figure 10.4a, and the P_1 solutions in Figure 10.4b.



(a) M_1 solutions compared with diffusion approximation.



(b) P_1 solutions compared with diffusion approximation.

Figure 10.4: Scalar fluxes in the diffuseive regime for the M_1 and P_1 moment model with exponential source term.

Here we observe a very close correspondence with the diffusion approximation, with errors again smaller than the expected $O(\varepsilon)$. We can therefore conclude that our numerical results show that the scheme shows the correct diffusion limit as $\varepsilon \to 0$, and that the correspondence with the diffusion approximation is closer if the external source term is smooth.

10.2 Qualitative properties of non-classical transport solutions

In this section we numerically investigate qualitative features of the non-classical transport solutions. We show that the tail of the corresponding path length distribution determines the decay of the scalar flux towards the spatial boundary. Furthermore, we present examples of non-classical cross sections and how they influence the shape of the solutions of the non-classical equation. This provides more intuition on the role of the non-classical cross section $\Sigma_t(s)$.

Decay towards the boundary

We compare the decay towards the spatial boundary of solutions of the non-classical transport equation with different underlying path length distributions. Therefore we choose a source term $Q(x) = 1$ if $-0.5 < x < 0.5$ and $Q(x) = 0$ else. That models particles being born in the center of an infinite domain, which then evolve towards the boundaries of the domain. For the transport calculations we now choose, on the one hand a classical exponential path length distribution $p_c(s) = 2\exp(-2s)$, and, on the other hand, path length distributions which decay algebraically as

$$p_{nc}(s) = \frac{C_\alpha \alpha}{(C_\alpha s + 1)^{\alpha+1}},$$

where C_α is chosen such that the mean free path is always $\langle s \rangle = 1/2$. The corresponding non-classical cross sections are $\Sigma_t(s) \equiv 2$ for the exponential case and

$$\Sigma_t(s) = \frac{C_\alpha(1+\alpha)}{1 + C_\alpha s}$$

for the algebraically decaying distributions. The non-classical equation is then solved numerically using the P_{11} system and an HLL scheme as presented in Chapter 8. We investigate the scalar flux, which is here calculated as

$$\phi(x) = \int_0^\infty \Psi^{(0)}(s, x)\, \mathrm{d}s.$$

The domains of the x and s-variable are chosen big enough to guarantee a sufficient decay. Furthermore, the scattering ratio is set to $c = 0.5$ and we assume isotropic scattering. Since we are only interested in the qualitative decay towards the boundary, all solutions are normalized in the L^∞-sense.

The resulting scalar fluxes are shown in Figure 10.5 on a logarithmic scale. While the classical solution that corresponds to a constant cross section shows a clear exponential decay towards the boundary, solutions with an underlying algebraic path length distribution decay significantly slower. In addition we note that the decay becomes faster as the exponent in the distribution func-

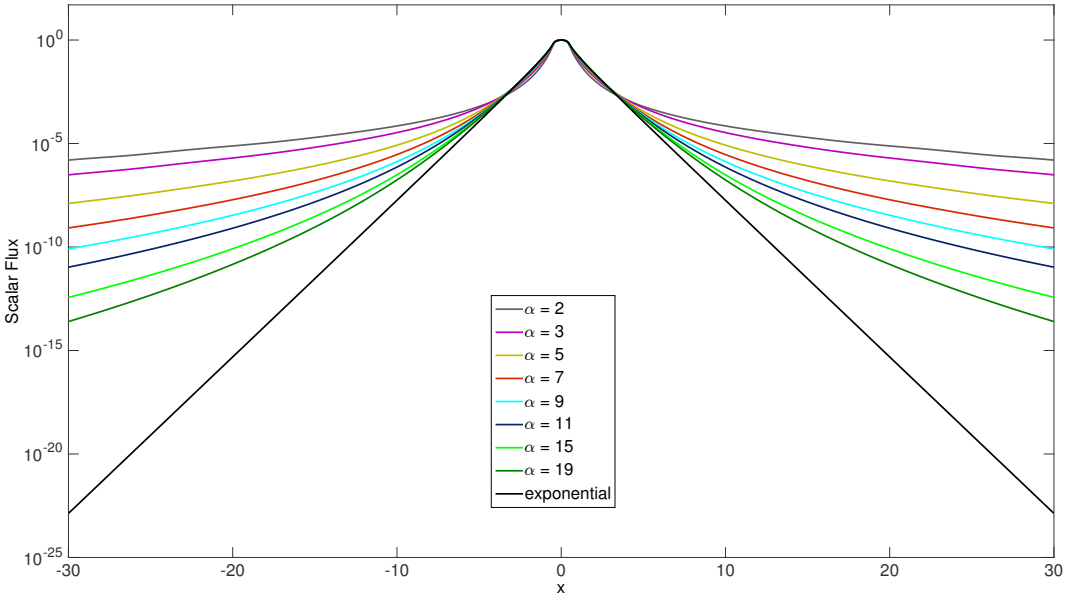


Figure 10.5: A comparison of non-classical solutions for different path length distributions with exponents α and scattering ratio $c = 0.5$. The exponential decay corresponds to a classical solution with $\Sigma_t \equiv 2$.

tion increases. In Figure 10.6 we quantify the decay of the scalar fluxes for exponents $\alpha = 3, 7, 11, 15$ on the domain $5 < x < 60$. The scalar fluxes are plotted on a log-log scale together with the corresponding slope. As it was to be expected, we find that the slope of the scalar fluxes equals the slope of the tail of underlying path length distribution p_{nc}. Because of the symmetry of the solution the same holds for the behavior of the scalar fluxes towards $-\infty$. We conclude that the qualitative decay of the scalar flux that results from the non-classical transport equation is determined by the underlying path length distribution. They show the same tail behavior in our numerical experiments.

Special shapes of the non-classical solution

This example is taken from the paper [104]. It nicely shows that non-classical transport equations can have additional qualitative properties compared to classical transport solutions. We consider a homogenized one dimensional periodic medium that consists of alternating solid and void layers, randomly placed in the infinite line $-\infty < x < \infty$. This means that we only know which material is present at any given point x in a probabilistic sense. Furthermore, we consider this problem in rod geometry; that is, particles are only allowed to travel in the horizontal directions $\mu = \pm 1$. We compare the results obtained by the non-classical transport equation and the results using a classical atomic mix approximation based on the classical transport equation. For a detailed investigation of this example we refer to [104], where the underlying path length distribution is calculated analytically. From this we can derive the non-classical cross section $\Sigma_t(s)$.

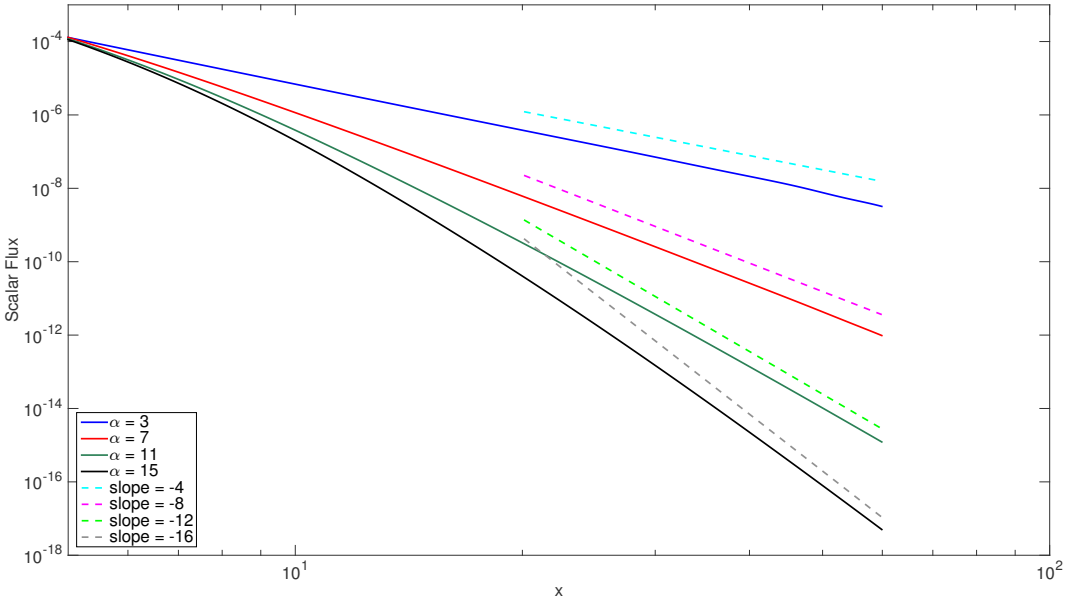
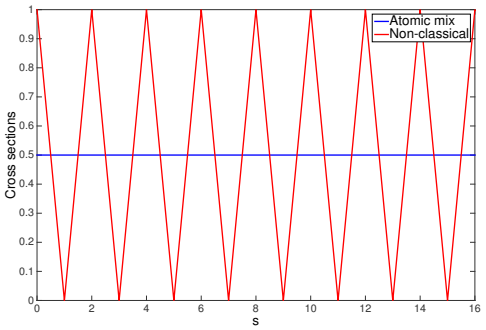
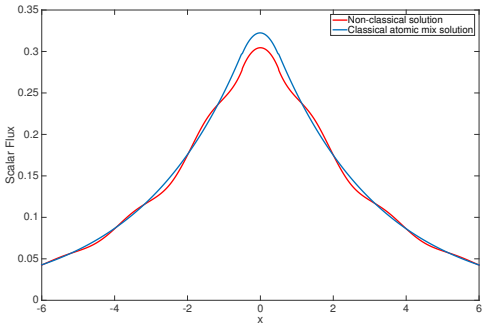


Figure 10.6: Non-classical solutions (on the positive x-interval) for different path length distributions exponents α and their slope on a log-log scale. The algebraic tail of the path length distribution determines the slope of the decay towards the boundary.

For the atomic mix strategy we simply volume-average the classical cross sections $\sigma_t^1 = 1$ and $\sigma_t^2 = 0$. This can also be modeled by non-classical transport with constant cross section $\Sigma_t(s) \equiv 0.5$. The s-dependent cross sections are shown in Figure 10.7a. Note that, over a suitable $s-$subinterval, both cross sections describe the same average collision rate.



(a) Classical atomic mix and corresponding non-classical cross sections $\Sigma_t(s)$.



(b) Scalar fluxes for the classical and non-classical homogenized cross section.

Figure 10.7: The cross sections and scalar fluxes for the homogenized periodic solid-void medium as described in [104].

We use again the HLL scheme to solve the non-classical transport equation, adapted to rod geometry, with 5120 points in the x domain and a CFL number 1/2. The scattering ratio is chosen as $c = 0.5$ and the external source term

is $Q(x) = 1/4$ for $-1/2 < x < 1/2$ and 0 elsewhere. The classical equation is obtained using the cross section $\Sigma_t(s) \equiv 1/2$ and the same computational parameters. We observe that the non-classical solution shows a sinoidal behavior, where the curvature changes at several points. The classical atomic mix solution is not able to reproduce this behavior, which is also shown by the true solution of this problem, as shown in [104]. Therefore, we conclude that non-classical transport solutions are capable of reproducing qualitatively more complex behavior than the classical transport equation.

10.3 Iterative methods

In this section, we present examples for the improvement of the speed of convergence by the Diffusion Synthetic Acceleration method compared to source iteration when the explicit HLL discretization as presented in Chapter 8 is used to solve the P_1 model of the non-classical transport equation. We compare the number of iterations and the maximal contraction rate for source iteration and DSA for different path length distributions $p(s)$. As in the previous chapter, we consider exponential path length distributions (Σ_t is constant) and algebraical distributions with exponents $\alpha = 3.5, 4, 4.5, 5, 6, 7$. For all cases we choose $\varepsilon = 0.1$ and a Gaussian source term $Q(x) = \exp(-0.05x^2)$. We use periodic boundary conditions and the domain of computation is chosen large enough to ensure a sufficient decline of the numerical solution; the number of grid points in the s-variable is 600 and 300 in the x-variable.

Iterations	Σ_t const.	$\alpha = 3.5$	$\alpha = 4$	$\alpha = 4.5$	$\alpha = 5$	$\alpha = 6$	$\alpha = 7$
SI	2330	2455	2525	2612	2720	3018	3483
DSA	11	22	15	14	15	20	26

Table 10.1: Number of iterations for the source iteration method (SI) and Diffusion Synthetic Acceleration (DSA) for different scenarios.

We see in Table 10.1 that the number of iterations decreases drastically. On average (over the presented scenarios), only 0.64% of the source iterations are needed using the Diffusion Synthetic Acceleration method.

ω	Σ_t const.	$\alpha = 3.5$	$\alpha = 4$	$\alpha = 4.5$	$\alpha = 5$	$\alpha = 6$	$\alpha = 7$
SI	0.99013	0.99067	0.99092	0.99126	0.9916	0.9925	0.99349
DSA	0.13397	0.42003	0.20192	0.11908	0.13618	0.21925	0.32289

Table 10.2: Contraction rates for the source iteration method (SI) and Diffusion Synthetic Acceleration (DSA) for different scenarios.

This is because of the significantly smaller contraction rates in the diffusion synthetic acceleration method. The contraction rate at the $(n+1)$-th iteration

is approximated by

$$\omega^{n+1} \approx \frac{\|\phi^{n+1} - \phi^n\|_2}{\|\phi^n - \phi^{n-1}\|_2}.$$

Table 10.2 shows the maximum contraction rate over all iterations. Note that the contraction rates predicted by the theory in Sections 9.3 and 9.4 are not attained consistently. That may be caused by different sources of error. Compared to the continuous equations there is an additional error introduced by the spatial and pseudo-time discretization. Furthermore, there is an error caused by the finite domain of computation, again in both independent variables.

Part III

Parameter Estimation

Introduction: Parameter Estimation

The main parameter of the non-classical transport equation is the $s-$dependent total cross section $\Sigma_t(s)$, which can be expressed in terms of the path length distribution $p(s)$ and vice versa. That means, the non-classical equation can be used to identify the path length distribution in practical applications. This is a typical so-called **inverse problem**: given measurement data and assuming that this data is the outcome of a process that can be described by a certain mathematical model, identify an optimal parameter that reproduces the data. We sketch this process in Figure 10.8. The reconstruction of model parameters from measurements is usually called **parameter estimation**. Mathematically, we can describe parameter estimation as an optimal control problem, where the constraints are given by the underlying model. In this part, we discuss mathematical and numerical techniques for parameter estimation based on the non-classical steady state transport equation.

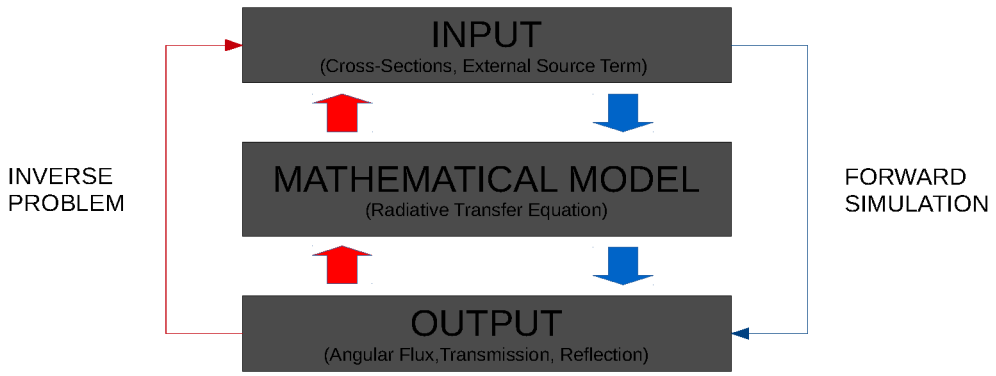


Figure 10.8: Inverse problem vs. forward simulation in the case of radiative transfer.

In the context of radiative transfer a large variety of inverse problems arise. In optical tomography [65, 64, 38] one is interested in the total cross section to retrieve information on the structure of the background medium. This is used to identify tumors or other irregularities in medical applications. The aim of optimal treatment planning for radiotherapy [92, 4, 5] is to find an optimal dose such that tumor cells are destroyed by the radiation while a maximal fraction of healthy tissue survives. Here, the dose is modeled by an external source term to the classical radiative transfer equation. Other applications in this field are neutron radiography [2, 95, 105], stellar identification of exo-planets [34, 6], or the optimization of semi-conductor devices [19, 77]. Nevertheless, the solution of inverse problems using kinetic equations is not common because of the high computational costs. Usually, very simplified mathematical models are used in this context.

A recent approach for the calculation of path length distributions in atmospheric clouds, as described in [48], is to neglect the effect of transport and

scattering, and concentrate on the absorption by the background medium only. Radiative transfer can then be described by the Beer-Lambert law. The transmission along a path of length s is then given by

$$T(s) = \exp(-s\sigma_t),$$

assuming that the total cross section σ_t is constant along the path. In a purely absorbing medium, the total cross section is also called the extinction coefficient, and is denoted by α. Then, the total transmission is obtained by integrating over all photon paths weighted with the total path length distribution p_{tot}. This distribution takes into account the full path length of a particle over all scattering events, in contrast to the distribution $p(s)$ of path lengths between collisions. The transmission for a given extinction coefficient α then reads

$$T(\alpha) = \int_0^\infty p_{tot}(s) \exp(-s\alpha)\, \mathrm{d}s = \mathcal{L}_{\alpha \leftarrow s} p_{tot}(s).$$

As indicated by the letter $\mathcal{L}$, this is the Laplace transform of the total path length distribution p_{tot} with respect to the extinction α. Hence, the problem of finding the function p_{tot} reduces to an inversion of the Laplace transform. Numerical methods for a Laplace inversion can be found in [9, 29]. Note that this is in principal an ill-posed problem.
The transmission $T(\alpha)$ with respect to the extinction coefficient can be measured by the so-called oxygen A-band spectroscopy. Important here is the fact that the extinction α depends on the frequency, while the total path length distribution is independent of frequency. Using this, many values for different $T(\alpha)$ with the same $p_{tot}(s)$ can be measured, from which the total path length distribution $p_{tot}(s)$ is then reconstructed by a numerical Laplace inversion.
Note that this method suffers from a clear modeling inconsistency. Although the underlying path length distribution is known to be non-exponential, the Beer-Lambert law, which does not hold in this case, is used to describe extinction along the path of a photon.
The connection between the distribution of total path lengths p_{tot} and the distribution of path lengths between collisions is established in [48] by a variable transformation and a shift of the distribution. That means, given $p(s)$, we could effectively calculate the distribution of total path lengths $p_{tot}(s)$, either analytically or using simulations.

The method that we propose in this part is based on the non-classical steady state transport equation. Finding the non-classical cross section Σ_t that governs the transport of photons through an atmospheric cloud, we can then directly calculate the corresponding distribution of path-lengths between collisions $p(s)$. In other words, given measurement data for certain physical quantities related to the angular radiative flux Ψ, we can rephrase the problem of finding the underlying photon path length distribution in the form of an optimal control problem, where the constraints are given by the non-classical equation. This yields a system of equations, the so-called first order optimality

system for this problem. We put our main focus on some simple test problems and focus on the numerical solution of these. Furthermore, we study important properties of the numerical schemes such as consistency of the continuous and the discrete optimality system, and their computational performance.

First, we formulate parameter estimation as an optimal control problem in Chapter 11. In addition to a realistic cost functional based on the transmittance, we also present simplified test problems. Furthermore, we introduce models for the non-classical parameter $\Sigma_t(s)$ in order to reduce the complexity of the optimal control problem. In Chapter 12 we present a formal analysis of the mathematical operators that are involved in this problem. We focus on the adjoints and Fréchet derivatives of these operators, and discuss the non-classical adjoint equation. In the following Chapter 13 we derive the first order optimality system of the optimal control problem that was introduced before, by following a Lagrange formalism. Numerical methods for this optimality system are the topic of Chapter 14. We apply the numerical schemes introduced in the previous part for the non-classical equation and its adjoint in order to solve the optimality system using gradient based methods, and present a number of numerical tests. In Chapter 15 we investigate the discrete form of the first order optimality system for the numerical schemes presented in the preceding chapter.

In all we present here, we always assume that both the parameter $\Sigma_t(s)$ and the solution Ψ of the non-classical equation are elements of a suitable Hilbert space. This is a reasonable assumption if the source term Q is bounded and has a certain regularity, and the cross section $\Sigma_t(s)$ decays fast enough as $s \to \infty$. The formal Hilbert space setting allows us to apply the techniques of the theory of PDE constrained optimization (see [102] for an extensive discussion).

11 Parameter estimation as an optimal control problem

The inverse problem we want to address here is to identify $\Sigma_t(s)$, or $p(s)$ respectively, from measured data. Recall that the classical (and also physical) angular flux Ψ_c can be obtained from the solution Ψ of the non-classical steady state transport equation as

$$\Psi_c(x,\Omega) = \int_0^\infty \Psi(s,x,\Omega)\,\mathrm{d}s.$$

As a possible field of application we consider the transport of photons in atmospheric clouds. The I3RC project [24] provides measurement data for radiative transfer through atmospheric clouds. Of special interest for us is that there exist measurements of the transmittance. That is the amount of radiation that leaves the cloud through the lower bounds (see Figure 11.1). That means the transmittance lives on the following subset of the classical phase space $G \times S^2$: spatially located at the lower boundary of our rectangular domain of computation $(\partial G)^l \subset \mathbb{R}^3$ with outer normal vector n, taking into account only outgoing directions

$$\Gamma^-(x) = \{\Omega \in S^{N-1} :\ \Omega \cdot n > 0\}.$$

Mathematically speaking, it can be written as a functional of the angular flux

$$T(x;\Psi) = \int_0^\infty \int_{\Gamma^-(x)} \Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}s, \quad x \in (\partial G)^l.$$

Assume that we are given measurement data $\bar{T}(x)$ for the transmittance. One way to find a parameter that reproduces this transmittance is to minimize the functional

$$J(\Psi,\Sigma_t) = \frac{A}{2}\int_{\partial G^l} (T(x;\Psi) - \bar{T}(x))^2\,\mathrm{d}x + \frac{B}{2}\,\mathrm{REG}(\Sigma_t), \tag{11.1}$$

with respect to Ψ and Σ_t, where REG is a regularization term that we discuss in the following section, and $A, B > 0$ are regularization parameters, under the condition

$$\begin{aligned}
&\partial_s \Psi(s,x,\Omega) + \Omega \cdot \nabla_x \Psi(s,x,\Omega) + \Sigma_t(s)\Psi(s,x,\Omega) = 0,\\
&\Psi(0,x,\Omega) = c\int_0^\infty \int \sigma(\Omega\cdot\Omega')\Sigma_t(s)\Psi(s,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s + Q(x,\Omega).
\end{aligned}$$

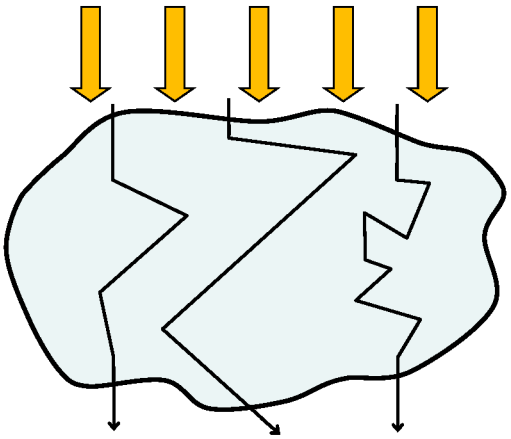


Figure 11.1: Sample paths of solar photons in an atmospheric cloud. The transmittance is the amount of radiation that leaves a cloud through the lower boundary.

This is an constrained optimization problem where the constraints are give in the form of an integro-differential equation. Formally, we also have the following additional constraints:

$$\Sigma_t \in L^2([0,\infty)),\ \Sigma_t \notin L^1([0,\infty)),\ \Sigma_t(s) \geq 0 \text{ a. e. on } [0,\infty).$$

The L^2 setting is of technical nature [102], the positivity constraint is standard, and the constraints on the L^1-norm prevent us from applying standard theory for existence and uniqueness of a solution.

Remark 11.1. Another possibility here is to use the transformed Form (6.9) of the non-classical equation and to formulate the optimal control problem in terms of the path length distribution $p(s)$. This leads to the additional constraints

$$p \in L^2([0,\infty)),\ p(s) \geq 0 \text{ a. e. on } [0,\infty), \|p\|_{L^1} = 1.$$

Especially the normalization condition is highly non-desirable, since the surface of the unit ball in L^1 is a set of measure zero. This is the reason why we only consider the optimal control problem in terms of the non-classical cross section $\Sigma_t(s)$.

11.1 Regularizations

Usually, optimal control problems have to be regularized in order to obtain well posedness [102]. Therefore, the cost functional is extended by an additional term $\mathrm{REG}(\Sigma_t)$ that depends on the control variable. The typical choice is a

Tikhonov regularization of the form

$$\mathrm{REG}(\Sigma_t) = \frac{B}{2}\|\Sigma_t(s) - \bar{\Sigma}_t(s)\|_2^2 = \frac{B}{2}\int_0^\infty (\Sigma_t(s) - \bar{\Sigma}_t(s))^2 \,\mathrm{d}s.$$

It can only be applied here if $\Sigma_t, \bar{\Sigma}_t \in L^2([0,\infty))$. We have seen that algebraically decreasing path length distributions lead to non-classical cross sections $\Sigma_t(s) \sim 1/s$, where this is satisfied. But this choice also excludes the case that Σ_t is constant, that means that we are in the case of classical transport. Instead of using the full cross section $\Sigma_t(s)$ as a control, we can also model the non-classical cross section by a function $\Sigma_t(s) = \Sigma_t(s; C)$, with a finite set of scalar parameters represented by a vector C. Then the parameters C of this model form a set of scalar controls. A finite dimensional set of controls also introduces well-posedness of the problem. For a finite set of N control parameters $C = (c_1, \ldots, c_N)$ the regularization then reads

$$\mathrm{REG}(\Sigma_t(C)) = \frac{B}{2}\|C - \bar{C}\|_2^2,$$

with appropriate $\bar{C}$.

11.2 Simplified cost functionals

The cost functional J has the disadvantage, that it is only defined on the spatial boundary of the computational domain. For test purposes we also introduce the following functionals:

$$J_1(\Psi, \Sigma_t) = \frac{A}{2}\int_0^\infty \int_{\mathbb{R}^3}\int_{S^2} \left(\Psi(s,x,\Omega) - \bar{\Psi}(s,x,\Omega)\right)^2 \mathrm{d}\Omega \,\mathrm{d}x \,\mathrm{d}s + \frac{B}{2}\,\mathrm{REG}(\Sigma_t),$$

and

$$J_2(\Psi, \Sigma_t) = \frac{A}{2}\int_{\mathbb{R}^3} (H(x;\Psi) - \bar{H}(x))^2 \,\mathrm{d}x + \frac{B}{2}\,\mathrm{REG}(\Sigma_t),$$

where

$$H(x;\Psi) = \int_0^\infty \int_{S^2} \Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}s.$$

Using these functionals, we eliminate boundary terms which appear later in the derivation of the optimality system. We use these functionals to verify the adjoint equation and derivatives of the functionals, and to test and compare the numerical methods. Note that these functionals are not realistic since they require the knowledge of the path length dependent flux. This data is not available in practice.

11.3 Models for the parameter

The introduction of finite dimensional models for the non-classical cross section helps to decrease the computational costs significantly and is therefore well suited for the verification of our method as well as for more complex calculations using real-world data. In the following, we present a one-parameter and two two-parameter models for the path length distribution, and derive the corresponding model of the cross section $\Sigma_t(s)$.

Algebraic decay
As a standard model for an algebraically decaying path length distribution we take again

$$p(s) = \frac{\alpha - 1}{(s+1)^\alpha}, \quad \alpha > 1,$$

which means

$$\Sigma_t(s) = \frac{\alpha - 1}{s+1}. \tag{11.2}$$

In this case α is our one dimensional control. This is well suited for simple numerical tests, but also includes the important case of an algebraically decaying tail.

Gamma distribution
Oftentimes, the path length distribution in atmospheric clouds is modeled as a gamma distribution. The gamma distribution is given by

$$p(s) = \frac{b^a}{\Gamma(a)} s^{a-1} \exp(-bs),$$

with Γ being the gamma function. The two free parameters $b, a > 0$ can be interpreted as

$$b = \frac{a}{\langle s \rangle}, \quad \text{and} \quad a = \frac{\langle s \rangle^2}{\mathrm{var}(s)}.$$

Then the $s-$dependent cross section is calculated as

$$\Sigma_t(s) = \frac{b^a s^{a-1} \exp(-bs)}{\Gamma(a, bs)},$$

where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function, defined by

$$\Gamma(a, bs) = \int_{bs}^{\infty} t^{a-1} \exp(-t)\, \mathrm{d}t.$$

The set of controls in this case are the two parameters a and b.

Log-normal distribution

Modeling the path length distribution $p(s)$ with a log-normal distribution yields

$$p(s) = \frac{1}{As\sqrt{2\pi}} \exp\left(-\frac{(\log(s) - B)^2}{2A^2}\right)$$

with free parameters $A > 0$ and $B \in \mathbb{R}$. Then we have for the non-classical cross section

$$\Sigma_t(s) = \frac{\exp\left(-\frac{(\log(s)-B)^2}{2A^2}\right)}{As\sqrt{2\pi}\left(\frac{1}{2} + \frac{1}{2}\operatorname{erf}(\frac{B-\log(s)}{\sqrt{2}A})\right)},$$

where $\operatorname{erf}(s)$ is the well known error function.

12 Adjoint calculus

In the following we analyze the operators which show up in this optimal control problem. We focus on the derivative and the adjoint of these operators. As mentioned in the introduction of this part, we do all the analysis formally in suitable L^2-spaces. This preliminary analysis is necessary in order to apply the Lagrange formalism [102] later on, where both the derivatives and the adjoints of the mathematical operators are needed. An exact analysis would be based on operators in suitable Banach spaces, and therefore much more complex. We rely on formal computations for the sake of simplicity. Nevertheless, we obtain the correct optimality conditions from our analysis (see also the comment in Section 2.10 in [102]). For the classical transport equation, a rigorous L^p theory exists for the parameter estimation of the total and scattering cross section [38].

We introduce the following operators

$$\begin{aligned} V\Psi(s,x,\Omega) &= (\partial_s + \Omega\nabla_x + \Sigma_t(s))\Psi(s,x,\Omega), \\ K\Psi(s,x,\Omega) &= c\int_0^\infty \int \sigma(\Omega,\Omega')\Sigma_t(s')\Psi(s',x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s', \end{aligned}$$

such that we can write Equation (6.4) in the following form

$$V\Psi(s,x,\Omega) = 0, \quad \Psi(0,x,\Omega) = K\Psi(s,x,\Omega) + Q(x,\Omega).$$

We denote an operator A acting on a function space X by writing the image of an arbitrary element $f \in X$. That means $(Af)(x)$ is the image of f under A as a function of x. In analogy to the finite dimensional case, the Fréchet derivative of an operator A acting on a function space X is a linear operator for each element $f \in X$, that acts on a (potentially different) space Y. We write this operator as A_f or $\partial_f A$. We also denote the Fréchet derivative of an operator A by writing the image of this linear operator at an arbitrary point $f \in X$ evaluated at an element $y \in Y$. Consequently, $(A_f f)(y)(x)$ is the image of y as a function of x under the derivative of A with respect to f at $f \in X$. The adjoint operator A^* of A is characterized by the property that

$$\langle Af, y\rangle = \langle f, A^* y\rangle \quad \text{for all } f \in X,\ y \in Y,$$

where $\langle\cdot\rangle$ denotes the scalar product

$$\langle f, g\rangle := \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} f(s,x,\Omega)\cdot g(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s,$$

which is the standard inner product in the Hilbert space $L^2([0,\infty) \times \mathbb{R}^3 \times S^2)$. We only calculate the derivatives and adjoint operators that are actually needed in the remainder of this work.

12.1 Control-to-state operator

In contrast to linear optimal control problems such as [45] or [46], where the external source term is used as a control variable, the control-to-state or solution operator

$$\mathcal{E}(\Sigma_t) = \Psi, \ \Psi \text{ satisfies (6.4)},$$

is neither linear nor affine linear in our setting. Therefore, we cannot calculate the derivative or the adjoint of this operator directly. This also restricts ourselves from introducing the reduced cost functional $J^*(\Sigma_t) = J(\Psi(\Sigma_t), \Sigma_t)$, since this requires an, at least formal, inversion of the control-to-state operator. This difficulty underlines the necessity to use the Lagrange formalism and an adjoint method.

12.2 Cost functionals

The operator T (here called the transmittance operator) is given by

$$(T\Psi)(x) = \int_0^\infty \int_{\Gamma^-} \Psi(s, x, \Omega) \, \mathrm{d}\Omega \, \mathrm{d}s.$$

This operator T is linear and bounded, hence its derivative with respect to Ψ is

$$(T_\Psi \bar{\Psi})(f)(x) = \int_0^\infty \int_{\Gamma^-} f(s, x, \Omega) \, \mathrm{d}\Omega \, \mathrm{d}s.$$

The residual term in the cost functional is also written in the operator form as

$$\int_{\partial\Omega^l} (T(x;\Psi) - \bar{T}(x))^2 \, \mathrm{d}x =: D(ST - \bar{T})\Psi.$$

where the operator D is the L^2-norm restricted to the lower spatial boundary:

$$Dt = \int_{(\partial G)^l} (t(x))^2 \, \mathrm{d}x,$$

and the operator S is the (partial) trace operator

$$(ST)(x) = T_{|(\partial G)^l}(x).$$

The trace operator S is linear and bounded while D is a bounded non-linear mapping, both being Fréchet differentiable. The derivative of the trace oper-

ator is

$$(S_T\bar{T})(V)(x) = V_{|(\partial G)^l}(x),$$

and the derivative of the operator D is given as

$$(D_t\bar{t})h = \int_{(\partial G)^l} 2\bar{t}(x)h(x)\,\mathrm{d}x.$$

Therefore, we obtain the derivative of the quadratic residual of the transmittance with respect to Ψ as

$$\begin{aligned} D(ST-\bar{T})_\Psi(\bar{\Psi})\Psi &= 2\int_{(\partial G)^l}(ST(x;\bar{\Psi})-\bar{T}(x))ST(x;\Psi)\,\mathrm{d}x \\ &= 2\int_{(\partial G)^l}\int_0^\infty\int_{\Gamma^-}(T(x;\bar{\Psi})-\bar{T}(x))S\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}s\,\mathrm{d}x. \end{aligned}$$

Remark 12.1. Note that the cost functional (11.1) can be written as

$$J(\Psi,\Sigma_t) = \frac{A}{2}D(ST-\bar{T}) + \frac{B}{2}\,\mathrm{REG}(\Sigma_t).$$

Since the second term is independent of Ψ, we consequently have

$$\partial_\Psi J(\bar{\Psi},\bar{\Sigma}_t)\Psi = A\int_{(\partial G)^l}\int_0^\infty\int_{\Gamma^-}(T(x;\bar{\Psi})-\bar{T}(x))S\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}s\,\mathrm{d}x.$$

12.3 Scattering-source operator

The operator K that models the scattering may be interpreted in different ways in our setting. We denote the different interpretations by introducing indices K^i, $i = 1,2,3$. It can be interpreted as an operator mapping the state variable Ψ to the initial function with respect to s for a given fixed cross section $\Sigma_t(s)$. Then it reads

$$(K^1\Psi)(x,\Omega) = c\int_0^\infty\int\sigma(\Omega\cdot\Omega')\Sigma_t(s)\Psi(s,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s.$$

Here, K_1 is a linear and bounded mapping with respect to Ψ, hence its derivative with respect to Ψ is given by

$$(K^1_\Psi\bar{\Psi})(h)(x,\Omega) = c\int_0^\infty\int\sigma(\Omega\cdot\Omega')\Sigma_t(s)h(s,x,\Omega)\,\mathrm{d}\Omega'\,\mathrm{d}s.$$

For a given fixed angular flux $\Psi(s,x,\Omega)$ the operator K^2 maps the non-classical cross section Σ_t to the scattering source term

$$(K^2\Sigma_t)(x,\Omega) = c\int_0^\infty \int \sigma(\Omega\cdot\Omega')\Sigma_t(s)\Psi(s,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s.$$

The operator K^2 is also linear and bounded, hence its derivative reads

$$(K^2_{\Sigma_t}\bar{\Sigma_t})(g)(x,\Omega) = c\int_0^\infty \int \sigma(\Omega\cdot\Omega')g(s)\Psi(s,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}s.$$

We need both derivatives in the following chapter to derive the first order optimality system. The adjoint operator of interest here is the adjoint of the extended operator K^3 given by

$$(K^3\Psi)(s,x,\Omega) = \delta(s)(K^1\Psi)(x,\Omega),$$

with symmetric scattering kernel σ. To calculate the adjoint we formally test with a function $\lambda \in L^2([0,\infty)\times\Omega\times S^2)$. Taking the scalar product yields:

$$\begin{aligned}
&\langle K^3\Psi,\lambda\rangle \\
&\quad= \int_0^\infty\int_G\int_{S^2} \lambda(s,x,\Omega)\delta(s)\left[c\int_0^\infty\int_{S^2}\sigma(\Omega\cdot\Omega')\Sigma_t(\tau)\Psi(\tau,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}\tau\right]\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \\
&\quad= \int_0^\infty\int_G\int_{S^2} \Psi(\tau,x,\Omega)\Sigma_t(\tau)\left[c\int_0^\infty\int_{S^2}\sigma(\Omega'\cdot\Omega)\delta(s)\lambda(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}s\right]\mathrm{d}\Omega'\,\mathrm{d}x\,\mathrm{d}\tau \\
&\quad= \int_0^\infty\int_G\int_{S^2} \Psi(\tau,x,\Omega)\Sigma_t(\tau)\left[c\int_{S^2}\sigma(\Omega'\cdot\Omega)\lambda(0,x,\Omega)\,\mathrm{d}\Omega\right]\mathrm{d}\Omega'\,\mathrm{d}x\,\mathrm{d}\tau \\
&\quad= \langle\Psi,(K^3)^*\lambda\rangle,
\end{aligned}$$

since we can interchange the roles of τ and s and Ω and Ω'. Hence, this operator is not self-adjoint.

12.4 Adjoint equation

In order to calculate the adjoint equation of Equation (6.4), we consider the operator

$$(V\Psi)(s,x,\Omega) = \partial_s\Psi(s,x,\Omega) + \Omega\nabla_x\Psi(s,x,\Omega) + \Sigma_t(s)\Psi(s,x,\Omega).$$

We write this equation in a weak form by multiplying a test function $\lambda \in L^2([0,\infty)\times\mathbb{R}^3\times S^2)$ and taking the scalar product. Additionally we assume that

$$\begin{aligned}
\lim_{s\to\infty}\lambda(s,x,\Omega) &= \lim_{s\to\infty}\Psi(s,x,\Omega) = 0 \quad \text{for all } x\in\mathbb{R}^3,\ \Omega\in S^2, \\
\lim_{|x|\to\infty}\lambda(s,x,\Omega) &= \lim_{|x|\to\infty}\Psi(s,x,\Omega) = 0 \quad \text{for all } s\in[0,\infty),\ \Omega\in S^2.
\end{aligned}$$

Partial integration then yields a representation of the adjoint equation. We have

$$\begin{aligned}\langle V\Psi,\lambda\rangle &= \int_0^\infty \int_G \int_{S^2} \lambda(s,x,\Omega)(V\Psi)(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \\ &= \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \lambda(s,x,\Omega)\partial_s\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \qquad \text{(A1)} \\ &\quad + \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \lambda(s,x,\Omega)\Omega\nabla_x\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \qquad \text{(A2)} \\ &\quad + \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \lambda(s,x,\Omega)\Sigma_t(s)\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s. \qquad \text{(A3)}\end{aligned}$$

Partially integrating the first Term (A1) yields

$$\begin{aligned}\int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \lambda(s,x,\Omega)\partial_s\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s = \\ \int_{\mathbb{R}^3}\int_{S^2}\Big[\lambda(s,x,\Omega)\Psi(s,x,\Omega)\Big]_{s=0}^{s=\infty}\mathrm{d}\Omega\,\mathrm{d}x \\ -\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\partial_s\lambda(s,x,\Omega)\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s.\end{aligned}$$

We note that, formally, $\Psi(\infty,x,\Omega)=\lambda(\infty,x,\Omega)=0$. We insert the initial condition for Ψ with respect to s into the remaining term to obtain

$$\begin{aligned}&-\int_{\mathbb{R}^3}\int_{S^2}\lambda(0,x,\Omega)\Psi(0,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x \\ &\quad = -\int_{\mathbb{R}^3}\int_{S^2}\lambda(0,x,\Omega)\big[c\int_0^\infty\int_{S^2}\sigma(\Omega\cdot\Omega')\Sigma_t(s)\Psi(s,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}\tau\big]\,\mathrm{d}\Omega\,\mathrm{d}x \\ &\quad = -\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\Psi(s,x,\Omega)\Sigma_t(s)\big[c\int_{S^2}\sigma(\Omega'\cdot\Omega)\lambda(0,x,\Omega)\,\mathrm{d}\Omega\big]\,\mathrm{d}\Omega'\,\mathrm{d}x\,\mathrm{d}s,\end{aligned}$$

by following the argumentation of the previous section. For the treatment of the second Term (A2) we use the divergence theorem. That yields

$$\begin{aligned}&\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\lambda(s,x,\Omega)\Omega\nabla_x\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \\ &\qquad = -\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\Omega\nabla_x\lambda(s,x,\Omega)\Psi(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s,\end{aligned}$$

since we consider an infinite spatial domain. The third Term (A3) is a simple multiplication of λ and Ψ with Σ_t where no more calculations are necessary.

Collecting all terms and factoring out the state variable Ψ gives

$$\begin{aligned}\langle V\Psi, \lambda\rangle = \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \Psi(s,x,\Omega) \Big[&- \partial_s\lambda(s,x,\Omega) - \Omega\nabla_x\lambda(s,x,\Omega) \\ &+ \Sigma_t(s)\lambda(s,x,\Omega) - c\Sigma_t(s)\int_{S^2}\sigma(\Omega\cdot\Omega')\lambda(0,x,\Omega')\,\mathrm{d}\Omega'\Big]\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \\ = \langle \Psi, V^*\lambda\rangle.\end{aligned}$$

Hence, the non-classical adjoint equation is given by $V^*\lambda = 0$ and reads

$$\begin{aligned} &- \partial_s\lambda(s,x,\Omega) - \Omega\nabla_x\lambda(s,x,\Omega) + \Sigma_t(s)\lambda(s,x,\Omega) && \text{(12.1a)} \\ &\qquad = c\Sigma_t(s)\int_{S^2}\sigma(\Omega\cdot\Omega')\lambda(0,x,\Omega')\,\mathrm{d}\Omega'. && \text{(12.1b)}\end{aligned}$$

Because of the negative sign of the pseudo-time derivative it needs to be solved backwards in s, and has to be equipped with initial condition

$$\lambda(\infty,x,\Omega) = 0 \qquad \text{for all } x\in\mathbb{R}^3,\quad \Omega\in S^2,$$

because of the choice of λ.

Source terms

One equation in the optimality system, that is derived in the following chapter, is the adjoint equation together with an additional source term. Both for the classical and the non-classical transport equation, existence of a solution also depend on the form of the source term. For example, if the source term is constant in time (or pseudo-time), the solution of the forward equation may blow up, since there are constantly particles inserted into the system. Therefore, it is necessary to discuss this issue for the adjoint equation with an additional source term, because it has to be solved over an infinite pseudo-time interval. In general, the source term Q of the adjoint equation depends on x, Ω and s and the non-classical adjoint equation then reads

$$\begin{aligned}&- \partial_s\lambda(s,x,\Omega) - \Omega\nabla_x\lambda(s,x,\Omega) + \Sigma_t(s)\lambda(s,x,\Omega) \\ &\qquad = c\Sigma_t(s)\int_{S^2}\sigma(\Omega\cdot\Omega')\lambda(0,x,\Omega')\,\mathrm{d}\Omega' + Q(s,x,\Omega).\end{aligned}$$

In order to discuss the boundedness of a solution for a source term that is constant in pseudo-time, we first look at the problem of a purely absorbing background medium ($c = 0$) with a constant source term $Q(s,x,\Omega) \equiv Q$. Then, for particles that are released in one specific direction we have $x = x(s) = s$, and therefore $\lambda(s,x) = \lambda(s,x(s)) = \lambda(s)$. Consequently, this can be described by the reduced ordinary differential equation

$$-\partial_s\lambda(s) + \Sigma_t(s)\lambda(s) = Q. \tag{12.2}$$

The solution of the homogeneous problem is easily calculated as

$$\lambda_h(s) = C \exp\left(-\int_s^\infty \Sigma_t(\tau)\,\mathrm{d}\tau\right).$$

Hence, the general solution of Equation (12.2) is given by

$$\lambda(s) = \lambda_h(s) + \exp\left(-\int_s^\infty \Sigma_t(\tau)\,\mathrm{d}\tau\right)\int_s^\infty Q \exp\left(\int_{s'}^\infty \Sigma_t(\tau)\,\mathrm{d}\tau\right)\,\mathrm{d}s',$$

which is obviously bounded for all $s \in [0, \infty)$. The reintroduction of the spatial transport term and the scattering operator (with a scattering ratio $0 < c < 1$) does not change this observation because they are both conservative. Therefore, we can conclude that the non-classical adjoint equation with source terms that emit particles constantly over pseudo time has a bounded solution.

13 Continuous first order optimality system

In this chapter, we derive the first order optimality system of the optimal control problem formulated in Chapter 11. Therefore, we follow the Lagrange formalism as introduced in [102], Section 2.10. Treating the state and control in the Lagrangian as independent variables, the system of equations for a (local) optimum is obtained by differentiating the Lagrangian with respect to state, Lagrange multiplier and control, and setting all partial derivatives to zero. The resulting coupled system of equations can then be solved iteratively, or using gradient based optimization methods.

13.1 Lagrange formalism

Following the Langrange formalism, the corresponding Lagrangian for the optimal control problem as introduced in Chapter 11 is given by

$$L(\Psi, \Sigma_t, \lambda) = J(\Psi, \Sigma_t) - \langle V\Psi, \lambda \rangle,$$

where J is the cost functional, $\langle \cdot \rangle$ is the scalar product in $L^2(\mathbb{R}_+ \times \mathbb{R}^3 \times S^2)$, and V the operator introduced in the previous chapter. Here, λ is the so-called Lagrange multiplier. It can also be interpreted as a test function of the right dimensionality that is sufficiently smooth. In the following we refer to it also as the adjoint variable. A formal integration by parts yields

$$\begin{aligned} L(\Psi, \Sigma_t, \lambda) &= J(\Psi, \Sigma_t) - \langle V\Psi, \lambda \rangle = J(\Psi, \Sigma_t) - \langle \Psi, V^*\lambda \rangle \\ &= J(\Psi, \Sigma_t) \\ &\quad - \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \Psi(s,x,\Omega) \Big[-\partial_s \lambda(s,x,\Omega) - \Omega \nabla_x \lambda(s,x,\Omega) \\ &\quad + \Sigma_t(s)\lambda(s,x,\Omega) - c\Sigma_t(s) \int_{S^2} \sigma(\Omega \cdot \Omega')\lambda(0,x,\Omega') \, \mathrm{d}\Omega' \Big] \, \mathrm{d}\Omega \, \mathrm{d}x \, \mathrm{d}s, \end{aligned}$$

where we have used the derivation of the adjoint operators from the previous chapter. Since Ψ is supposed to be a solution of the non-classical transport equation we have $V\Psi = 0$, hence $\langle V\Psi, \lambda \rangle = \langle \Psi, V^*\lambda \rangle = 0$. Then it is easy to see that a necessary condition for a local minimum of L at $(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t)$ is

$$\partial_\Psi L(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t) = 0, \quad \partial_{\Sigma_t} L(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t) = 0 \quad \text{and} \quad \partial_\lambda L(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t) = 0.$$

These three conditions form the so-called optimality system. They are also a necessary condition for a local minimum of J at $(\bar{\Psi}, \bar{\Sigma}_t)$ under the constraint $V\Psi = 0$.

13.2 Forward, adjoint and update equation

The necessary conditions for a local optimum yield a coupled system of three equations. They are usually called the forward, the adjoint, and the update equation. In the following we derive these equations for the optimal control problem as described above.

13.2.1 Forward equation

We obtain the forward equation by formally differentiating the Lagrangian with respect to the adjoint variable or Lagrange multiplier λ. We have

$$\partial_\lambda L(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t)\lambda = \langle V\Psi, \lambda\rangle = 0,$$

for all λ. That means that $V\Psi = 0$ holds in a weak sense. Consequently, the forward equation is simply the weak formulation of Equation (6.4) together with the corresponding initial conditions:

$$\partial_s \Psi(s, x, \Omega) + \Omega\nabla_x \Psi(s, x, \Omega) + \Sigma_t(s)\Psi(s, x, \Omega) = 0, \tag{13.1a}$$

$$\Psi(0, x, \Omega) = c\int_0^\infty \int_{S^2} \sigma(\Omega, \Omega')\Sigma_t(s)\Psi(s, x, \Omega)\,\mathrm{d}\Omega\,\mathrm{d}s + Q(x, \Omega). \tag{13.1b}$$

13.2.2 Adjoint equation

We obtain the adjoint equation by formally differentiating the Lagrangian with respect to the state variable Ψ. From

$$\partial_\Psi L(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t)\Psi = \partial_\Psi J(\Psi, \Sigma_t)\Psi - \langle \Psi, V^*\lambda\rangle = 0,$$

it follows that

$$\langle \Psi, V^*\lambda\rangle = \partial_\Psi J(\Psi, \Sigma_t)\Psi, \tag{13.2}$$

for all Ψ. This is a weak formulation of the non-classical adjoint Equation (12.1), where the derivative of the cost functional with respect to the state variable serves as an additional source-term. We recall that

$$\partial_\Psi J(\bar{\Psi}, \bar{\Sigma}_t)\Psi = A\int_{(\partial G)^l} \int_0^\infty \int_{\Gamma^-} (T(x; \bar{\Psi}) - \bar{T}(x))S\Psi(s, x, \Omega)\,\mathrm{d}\Omega\,\mathrm{d}s\,\mathrm{d}x,$$

where S is the trace operator that restricts Ψ to the boundary. Hence, we can write the corresponding adjoint Equation (13.2) as

$$\begin{aligned} -\partial_s \lambda(s,x,\Omega) - \Omega \nabla_x \lambda(s,x,\Omega) + \Sigma_t(s)\lambda(s,x,\Omega) \\ = c\Sigma_t(s) \int_{S^2} \sigma(\Omega,\Omega')\lambda(0,x,\Omega')\,\mathrm{d}\Omega' + R(x), \end{aligned}$$

where

$$R(x) = A(T(x;\bar{\Psi}) - \bar{T}(x))$$

is non-zero only on the spatial boundary ∂G^l and reemits particles isotropically over the full pseudo-time domain. This equation is again equipped with the initial condition $\lambda(\infty, x, \Omega) = 0$ for all $x \in \mathbb{R}^3$ and $\Omega \in S^2$.

13.2.3 Update equation

We obtain the update equation by formally differentiating the Lagrangian with respect to the control variable Σ_t. This can also be interpreted as the gradient of the reduced cost functional

$$\bar{J}(\Sigma_t) = J(\Psi(\Sigma_t), \Sigma_t),$$

with respect to the control variable. Here it makes a difference if the control variable is the full function $\Sigma_t(s)$ (infinite dimensional control), or if we use a model for Σ_t that can be represented by a finite set of scalar controls (finite dimensional control). We derive the update equation for both cases, including the models for Σ_t presented in Section 11.3.

Infinite dimensional control

The derivative of the Lagrangian with respect to Σ_t at the local minimum $(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t)$ is given by

$$\begin{aligned} \partial_{\Sigma_t} L(\bar{\Psi}, \bar{\lambda}, \bar{\Sigma}_t)\Sigma_t = &- \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \Sigma_t(s)\bar{\Psi}(s,x,\Omega)\bar{\lambda}(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \\ &+ \int_0^\infty \int_{\mathbb{R}^3} \int_{S^2} \Sigma_t(s)\bar{\Psi}(s,x,\Omega) c \int_{S^2} \sigma(\Omega,\Omega')\bar{\lambda}(0,x,\Omega')\,\mathrm{d}\Omega'\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \\ &+ B \int_0^\infty \Sigma_t(s)\bar{\Sigma}_t(s)\,\mathrm{d}s, \end{aligned}$$

for the standard Tykhonov regularization. The function Σ_t is used as a test function, hence we can interpret the above as a gradient

$$\partial_{\Sigma_t} L(s) = -\int_\Omega \int_{S^2} \Psi(s,x,\Omega)\lambda(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x \tag{13.3a}$$

$$+\int_{\mathbb{R}^3}\int_{S^2} \Psi(s,x,\Omega)\left[c\int_{S^2}\sigma(\Omega,\Omega')\lambda(0,x,\Omega')\,\mathrm{d}\Omega'\right]\mathrm{d}\Omega\,\mathrm{d}x + B\Sigma_t(s). \tag{13.3b}$$

Finite dimensional control

Using models for the the non-classical cross section $\Sigma_t(s)$ as introduced in Section 11.3, the corresponding Lagrangian and hence the forward and adjoint equation are the same as above, only the update equation, respectively the gradient, changes. Assume that $\Sigma_t(s)$ can be modeled by a finite set of N control parameters $c_1, \dots c_N$, and that $\Sigma_t(s)$ depends on each control c_i in a differentiable way. That means, we assume that the gradient $\partial_{c_i}\Sigma_t$ exists for all i. Then the gradient with respect to each control parameter c_i, $i \in \{1,\dots,N\}$, is given by

$$\partial_{c_i} L(\bar\Psi,\bar\lambda,\bar c_i)c_i = -\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2} \left[\partial_{c_i}\Sigma_t(s)(\bar c_i)\right] c_i\bar\Psi(s,x,\Omega)\bar\lambda(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \tag{13.4a}$$

$$+\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2} \left[\partial_{c_i}\Sigma_t(s)(\bar c_i)\right] c_i\bar\Psi(s,x,\Omega)\Big[c\int_{S^2}\sigma(\Omega,\Omega')\bar\lambda(0,x,\Omega')\,\mathrm{d}\Omega'\Big]\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s \tag{13.4b}$$

$$+ Bc_i\bar c_i. \tag{13.4c}$$

For the simple case of an algebraically decaying function with α as a one-dimensional control

$$\Sigma_t(s) = \frac{\alpha-1}{s+1}, \quad \alpha \geq 3,$$

we obtain the update equation as

$$\partial_\Psi L(\bar\Psi,\bar\lambda,\bar\alpha)\alpha = -\alpha\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\frac{1}{s+1}\bar\Psi(s,x,\Omega)\bar\lambda(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s$$

$$+\alpha\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\frac{1}{s+1}\bar\Psi(s,x,\Omega)\Big[c\int_{S^2}\sigma(\Omega,\Omega')\bar\lambda(0,x,\Omega')\,\mathrm{d}\Omega'\Big]\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s + B\alpha\bar\alpha.$$

Here, the control α can be factored out, which yields the gradient

$$\partial_\alpha L = -\int_0^\infty\int_{R^3}\int_{S^2}\frac{1}{s+1}\Psi(s,x,\Omega)\lambda(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s$$

$$+\int_0^\infty\int_{\mathbb{R}^3}\int_{S^2}\frac{1}{s+1}\Psi(s,x,\Omega)\Big[c\int_{S^2}\sigma(\Omega,\Omega')\lambda(0,x,\Omega')\,\mathrm{d}\Omega'\Big]\,\mathrm{d}\Omega\,\mathrm{d}x\,\mathrm{d}s\Big) + B\alpha.$$

If the underlying path length distribution is modeled by a gamma or lognormal distribution (as in Section 11.3), the derivatives with respect to the parameters a, b and A, B exist, but in contrast to the above presented case, they are nonlinear and more complex. The derivative of Σ_t for the gamma distribution can be expressed in terms of the incomplete gamma function and Meijer G-functions [8], the one for the lognormal distribution in terms of the error function. We do not state them here explicitly, but they can be easily calculated using a computer algebra system such as Mathematica by Wolfram Research. Alternatively, the derivatives with respect to the parameters can be calculated approximately using finite differences.

13.2.4 Optimality system for simplified test problems

Using simplified cost functionals as introduced in Section 11.2, the first order optimality system only differs in the form of the source term of the adjoint equation. Consequently, the forward equation is again Equation (13.1), while the corresponding adjoint equation is given by

$$\begin{aligned} -\partial_s \lambda(s,x,\Omega) - \Omega \cdot \nabla_x \lambda(s,x,\Omega)\,\mathrm{d}\tau + \Sigma_t(s)\lambda(s,x,\Omega) = \\ \Sigma_t(s) c \int_{S^2} \sigma(\Omega',\Omega)\lambda(0,x,\Omega')\,\mathrm{d}\Omega' + R(s,x,\Omega), \end{aligned}$$

with residual term $R(s,x,\Omega) = A(\Psi(s,x,\Omega) - \bar{\Psi}(s,x,\Omega))$, if we consider the simplified cost functional J_1, and $R(s,x,\Omega) = R(x) = A(H(x;\Psi) - \bar{H}(x))$, if the cost functional J_2 is used. The gradient, derived from the update equation is given by Equation (13.3) for the infinite dimensional control, and the equations for the gradient in the case of finite dimensional control also remain valid.
The main advantage here is that source terms are simple and defined everywhere. This is why we focus on the numerical solutions of the simpler optimality systems in the following chapter in order to test the numerical methods.

13.3 Gradient based solution methods

We can make use of the fact that the derivative of the Lagrangian yields the gradient of the reduced cost functional

$$\partial_{\Sigma_t} L(\Psi, \lambda, \Sigma_t) = \partial_{\Sigma_t} \bar{J}(\Sigma_t),$$

where the explicit form of this gradient is given by Equation (13.3). Hence, we use gradient based optimization techniques to minimize the reduced cost functional, see [102] for an overview. That means, we can solve the optimality system by a gradient descent method, that can be formulated in the following way:

(0) Start with an initial choice $\Sigma_t^0(s)$. Then proceed as follows.

(1) Solve the forward equation for Ψ^n using Σ_t^n.

(2) Solve the adjoint equation for λ^n using Σ_t^n and Ψ^n.

(3) Update the non-classical cross section by

$$\Sigma_t^{n+1}(s) = \Sigma_t^n - h^n \partial_{\Sigma_t} L(\Psi^n, \lambda^n, \Sigma_t^n),$$

and return to (1). Repeat until this is converged.

The positivity constraint for Σ_t can be included by an additional projection in the update step (3). The update then reads

$$\Sigma_t^{n+1}(s) = \mathrm{proj}_{\geq 0}\Big(\Sigma_t^n(s) + h^n \int_{\mathbb{R}^3}\int_{S^2} \Psi^n(s,x,\Omega)\lambda^n(s,x,\Omega)\,\mathrm{d}\Omega\,\mathrm{d}x \tag{13.5a}$$

$$- h^n \int_{\mathbb{R}^3}\int_{S^2} \Psi^n(s,x,\Omega)\Big[c\int_{S^2}\sigma(\Omega\cdot\Omega')\lambda^n(0,x,\Omega')\,\mathrm{d}\Omega'\Big]\,\mathrm{d}\Omega\,\mathrm{d}x + B\Sigma_t^n(s)\Big), \tag{13.5b}$$

where $\mathrm{proj}_{\geq 0}$ denotes the pointwise projection to the positive half-space. The step size $h^n > 0$ has to be chosen appropriately, this is discussed in the following chapter.

Note also, that the gradient depends on both the state and the adjoint variable. This is in contrast to the standard optimal control problems discussed in [102, 59]. Knowledge of the gradient also enables the use of more advanced gradient based methods such as the BFGS quasi-Newton method [17, 42, 50, 97], where the inverse of the Hessian is approximated using the gradient information (see also [90] for quasi-Newton methods in PDE constrained optimization). Another advantage of this method is, that it is implemented very effectively in MATLAB, and therefore it can be used without extra effort once the gradient is calculated. If the positivity constraints have to be taken into account, we can also use SQP (sequential quadratic programming) methods [90], which are again implemented efficiently in MATLAB.

14 Numerical methods for the first order optimality system

In this chapter we use a discrete ordinate S_N method or the spherical harmonics P_N expansion, together with an HLL scheme as introduced in Chapters 7 and 8, to solve the first order optimality system numerically. We restrict ourselves to the 1D slab geometry case with isotropic scattering to simplify the presentation. Equation (6.4) then reads

$$\partial_s \Psi(s,x,\mu) + \mu \partial_x \Psi(s,x,\mu) + \Sigma_t(s)\Psi(s,x,\mu) = 0 \tag{14.1a}$$

$$\Psi(0,x,\mu) = \frac{c}{2}\int_0^\infty \int_{-1}^1 \Sigma_t(s')\Psi(s',x,\mu)\,\mathrm{d}\mu\,\mathrm{d}s' + Q(x,\mu), \tag{14.1b}$$

with $\mu \in [-1,1]$ being the cosine of the polar angle. In the same way the slab geometry version of the adjoint equation is given by

$$-\partial_s \lambda(s,x,\mu) - \mu\partial_x \lambda(s,x,\mu) + \Sigma_t(s)\lambda(s,x,\mu) = \tag{14.2a}$$

$$\Sigma_t(s)\frac{c}{2}\int_{-1}^1 \lambda(0,x,\mu)\,\mathrm{d}\mu + R(s,x,\mu), \tag{14.2b}$$

where R is the residual and serves as a source term for the adjoint equation.

We first discuss the discretization of the optimality system by the S_N-HLL method in Section 14.1, and by the P_N-HLL method in Section 14.2. In the following Section 14.3 we describe line search strategies for gradient based solution methods. Finally, in Section 14.4 we present numerical tests that illustrate that the here developed method can be used to solve parameter estimation problem numerically with a good accuracy at a reasonable computational cost.

14.1 S_N-HLL discretization

It is straight forward to use a S_N approximate model for the discretization of the angular variable and the HLL-scheme as described in Chapter 8 for the discretization of the first order optimality system. Recall that for the discrete ordinate approximation, the angular variable is discretized at N points $\mu_i,\ i = 1,..,N$, such that

$$\sum_{i=1}^N \omega_i \Psi(s,x,\mu_i) \approx \int_{-1}^1 \Psi(s,x,\mu)\,\mathrm{d}\mu,$$

where μ_i are usually chosen to be Gauss points with corresponding weights ω_i. So in slab geometry, we obtain the following approximate system of equations for Equation (14.1):

$$\partial_s \Psi_i(s,x) + \mu_i \partial_x \Psi_i(s,x) + \Sigma_t(s)\Psi_i(s,x) = 0,$$
$$\Psi_i(0,x) = \frac{c}{2}\int_0^\infty \sum_{i=1}^N \omega_i \Sigma_t(s')\Psi_i(s',x)\,\mathrm{d}s' + Q_i(x),$$

where $\Psi_i(s,x) = \Psi(s,x,\mu_i)$ and $Q_i(x) = Q(x,\mu_i)$. The adjoint Equation (14.2) is approximated by

$$-\partial_s \lambda_i(s,x) - \mu_i \partial_x \lambda_i(s,x) + \Sigma_t(s)\lambda_i(s,x) = \Sigma_t(s)\frac{c}{2}\sum_{i=1}^N \omega_i \lambda_i(0,x) + R_i(s,x),$$

in the case of isotropic scattering, where $\lambda_i(s,x) = \lambda(s,x,\mu_i)$ and $R_i(s,x) = R(s,x,\mu_i)$.

For each ordinate μ_i we can adapt the HLL scheme presented in Chapter 8. Setting $\Psi_i^{n,j} = \Psi(s^n, x_j, \mu_i)$, $\Sigma_t^n = \Sigma_t(s^n)$, and $b = \max|\mu_i|$ yields, in the isotropic case, the following explicit scheme

$$\frac{\Psi_i^{j,n+1} - \Psi_i^{j,n}}{\Delta s} + \mu_i \frac{\Psi_i^{j+1,n} - \Psi_i^{j-1,n}}{2\Delta x} - b\frac{\Psi_i^{j+1,n} - 2\Psi_i^{j,n} + \Psi_i^{j-1,n}}{2\Delta x} + \Sigma_t^n \Psi_i^{j,n} = 0,$$
$$\Psi_i^{j,0} = \frac{c}{2}\sum_{n=0}^\infty \sum_{i=1}^N \alpha_n \Sigma_t^n \omega_i \Psi_i^{j,n} + Q_i^j.$$

Note that, in contrast to the scheme in Chapter 8, this scheme is not necessarily asymptotic preserving. In the same fashion, an HLL scheme for the adjoint equation can be derived. This runs backwards in (pseudo-)time and reads

$$-\frac{\lambda_i^{j,n} - \lambda_i^{j,n+1}}{\Delta s} - \mu_i \frac{\lambda_i^{j+1,n+1} - \lambda_i^{j-1,n+1}}{2\Delta x} - b\frac{\lambda_i^{j+1,n+1} - 2\lambda_i^{j,n+1} + \lambda_i^{j-1,n+1}}{2\Delta x} + \Sigma_t^{n+1}\lambda_i^{j,n+1} = \frac{c}{2}\Sigma_t^{n+1}\sum_{i=1}^N \omega_i \lambda_i^{j,0} + R_i^{j,n+1},$$

where $R_i^{j,n+1} = R(s^n, x_j, \mu_i)$. Both equations can either be solved in a source-iteration like manner or directly as one linear system. If N is large or the spatial discretization is fine, the source-iteration should be preferred. Note also, that the sign of the numerical diffusion is the same for both equations. That plays a role in the proof of the consistency of the scheme with the discrete optimality system.
The discretization of the update equation is straightforward in this case since

the pointwise discretization of the angular variable directly yields

$$[\partial_{\Sigma_t} J](s) \approx -\int_{\mathbb{R}} \sum_{i=1}^{N} \omega_i \Psi_i(s,x)\lambda_i(s,x)\,\mathrm{d}x + \int_{\mathbb{R}} \sum_{i=1}^{N} \omega_i \Psi_i(s,x) \left[\frac{c}{2}\sum_{i'=1}^{N} \omega_{i'}\lambda_i(0,x)\right] \mathrm{d}x.$$

The integration in x needs to be carried out numerically.

14.2 P_N-HLL discretization

The P_N closure for the non-classical equation was introduced in Chapter 7. Let $P_k(\mu)$ be the $k-$th Legendre polynomial, then we define the $k-$th Legendre moment of the forward and the adjoint variable as

$$\Psi^{(k)}(s,x) := \int_{-1}^{1} P_k(\mu)\Psi(s,x,\mu)\,\mathrm{d}\mu, \quad \lambda^{(k)}(s,x) := \int_{-1}^{1} P_k(\mu)\lambda(s,x,\mu)\,\mathrm{d}\mu.$$

The P_N system that approximates Equation (14.1) with isotropic scattering then reads

$$\partial_s \Psi^{(k)}(s,x) + \partial_x\left[\frac{k}{2k+1}\Psi^{(k-1)}(s,x) + \frac{k+1}{2k+1}\Psi^{(k+1)}(s,x)\right] + \Sigma_t(s)\Psi^{(k)}(s,x) = 0,$$
$$\Psi^{(0)}(0,x) = c\int_0^\infty \Sigma_t(s)\Psi^{(0)}(s,x)\,\mathrm{d}s + Q(x), \qquad \Psi^{(k)}(0,x) = 0, \quad \text{for } k = 1,\dots,N.$$

Introducing $\Psi_k^{j,n} = \Psi^{(k)}(s^n, x_j)$, the explicit HLL scheme reads

$$\frac{\Psi_k^{j,n+1} - \Psi_k^{j,n}}{\Delta s} + \frac{k}{2k+1}\cdot\frac{\Psi_{k-1}^{j+1,n} - \Psi_{k-1}^{j-1,n}}{2\Delta x} + \frac{k+1}{2k+1}\cdot\frac{\Psi_{k+1}^{j+1,n} - \Psi_{k+1}^{j-1,n}}{2\Delta x}$$
$$- \frac{\Psi_k^{j+1,n} - 2\Psi_k^{j,n} + \Psi_k^{j-1,n}}{2\Delta x} + \Sigma_t^n \Psi_k^{j,n} = 0,$$
$$\Psi_0^{j,0} = c\sum_{n=0}^{\infty} \omega_n \Sigma_t^n \Psi_0^{j,n} + Q^j, \qquad \Psi_k^{j,0} = 0, \quad \text{for } k = 1,\dots,N,$$

for some infinite quadrature rule for the s-integration given by the weights ω_n. The P_N system for Equation (14.2) with isotropic scattering reads

$$-\partial_s \lambda^{(k)}(s,x) - \partial_x\left[\frac{k}{2k+1}\lambda^{(k-1)}(s,x) + \frac{k+1}{2k+1}\lambda^{(k+1)}(s,x)\right]$$
$$+ \Sigma_t(s)\lambda^{(k)}(s,x) = c\Sigma_t(s)\lambda^{(k)}(0,x)\delta(k=0) + R^{(k)}(s,x),$$
$$\lambda^{(k)}(\infty,x) = 0, \quad \text{for all } k = 0,\dots,N,$$

where $R^{(k)}(s,x)$ is the k-th Legendre moment of the residual term R. Setting $\lambda_k^{j,n} = \lambda^{(k)}(s^n, x_j)$ and $R_k^{j,n+1} = R^{(k)}(s^n, x_j)$, the HLL scheme for this system

of equations is given by

$$
\begin{aligned}
&-\frac{\lambda_k^{j,n+1}-\lambda_k^{j,n}}{\Delta s}-\frac{k}{2k+1}\cdot\frac{\lambda_{k-1}^{j+1,n+1}-\lambda_{k-1}^{j-1,n+1}}{2\Delta x}-\frac{k+1}{2k+1}\cdot\frac{\lambda_{k+1}^{j+1,n+1}-\lambda_{k+1}^{j-1,n+1}}{2\Delta x}\\
&-\frac{\lambda_k^{j+1,n+1}-2\lambda_k^{j,n+1}+\lambda_k^{j-1,n+1}}{2\Delta x}+\Sigma_t^{n+1}\lambda_k^{j,n+1}=c\Sigma_t^{n+1}\lambda_k^{j,0}\delta_{k=0}+R_k^{j,n+1},\\
&\lambda_k^{j,\infty}=0,\quad \text{for all } k=0,\dots,N.
\end{aligned}
$$

Note that this system needs to be solved backwards in pseudo-time, that explains the, at least on the first look, implicit discretization of this variable.

The discretization of the update equation and the cost functional requires a reconstruction of the forward and the adjoint variable because of the non-linear nature of Equation (13.3) and J. Inserting the reconstructions and integrating with respect to μ, the gradient simplifies to

$$
\left[\partial_{\Sigma_t}J\right](s)\approx-\int_{\mathbb{R}}\sum_{k=0}^{N}\frac{2k+1}{2}\Psi^{(k)}(s,x)\lambda^{(k)}(s,x)\,\mathrm{d}x+\frac{c}{2}\int_{\mathbb{R}}\Psi^{(0)}(s,x)\lambda^{(0)}(0,x)\,\mathrm{d}x.
$$

If a regularization is used, an additional term appears that is independent of Ψ and λ. Inserting the reconstructions of Ψ and the desired state $\overline{\Psi}$ into the first term of the cost functional yields

$$
\begin{aligned}
\|\Psi-\overline{\Psi}\|_{L^2}^2&=\int_0^\infty\int_{\mathbb{R}}\int_{-1}^{1}(\Psi(s,x,\mu)-\overline{\Psi}(s,x,\mu))^2\,\mathrm{d}\mu\,\mathrm{d}x\,\mathrm{d}s\\
&\approx\int_0^\infty\int_{\mathbb{R}}\int_{-1}^{1}\left[\sum_{k=0}^{N}\frac{2k+1}{2}P_k(\mu)\left(\Psi^{(k)}(s,x)-\overline{\Psi}^{(k)}(s,x)\right)\right]^2\mathrm{d}\mu\,\mathrm{d}x\,\mathrm{d}s\\
&=\int_0^\infty\int_{\mathbb{R}}\sum_{k=0}^{N}\frac{2k+1}{2}\left(\Psi^{(k)}(s,x)-\overline{\Psi}^{(k)}(s,x)\right)^2\mathrm{d}x\,\mathrm{d}s.
\end{aligned}
$$

In both cases, the integration in x needs to be carried out numerically, while the integrals with respect to μ could be solved analytically.

14.3 Line search

As described in Section 13.3, we solve the first order optimality system using gradient based optimization methods. The derivative of the reduced cost functional $\bar{J}(\Sigma_t)$ with respect to the non-classical cross section Σ_t equals the corresponding derivative of the Lagrangian with respect to Σ_t. Hence, we can use a gradient descent method to update the parameter after each step, that means we set

$$
\Sigma_t^{n+1}=\Sigma_t^n+h_nv_n,\tag{14.3}
$$

where v_n is the anti gradient given by

$$v_n = \int_{\mathbb{R}} \int_{-1}^{1} \bar{\Psi}(s,x,\mu)\bar{\lambda}(s,x,\mu)\,\mathrm{d}\mu\,\mathrm{d}x \\ - \int_{\mathbb{R}} \int_{-1}^{1} \bar{\Psi}(s,x,\mu)\left[\frac{c}{2}\int_{-1}^{1} \bar{\lambda}(s,x,\mu')\,\mathrm{d}\mu'\right]\,\mathrm{d}\mu\,\mathrm{d}x - B\Sigma_t^n,$$

and the step size h_n is determined by the following line search method.

(0) Choose an initial step size h_0.

(1) For a given step size h_k, calculate Σ_t^{n+1} by (14.3) and the corresponding value of the reduced cost functional $\bar{J}(\Sigma_t^{n+1})$.

(2) If $\bar{J}(\Sigma_t^{n+1}) < \bar{J}(\Sigma_t^n)$, the update is accepted, and we proceed with step (0). If $\bar{J}(\Sigma_t^{n+1}) \geq \bar{J}(\Sigma_t^n)$, we set $h_{k+1} = h_k/2$ and return to step (1).

(3) If the relative error in $\bar{J}$ falls below a predefined threshold, we stop the iteration.

In order to guarantee positivity of the cross section in every step, Σ_t^{n+1} is pointwise projected to the positive half space if it becomes negative. The above presented line search method is the simplest possible method. More sophisticated methods that take into account the Armijo condition or Wolfe's conditions [59, 90] can also be applied in our setting.

14.4 Numerical tests

In this section, we present the results of several numerical test cases, where we study and test the above introduced numerical methods to solve the first order optimality system. First, we present a verification of our method, showing that the calculated gradient agrees with the one obtained from a finite difference approximation. Then we calibrate the regularization parameter B and investigate the effect of the scattering ratio c. In the following, we present testcases where parametrized and full cross sections are reconstructed from perturbed data for the non-classical scalar flux.

14.4.1 Verification and parameter testing

The methods to solve the first order optimality system all rely on a calculation of the gradient of the cost functional with respect to the non-classical cross section. Therefore, we have to make sure that the numerical methods are able to calculate the true gradient of this quantity. Furthermore, we study the regularization parameters for the Tikhonov regularization.

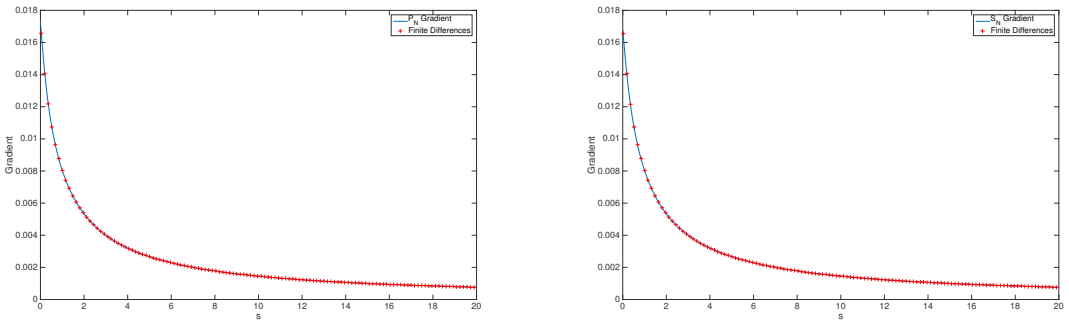
Verification of the gradient

We verify the calculation of the gradient given by Equation (13.3) by comparing the numerical solution to a finite difference approximation of the gradient.

We use the simplified cost functional J_1 as introduced in Section 11.2, with parameters $A = 1$ and $B = 0.1$. The cost functional is evaluated at Ψ and the desired state $\overline{\Psi}$, which correspond to solutions of the non-classical transport equation with underlying path length distributions p, and $\overline{p}$ respectively, given by

$$p(s) = \frac{4}{(s+1)^5}, \qquad \overline{p}(s) = \frac{3}{(s+1)^4},$$

and corresponding non-classical cross sections $\Sigma_t(s)$ and $\bar{\Sigma}_t(s)$.



(a) Gradient by P_N (b) Gradient by S_N

Figure 14.1: Comparison of the gradient calculated by the analytic formulation derived in the previous chapter, and the corresponding finite difference approximation, using the P_N method (left), or the S_N method (right).

The finite difference approximation to the gradient is calculated using a central difference scheme, where the discretized non-classical cross section $\Sigma_t(s)$ is perturbed pointwise

$$\partial J_1^i \approx \frac{J_1^+ - J_1^-}{2\Delta\Sigma_t^i},$$

with J_1^+ being the value of the cost functional for $\tilde{\Sigma}_t^i = \Sigma_t^i + \Delta\Sigma_t^i$, and J_1^- being the value of the cost functional for $\tilde{\Sigma}_t^i = \Sigma_t^i - \Delta\Sigma_t^i$. The remaining entries are left unperturbed. The perturbation is chosen relative to the current value $\Delta\Sigma_t^i = 10^{-4} \cdot \Sigma_t^i$.
We see in Figure 14.1 that the gradient agrees very well with the finite difference approximation at every entry for the P_N method, as well as for the S_N method. The relative L^2-error is $5.02 \cdot 10^{-4}$ for the P_N-method, and $8.81 \cdot 10^{-3}$ for the S_N method. Furthermore, we note that the gradients for the P_N and S_N method also agree very well with a relative L^2-error of $1.07 \cdot 10^{-2}$. Consequently, for the following test cases it is sufficient to only consider one of these methods.

Regularization parameters

We study the effect of the regularization parameter B in the Tikhonov regularization

$$\mathrm{REG}(\Sigma_t) = B \int_0^\infty \left(\Sigma_t(s) - \bar{\Sigma}_t(s)\right)^2 \mathrm{d}s$$

for the simplified cost functionals J_1 and J_2. Therefore, we calculate a forward solution with non-classical cross section and underlying path length distribution

$$p(s) = \frac{3}{(s+1)^4}, \text{ hence } \Sigma_t(s) = \frac{3}{s+1},$$

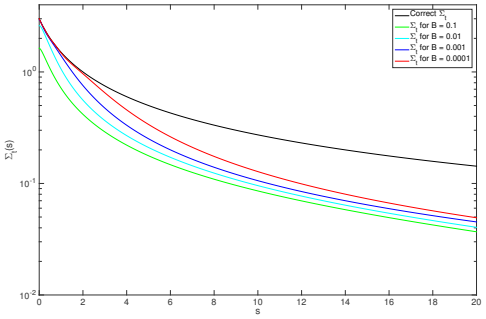
and then solve the first order optimality system in order to reconstruct the cross section using a different algebraically decaying initial guess Σ_t^0. We expect that the quality of the solution increases as B decreases. The first regularization parameter is chosen as $A = 1$ for all cases. We also compare the quality of the approximation of the underlying cumulative distribution function

$$q(s) = \exp\left(-\int_0^s \Sigma_t(\tau)\,\mathrm{d}\tau\right).$$

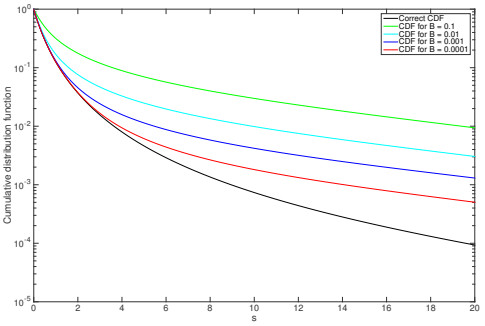
For all calculations we use a P_5-HLL scheme on the spatial interval $x \in [-20, 20]$ with Dirichlet boundary conditions, $\Delta x = 0.08$ and $\Delta s = 0.04$, where the s-variable is cut off at $s_{\max} = 40$. We choose a Gaussian source term $Q(x)$ and the scattering ratio is fixed to $c = 0.5$. For the regularization parameter we choose the values $B = 0.1, 0.01, 0.001, 0.0001$.
Figure 14.2 shows the reconstructions of $\Sigma_t(s)$ and the corresponding cumulative distribution functions (CDF) for the different choices of the regularization parameter B. The results for the cost functional J_1 are shown in the top Figures 14.2a and 14.2b, the results for J_2 are shown in the bottom Figures 14.2c and 14.2d. We see that the reconstructions approach the correct cross section very good for small s as the parameter B decreases. For larger s this is not that clear in Figures 14.2a and 14.2c. For the cumulative distribution functions (Figures 14.2b and 14.2d), on the other hand, the correct solution is approximated increasingly better as B decreases. Furthermore, the reconstructions from the integrated cost functional J_2 are visibly better than those for J_1. Especially the reconstruction of the CDF for $B = 0.0001$ from J_2 is very good. In the following we therefore rely on a regularization parameter $B = 0.0001$, and we prefer the simplified cost functional J_2 for further calculations. Note that the integrated cost functional J_2 is closer to realistic applications, because of the independence of the angular and the path length variable.

Effect of the scattering ratio c

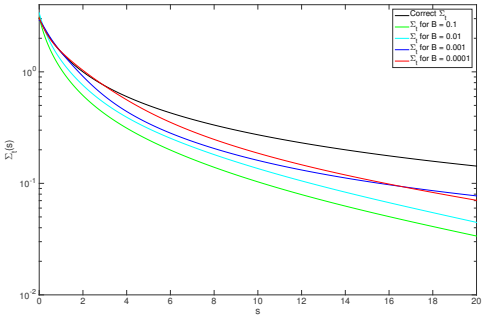
Here, we study the effect of the scattering ratio c on the quality of the reconstructions. Therefore, we consider the same computational setup as above, only



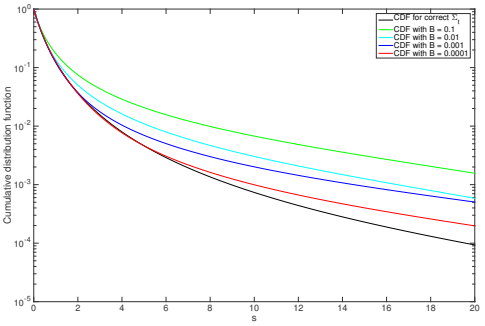
(a) Σ_t for cost functional J_1, and different choices of B.



(b) CDF for cost functional J_1, and different choices of B.



(c) Σ_t for cost functional J_2, and different choices of B.



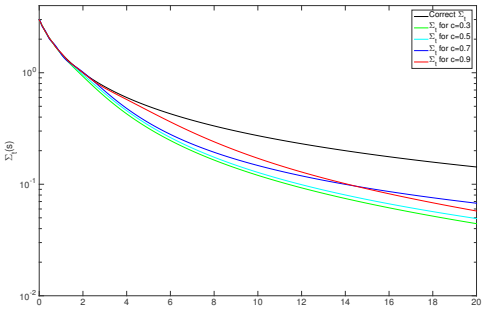
(d) CDF for cost functional J_2, and different choices of B.

Figure 14.2: Comparison of the calculated and the exact non-classical cross section Σ_t (left) and the corresponding cumulative distribution functions (right) for the cost functional J_1 (top) and J_2 (bottom) for varying regularization parameter B.

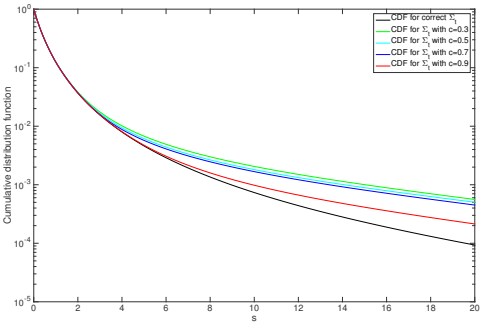
the scattering ratio c varies, and we consider the values $c = 0.3, 0.5, 0.7, 0.9$. We again compare the reconstructions of the cross section and the underlying cumulative distribution functions for the cost functionals J_1 and J_2.
The results are shown in Figure 14.3. The reconstructions of $\Sigma_t(s)$ and the CDF for J_1 in Figure 14.3a and 14.3b are again very close to the correct solution for small s, and a gap opens as s increases. The quality of the reconstruction is notably the best for $c = 0.9$. The reconstructions for the cost functional J_2 lie above the correct solution for $c = 0.7, 0.9$, and below for $c = 0.3, 0.5$. The CDFs for the values $c = 0.3, 0.5$ are slightly closer to the correct solution than for the other values.
Overall, we cannot determine a consistent effect of the scattering ratio c on the quality of the reconstructions obtained from the cost functional J_2. For results obtained from the cost functional J_1, however, we observe that the quality of the reconstruction obviously increases as the scattering ratio c approaches 1.

14.4.2 Finite dimensional controls

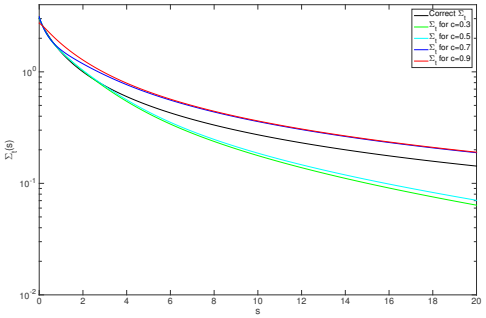
In this subsection we reconstruct parametrized non-classical cross sections for underlying algebraic, gamma, or log-normal path length distributions, as introduced in Section 11.3. We now fix the regularization parameters to $A = 1$ and



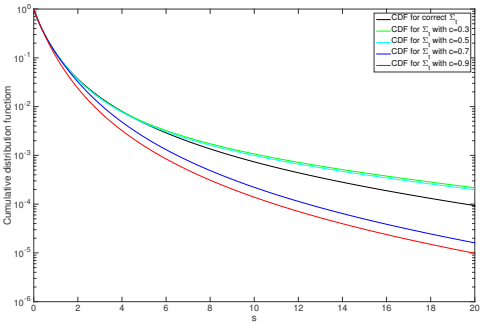
(a) Σ_t for cost functional J_1, and different choices of c.



(b) CDF for cost functional J_1, and different choices of c.



(c) Σ_t for cost functional J_2, and different choices of c.



(d) CDF for cost functional J_2, and different choices of c.

Figure 14.3: Comparison of the reconstructed and the exact non-classical cross section Σ_t (left) and the corresponding cumulative distribution functions (right) for the cost functional J_1 (top) and J_2 (bottom) for varying scattering ratio c.

$B = 0.0001$, and the scattering ratio is set to $c = 0.5$. For all reconstructions we we use a P_5-HLL scheme on the spatial interval $x \in [-20, 20]$, with 1000 discrete points in x and a CFL number $1/2$. The s-variable is cut off at $s_{\max} = 40$. We choose a Gaussian source term $Q(x)$, and the simplified cost functional J_2. The desired state $\bar{\phi}(x)$ is calculated using a P_{29} discretization with the same computational parameters. For each test we run 16 independent calculations, where $\bar{\phi}$ is perturbed pointwise by $\pm 5\%$, and the error is uniformly distributed. The optimal parameters for the particular cross sections are then reconstructed using a gradient based method, where the gradients are calculated using Equation (13.4a). We choose the same initial guess for the parameters each time, such that all calculations are comparable.

The results for an algebraic non-classical cross section are summarized in Table 14.1. The average reconstructions all underestimate the correct parameters $\alpha = 4, 5, 6, 7$. The average absolute relative error increases as α increases, but it is reasonably small (maximal 2%). Especially for $\alpha = 4$ and $\alpha = 5$ the correct parameter is reconstructed almost perfectly. As gradient method we choose an unconstrained quasi-Newton method. The small number of iterations needed is not surprising, since in the one dimensional parameter space the correct descent direction is fully determined by the sign of the gradient.

Type	correct α	avgerage α	relative error	avg. # iterations	avg. # grad. eval.
algebraic	4	3.9959	0.1017%	1.8125	13.688
algebraic	5	4.9727	0.5453%	5.4375	44.250
algebraic	6	5.9389	1.0179%	4.3125	75.563
algebraic	7	6.8532	2.0976%	4.5625	97.563

Table 14.1: Overview for the reconstruction of the parameter α for an underlying algebraic distribution.

Instead, the average number of gradient evaluations is higher. We also observe that the number of gradient evaluations increases as α increases.

Type	corr. a	corr. b	avg. a	avg. b	avg. rel. error	avg. # iterations	avg. # grad. eval.
gamma	2	1	1.9981	1.0011	2.4700%	20.188	59.125
gamma	3	2	3.3654	1.9414	4.6247%	11.063	36.063
log-normal	2	1	2.0004	1.0004	0.3902%	10.063	32.375
log-normal	1	2	0.8927	2.0146	4.8481%	12.188	28.375

Table 14.2: Overview over reconstruction of the parameters for underlying gamma and log-normal distributions.

The reconstruction of parametrized gamma and log-normal cross sections is a more complex task because of the two dimensional parameter space. We parametrize the gamma distribution p_{gam} and the log-normal distribution p_{logn} by

$$p_{gam}(s) = \frac{b^a}{\Gamma(a)} s^{a-1} \exp(-bs), \quad \text{and} \quad p_{logn}(s) = \frac{1}{as\sqrt{2\pi}} \exp\left(-\frac{(\log(s)-b)^2}{2a^2}\right).$$

The corresponding non-classical cross sections are specified in Section 11.3. For an initial guess for the parameters we always add 1 to the correct parameter. The parameters are then reconstructed using a SQP method that considers the positivity constraints on a and b. Otherwise, we observed negative parameters for which a forward solution does not exist.
The results for the tests are summarized in Table 14.2. We see again good reconstructions of the correct parameters for both gamma and log-normal distributions. The average absolute relative error for the log-normal and gamma distribution are slightly higher than in the algebraic one parameter case, but still yield good results (maximal absolute relative error around 4%). On average we need more iterations and gradient evaluations to reconstruct the parameters for the gamma distribution as for the log-normal distribution, but we cannot determine a systematic difference between the two distributions. In both cases the number of iterations is higher than for the one dimensional

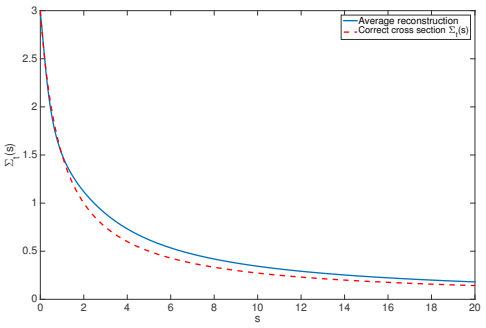
parameter space.

Overall, we can conclude that by our method we are able to reconstruct one and two dimensional parameters with high precision for reasonably perturbed data. The computational costs increase significantly when the dimensionality of the parameter space is increased. Positivity constraints become important even for the relative low parameter space dimension 2.

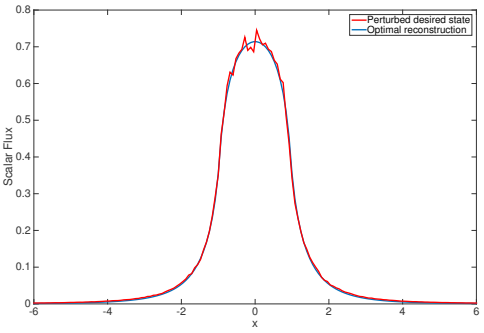
14.4.3 Infinite dimensional controls

The reconstruction of the full non-classical cross section from perturbed data for the scalar flux is the most complex testcase that we consider here. The computational setup is the same as for the previous testcases, and we use again a scattering ratio $c = 0.5$. We choose a discontinuous external source term with $Q(x) = 1$ for $-1 < x < 1$, and 0 elsewhere. A desired scalar flux $\bar{\phi}(x)$ is calculated using a P_{29} approximation, where the underlying path length distribution is algebraic with different parameters.

For **Testcase 1** we choose $\alpha = 4$, for **Testcase 2** we choose $\alpha = 5$, for **Testcase 3** we choose $\alpha = 6$, and for **Testcase 4** we choose $\alpha = 7$. For the reconstructions we we use again a P_5-HLL scheme. The initial guess for $\Sigma_t(s)$ corresponds to an algebraic distribution with $\alpha_0 = \alpha + 1$. Then we perform 16 independent reconstructions, where the desired scalar flux $\bar{\phi}$ is perturbed pointwise by $\pm 5\%$, and the error that we introduce is again equally distributed. The gradient based method that we use is an unconstrained quasi-Newton method.



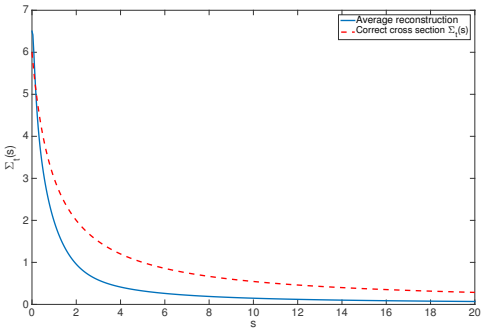
(a) Average reconstruction and correct $\Sigma_t(s)$.



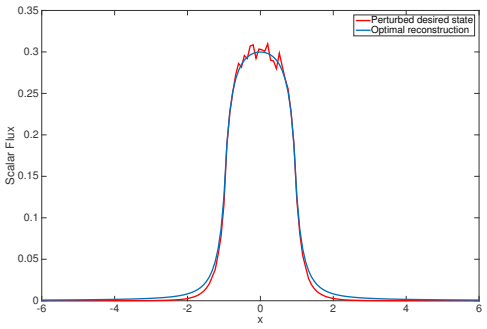
(b) Optimal reconstruction of the scalar flux.

Figure 14.4: Testcase 1: average reconstruction of an algebraic non-classical cross section with $\alpha = 4$ (left), and an optimal reconstruction of the scalar flux (right).

Figure 14.4 shows the results for Testcase 1. The average reconstruction of the non-classical cross section fits the correct one very well over the full s-interval (Figure 14.4a). A reconstruction of the underlying perturbed scalar flux is shown in Figure 14.4b, where we observe a very good qualitative fit. In Figure 14.5 we show the corresponding results for Testcase 4. The average of the reconstructions is again very good for small s, but for decreasing s the non-



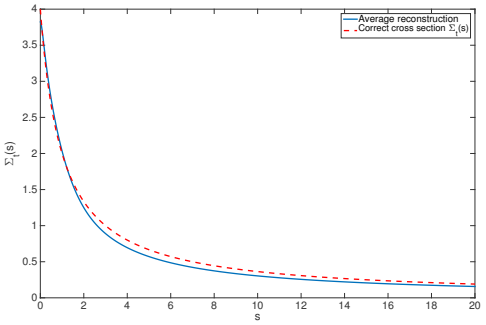
(a) Average reconstruction and correct $\Sigma_t(s)$.



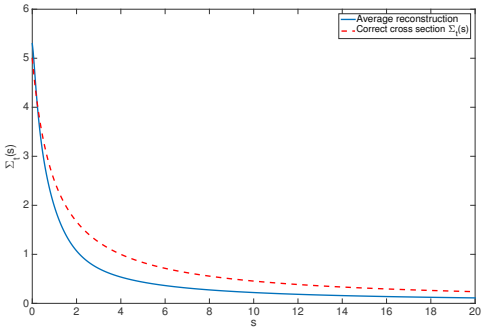
(b) Optimal reconstruction of the scalar flux.

Figure 14.5: Testcase 4: average reconstruction of an algebraic non-classical cross section with $\alpha = 7$ (left), and an optimal reconstruction of the scalar flux (right).

classical cross section is notably underestimated (Figure 14.5a). Qualitatively, that shows only a small effect in the reconstruction of the scalar flux in Figure 14.5b. Around the origin the reconstruction nicely fits the curve of the scalar flux, only towards the boundary we note a small discrepancy.



(a) Average reconstruction and correct $\Sigma_t(s)$ for $\alpha = 5$.



(b) Average reconstruction and correct $\Sigma_t(s)$ for $\alpha = 6$.

Figure 14.6: Testcases 2 and 3: average reconstruction of an algebraic non-classical cross section with $\alpha = 5$ (left), and $\alpha = 6$ (right).

Figure 14.6 shows the reconstructions of the non-classical cross sections for the Testcases 2 and 3. We again observe a good qualitative fit, although in Testcase 3 the correct cross section is notably underestimated for larger s.

A summary of the Testcases 1-4 is given in Table 14.3. The average relative errors (in a l^2-sense) to the correct cross section are high at the first sight, but we see that our results yield qualitatively satisfactory approximations of the underlying non-classical cross sections, and especially of the perturbed scalar fluxes. We observe that the number of iterations and gradient evaluations increases as α increases, while the relative absolute error also increases. Therefore, the increased computational cost for higher exponents α comes together with a notably loss of accuracy. For smaller values of α we observe

qualitatively and quantitatively better reconstructions (that was already observed in the previous subsection in Table 14.1). That may be related with a slower decay of the solutions in these cases, since we integrate over the full s domain to obtain the gradient.

	correct α	avg. rel. error	avg. # iterations	avg. # grad. eval.
Testcase 1	4	14.960%	10.688	37.938
Testcase 2	5	18.398%	14.688	40.813
Testcase 3	6	30.781%	18.875	47.750
Testcase 4	7	43.215%	20.438	64.563

Table 14.3: Overview for the reconstruction of the full cross section $\Sigma_t(s)$.

We conclude that our method can be used to reconstruct full non-classical cross sections from perturbed data for the scalar fluxes at a reasonable computational cost. These testcases are relatively close to realistic applications like the transport of photons in atmospheric clouds, which we have used as a motivational example throughout this work. Therefore, we expect that the solution of the parameter estimation problem based on the non-classical transport equation can be successfully applied there in the future.

15 Discrete first order optimality system

In this chapter we address the paradigms of first-discretize-then-optimize and first-optimize-then-discretize (also known as DO vs. OD). When we solve the discretized first order optimality system, there is an underlying discrete optimization problem that we solve implicitly. If this discrete problem is not an admissible discretization of the continuous optimality system, we expect that the resulting gradient is inaccurate. That may lead to a slower convergence or even incorrect solution of the optimality system. Therefore, we first investigate the underlying discrete first order optimality system of the S_N-HLL method. We compare the discretization of the continuous first order optimality system with the first order optimality system of the discretized minimization problem. We show that, for the S_N-HLL method, they coincide up to an error of the order of the discretization error. Treating the collision source term implicitly, this error can be further reduced. Then we show that these results also hold for the P_N-HLL discretization.

15.1 Consistency of a numerical scheme

We study the simplified parameter estimation problem as introduced in Chapter 11, without regularization term. That means, we consider Equation (6.4) in slab geometry with isotropic scattering

$$\begin{aligned}
&\partial_s \Psi(s,x,\mu) + \mu \partial_x \Psi(s,x,\mu) + \Sigma_t(s)\Psi(s,x,\mu) = 0,\\
&\Psi(0,x,\mu) = \frac{c}{2}\int_0^\infty \int_{-1}^1 \Sigma_t(s')\Psi(s',x,\mu)\,\mathrm{d}\mu\,\mathrm{d}s' + Q(x,\mu),
\end{aligned}$$

being the constraints of the simplified minimization problem

$$\begin{aligned}
&\min_{\Psi,\Sigma_t} J_1(\Psi,\Sigma_t), \qquad \text{with}\\
&J_1(\Psi,\Sigma_t) := \frac{1}{2}\int_0^\infty \int_{\mathbb{R}} \int_{-1}^1 \left(\Psi(s,x,\mu) - \bar{\Psi}(s,x,\mu)\right)^2 \mathrm{d}\mu\,\mathrm{d}x\,\mathrm{d}s,
\end{aligned}$$

where $\bar{\Psi}$ is known data that we want to fit with our model with a suitable parameter $\Sigma_t(s)$.

In the discrete case, the solution Ψ is a vector with length $N \cdot S_{\max} \cdot (2X+1)$, where N is the number of ordinates in the S_N method, $S_{\max}$ is the number of discrete points in pseudo-time, and $2X+1$ is the number of spatial points. In

the following, we choose the following ordering in the solution vector:

$$\Psi = \begin{pmatrix} \Psi_1^{-X,0} \\ \vdots \\ \Psi_N^{-X,0} \\ \Psi_1^{-X+1,0} \\ \vdots \\ \Psi_N^{X,0} \\ \Psi_1^{-X,1} \\ \vdots \\ \vdots \\ \Psi_N^{X,S_{\max}} \end{pmatrix}.$$

For a fixed pseudo time step and spatial cell, the neigboring entries vary by the ordinates, for a fixed pseudo time step, the spatial cells are ordered from $-X$ to X, and from top to the bottom of the vector the pseudo time step increases. The solution vector Ψ depends on the parameter Σ_t, which is a vector in the discrete case with entries Σ_t^n, $n = 0, ..., S_{\max}$. Therefore, the reduced cost functional which we want to minimize can be written as

$$F(\Sigma_t) = \frac{1}{2}\|\Psi(\Sigma_t) - \bar{\Psi}\|_c^2 = \frac{1}{2}\sum_{n=1}^{S_{\max}}\sum_{i=1}^{N}\sum_{j=-X}^{X}\alpha_n\omega_i\beta_j(\Psi_i^{j,n} - \bar{\Psi}_i^{j,n})^2,$$

with $\alpha_n, \omega_i, \beta_j$ being the weights of a quadrature rule for the integration with respect to s, μ and x. We do not consider regularization terms in this chapter, since they do not contribute to the question of consistency between the discrete and the continuous optimality system. For each numerical method that we consider, Ψ is the solution to the following fixed-point problem

$$\Psi = A\Psi + q$$

with Ψ as indicated above, the source term q being a vector given by

$$q = \begin{pmatrix} Q_{-X} \\ \vdots \\ Q_X \\ 0 \\ \vdots \\ 0 \\ \vdots \end{pmatrix},$$

and A is a matrix whose structure depends on the numerical method. Note that the matrix A and therefore also Ψ depend on the parameter Σ_t. The

solution can therefore, in theory, be directly calculated as

$$\Psi = Rq, \qquad R = \big(\mathrm{Id} - A\big)^{-1}.$$

where Id is the identity matrix of the same dimension as A. In practice, this is solved iteratively. Nevertheless, the final solution satisfies the fixed point equation. The discrete optimization problem then reads

$$\min_{\Sigma_t} F(\Sigma_t) \quad \text{under} \quad (\mathrm{Id} - A)\Psi = q, \quad \Sigma_t \geq 0. \tag{15.1}$$

In order to derive the first order optimality system, we have to take the derivative of the cost functional F with respect to the parameter Σ_t. That yields

$$\frac{\partial F}{\partial \Sigma_t} = (\Psi(\Sigma_t) - \bar{\Psi}) \cdot C \frac{\partial \Psi}{\partial \Sigma_t},$$

where C is a diagonal matrix containing the products of the quadrature weights (of the form $\alpha_n \omega_i \beta_j$), and

$$\frac{\partial \Psi}{\partial \Sigma_t} = \big(\mathrm{Id} - A\big)^{-1} \frac{\partial A}{\partial \Sigma_t} \Psi = R \frac{\partial A}{\partial \Sigma_t} \Psi.$$

Rearranging this term, we find the discrete first order optimality system from the necessary condition

$$\frac{\partial F}{\partial \Sigma_t} = \Big(C \frac{\partial A}{\partial \Sigma_t} \Psi\Big)^T R^T \big(\Psi - \bar{\Psi}\big) = 0.$$

Introducing an adjoint state vector λ, and writing $C = C_1 \cdot C_2$, with C_1 being a diagonal matrix whose entries are the corresponding quadrature weights α_n, we find the forward and adjoint equations to be

$$\begin{aligned} &\Psi = Rq, && \text{(forward equation)} \\ &\lambda = R^T C_1 (\Psi - \bar{\Psi}). && \text{(adjoint equation)} \end{aligned}$$

The adjoint eqation can also be written in fixed point form as

$$\lambda = A^T \lambda + C_1 (\Psi - \bar{\Psi}),$$

since inverse and transpose interchange and taking the transpose is a linear operation. We see later on, why the splitting of the matrix C is important here. Hence, the discrete adjoint and the discretization of the continuous equation is consistent if the matrix A^T represents the discretization of the continuous adjoint equation. The missing update equation is then given by the necessary condition

$$\frac{\partial F}{\partial \Sigma_t} \Big(C_2 \frac{\partial A}{\partial \Sigma_t} \Psi\Big)^T \lambda = \Psi^T \frac{\partial A}{\partial \Sigma_t}^T C_2 \lambda = 0, \tag{15.2}$$

which yields a discretization of the integral of the products of Ψ and λ as in

Chapter 13.
In the following, we use these notations and calculations to show that for the S_N-HLL method the first order optimality system of the discrete problem coincides with the discretization of the first order optimality system of the continuous problem under some mild conditions, up to an error of the order of the discretization. These results can be then directly carried over to the P_N-HLL discretization of the first order optimality system.

15.2 S_N-HLL discretization

The S_N-HLL discretization of the forward equation, as introduced in the previous Chapter 14, reads as

$$\frac{\Psi_i^{j,n+1}-\Psi_i^{j,n}}{\Delta s}+\mu_i\frac{\Psi_i^{j+1,n}-\Psi_i^{j-1,n}}{2\Delta x}-b\frac{\Psi_i^{j+1,n}-2\Psi_i^{j,n}+\Psi_i^{j-1,n}}{2\Delta x}+\Sigma_t^n\Psi_i^{j,n}=0, \tag{15.3a}$$

$$\Psi_i^{j,0}=\frac{c}{2}\sum_{n=0}^{\infty}\sum_{i=1}^{N}\alpha_n\Sigma_t^n\omega_i\Psi_i^{j,n}+Q_i^j, \tag{15.3b}$$

where $\Psi_i^{n,j}=\Psi(s^n,x_j,\mu_i)$ and $b=\max|\mu_i|$. Then, the solution of the forward equation satisfies the fixed-point equation

$$\Psi=A\Psi+q,$$

where A is a matrix of the following form:

$$A=\begin{pmatrix} S_0 & S_1 & S_2 & \dots & \dots & S_{S_{\max}} \\ E_0 & 0 & 0 & \dots & \dots & 0 \\ 0 & E_1 & 0 & \dots & \dots & 0 \\ 0 & 0 & E_2 & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & E_{S_{\max}-1} & 0 \end{pmatrix}.$$

The subblocks of A are given by

$$S_n=\frac{c}{2}\alpha_n\Sigma_t^n\begin{pmatrix} T & 0 & \dots & 0 \\ 0 & T & \dots & 0 \\ 0 & \dots & \ddots & 0 \\ 0 & \dots & \dots & T \end{pmatrix}, \quad \text{with} \quad T=\begin{pmatrix} \omega_1 & \dots & \dots & \omega_N \\ \vdots & \dots & \dots & \vdots \\ \omega_1 & \dots & \dots & \omega_N \end{pmatrix},$$

and, in case of Dirichlet boundary conditions,

$$E_n = \begin{pmatrix} E_n^0 & E_n^1 & 0 & \dots & \dots & 0 \\ E_n^{-1} & E_n^0 & E_n^1 & \dots & \dots & 0 \\ 0 & E_n^{-1} & E_n^0 & E_n^1 & \dots & 0 \\ 0 & \dots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & E_n^{-1} & E_n^0 & E_n^1 \\ 0 & \dots & \dots & 0 & E_n^{-1} & E_n^0 \end{pmatrix}.$$

with diagonal submatrices

$$E_n^0 = \big(1 - \Delta s \Sigma_t^n - b\frac{\Delta s}{2\Delta x}\big)\,\mathrm{Id}_{N\times N}, \quad E_n^{-1} = \big(\frac{\Delta s}{2\Delta x}\mu_i + b\frac{\Delta s}{\Delta x}\big)\,\mathrm{Id}_{N\times N},$$
$$E_n^1 = \big(-\frac{\Delta s}{2\Delta x}\mu_i + b\frac{\Delta s}{\Delta x}\big)\,\mathrm{Id}_{N\times N}\,.$$

The indexed term μ_i means that this value varies over the diagonal of these matrices, i.e. $E_n^1(i,i) = -\frac{\Delta s}{2\Delta x}\mu_i + b\frac{\Delta s}{\Delta x}$.

In order to be consistent with the discrete first order optimality system, we need the following additional conditions on the S_N-HLL discretization to be satisfied.

(1) The quadrature rule for the $s-$integration is given by the trapezoidal rule, i.e. the corresponding weights are

$$\alpha_0 = \frac{\Delta s}{2}, \quad \alpha_n = \Delta s \quad \text{for} \quad 1 < n < S_{\max}, \quad \alpha_{S_{\max}} = \frac{\Delta s}{2};$$

(2) For the weights ω_i in the quadrature rule for the angular variable μ we have

$$\omega_1 = \dots = \omega_N = \frac{2}{N}.$$

(3) The source term $\Sigma_t(s)\Psi$ is discretized implicitly, that means we replace the term $\Sigma_t^n \Psi_i^{j,n}$ by $\Sigma_t^{n+1}\Psi_i^{j,n+1}$ in the first line of the Scheme (15.3).

Assuming this, the following theorem holds.

Theorem 15.1. *(Consistency with discrete optimality system)*
Under the Assumptions (1)-(3), the solution of the first order optimality system of the S_N-HLL discretized minimization problem (15.1) and the solution of the S_N-HLL discretization of the continuous optimality system coincide up to an error of the order of Δs.

Proof. From the introductory discussion it is clear that it remains to study if the transpose matrix A^T yields an admissible discretization of the adjoint equation, and its derivative with respect to Σ_t an admissible discretization of the update equation.

Taking into account Condition (3), the transpose matrix A^T has the form

$$A^T = \begin{pmatrix} \tilde{S}_0^T & \tilde{E}_0^T & \dots & \dots & \dots & 0 \\ S_1^T & D_1 & \tilde{E}_1^T & \dots & \dots & \vdots \\ S_2^T & 0 & D_2 & \tilde{E}_2^T & \dots & \vdots \\ \vdots & 0 & 0 & D_3 & \ddots & 0 \\ \vdots & 0 & \dots & \dots & \ddots & \tilde{E}_{S_{\max}-1}^T \\ S_{S_{\max}}^T & 0 & \dots & \dots & \dots & D_{S_{\max}} \end{pmatrix},$$

where $D_n = (-\Delta s \Sigma_t^n)\,\mathrm{Id}_{N\cdot(2X+1)\times N\cdot(2X+1)}$, $\tilde{S}_0^T = S_0^T + D_0$, and $\tilde{E}_n^T = E_n^T - D_n$. Defining $r := \Psi - \bar{\Psi}$, the resulting discrete adjoint equation is

$$\lambda = A^T\lambda + C_1 r.$$

Line by line, this reads

$$\begin{aligned}\lambda_i^{j,n} =& -\Delta s\lambda_i^{j,n} + \big(1 - b\frac{\Delta s}{2\Delta x}\big)\lambda_i^{j,n+1} + \big(\frac{\Delta s}{2\Delta x}\mu_i + b\frac{\Delta s}{\Delta x}\big)\lambda_i^{j-1,n+1} \\ &+ \big(-\frac{\Delta s}{2\Delta x}\mu_i + b\frac{\Delta s}{\Delta x}\big)\lambda_i^{j+1,n+1} + \frac{c}{2}\Sigma_t^n\alpha_n\sum_{l=1}^{N}(\omega_i\lambda_l^{j,0}) + \alpha_n r_i^{j,n}.\end{aligned}$$

Assumption (1) yields $\alpha_n = \Delta s$ and Assumption (3) gives $\omega_i = \omega_l$ for $= 1, \dots, N$. Then rearranging the terms above gives

$$\begin{aligned}-\frac{\lambda_i^{j,n+1} - \lambda_i^{j,n}}{\Delta s} - \mu_i\frac{\lambda_i^{j+1,n+1} - \lambda_i^{j-1,n+1}}{2\Delta x} - b\frac{\lambda_i^{j+1,n+1} - 2\lambda_i^{j,n+1} + \lambda_i^{j-1,n+1}}{2\Delta x} \\ +\Sigma_t^n\lambda_i^{j,n} = \frac{c}{2}\Sigma_t^n\sum_{l=1}^{N}(\omega_l\lambda_l^{j,0}) + r_i^{j,n},\end{aligned}$$

which is an admissible S_N-HLL discretization of the adjoint equation. For $n = 0$ we have $\alpha_0 = \frac{\Delta s}{2}$ instead of Δs, hence that accounts for an error of $O(\Delta s)$. On the other hand, for $n = S_{\max}$ we have

$$\lambda_i^{j,S_{\max}} = \frac{\Delta s}{1 - \Delta s}\frac{c}{2}\Sigma_t^{S_{\max}}\sum_{l=1}^{N}(\omega_l\lambda_l^{j,0}),$$

which is again an error of $O(\Delta s)$ when we compare with the formal condition $\lambda(\infty, x, \mu) = 0$. Note that often we can assume, that $\Sigma_t \to 0$ as $s \to \infty$. That attenuates this error additionally.
Now we consider the discrete gradient (15.2). The quantity of interest there is the derivative of the matrix A with respect to the non-classical cross section

Σ_t . We write this componentwise as

$$\frac{\partial A^T}{\partial \Sigma_t^n} = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ \vdots & 0 & \dots & \dots & \vdots \\ \frac{\partial}{\partial \Sigma_t^n} S_n^T & 0 & \frac{\partial}{\partial \Sigma_t^n} D_n & \frac{\partial}{\partial \Sigma_t^n} \tilde{E}_n^T & 0 \\ 0 & \dots & \dots & \dots & \vdots \\ \vdots & 0 & \dots & \dots & 0 \end{pmatrix}$$

since the subblocks are ordered by pseudo time steps in the matrix A. The subblocks of this matrix are given by

$$\frac{\partial}{\partial \Sigma_t^n} S_n^T = \frac{c}{2}\alpha_n T, \ \frac{\partial}{\partial \Sigma_t^n} D_n^T = -\Delta s \, \mathrm{Id}_{N\cdot(2X+1)\times N\cdot(2X+1)}, \quad \text{and} \quad \frac{\partial}{\partial \Sigma_t^n} E_n^T = 0.$$

When we write down the necessary Condition (15.2) line by line, we obtain

$$\begin{aligned} \frac{\partial F}{\partial \Sigma_t^n} = \ \Psi^T \frac{\partial A^T}{\partial \Sigma_t^n} C_2 \lambda = \frac{c}{2}\alpha_n \sum_j \sum_i \omega_i \beta_j \Psi_i^{j,n} \sum_{i'} \omega_{i'} \lambda_{i'}^{j,0} \\ - \Delta s \sum_j \sum_i \omega_i \beta_j \Psi_i^{j,n} \lambda_i^{j,n} = 0. \end{aligned}$$

For every $1 \leq n < S_{\max}$ we have that $\alpha_n = \Delta s$ because of Assumption (1), hence, we can multiply Equation (15.2) by $1/\Delta s$ and obtain a discretization of the continuous Gradient (13.3). For $n = 0$ and $n = S_{\max}$ we have $\alpha_0 = \Delta s/2$ by Assumption (1), which again accounts for an error of $O(\Delta s)$. That finishes the proof.

□

Remark 15.1. To finish the discussion of the consistency of the S_N-HLL discretization, we want to point out the following:

(a) This result extends to more complicated cost functionals as introduced in Chapter 11, since the cost functional only effects the external source term of the adjoint equation.

(b) A generalization to the case of anisotropic scattering is possible as long as the scattering kernel is symmetric. Then we have $T = T^T$ under Assumption (2), and the scheme is still consistent.

(c) Assumption (2) is a comparably strong restriction for the discretization of the angular variable. As we have pointed out in Chapter 7, the discrete ordinates method is equivalent to the P_N-method only if Gauss points (with corresponding weights) are chosen for the angular integration. Hence, there is a trade-off here between consistency of the numerical scheme and accuracy of the integration in the angular variable.

(d) Assumption (3) is necessary in order to obtain a correct discretization of the the second integral in Equation (13.3). It brings the entries that

depend Σ_t on the diagonal line of A, hence of A^T.

(e) Note that our calculation shows, that the quadrature rule for the $s-$integration in the initial condition in forward calculation yields the pseudo-time steps of the the discretized adjoint calculations. That means, that we cannot use an arbitrary quadrature rule in order to obtain consistency of the numerical schemes. Hence, assumption (1) cannot be weakened without a further loss of consistency.

(f) The Assumptions (1)-(3) yield a certain symmetry of the submatrices of A. Since the $s-$variable is not symmetric (initial value $s = 0$ but cut-off at some $S_{\max}$), we cannot avoid an additional error of the order $O(\Delta s)$.

15.3 P_N-HLL discretization

In order to analyze the consistency of the P_N-HLL discretization, we restrict ourselves to the analysis of the P_1 case for reasons of simplicity and clearity. Everything remains valid for the general case $N > 1$. We assume again, that Assumptions (1) and (3) from the previous section hold. Then we denote by E and F the zeroth and first angular moment of the underlying distribution function, and order the discretized vector as follows:

$$\Psi = \begin{pmatrix} E^0_{-X} \\ F^0_{-X} \\ \vdots \\ E^0_X \\ F^0_X \\ E^1_{-X} \\ \vdots \\ \vdots \\ F^{S_{\max}}_X \end{pmatrix}.$$

Then we can write the resulting linear system in fixed-point form as

$$\Psi = A_{PN}\Psi + q,$$

where q is the discretized source term as in the previous section, and A_{PN} has the following structure:

$$A_{PN} = \begin{pmatrix} \bar{S}_0 & \bar{S}_1 & \bar{S}_2 & \dots & \dots & \bar{S}_{S_{\max}} \\ \bar{E}_0 & \bar{D}_1 & 0 & \dots & \dots & 0 \\ 0 & \bar{E}_1 & \bar{D}_2 & \dots & \dots & 0 \\ 0 & 0 & \bar{E}_2 & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \bar{E}_{S_{\max}-1} & \bar{D}_{S_{\max}} \end{pmatrix}.$$

Here, the subblocks $\bar{S}_n$ and $\bar{D}_n$ are diagonal matrices with

$$\bar{S}_n = c\alpha_n \Sigma_t^n \begin{pmatrix} 1 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & \dots & \dots & \ddots & 0 & 0 \\ 0 & \dots & \dots & \dots & 1 & 0 \\ 0 & \dots & \dots & \dots & 0 & 0 \end{pmatrix},$$

and $\bar{D}_n = (-\Delta s \Sigma_t^n)\,\mathrm{Id}_{2(2X+1)\times 2(2X+1)}$. The subblocks $\bar{E}_n$ on the lower diagonal have the following tridiagonal form

$$\bar{E}_n = \begin{pmatrix} H_0 & H_1 & 0 & \dots & \dots & 0 \\ H_{-1} & H_0 & H_1 & \dots & \dots & 0 \\ 0 & H_{-1} & H_0 & H_1 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & H_1 \\ 0 & \dots & \dots & \dots & H_{-1} & H_0 \end{pmatrix},$$

with

$$H_{-1} = \begin{pmatrix} -\frac{\Delta s}{2\Delta x} & \frac{\Delta s}{2\Delta x} \\ \frac{\Delta s}{6\Delta x} & -\frac{\Delta s}{2\Delta x} \end{pmatrix}, \quad H_0 = \begin{pmatrix} 1+\frac{\Delta s}{\Delta x} & 0 \\ 0 & 1+\frac{\Delta s}{\Delta x} \end{pmatrix}, \quad H_1 = \begin{pmatrix} -\frac{\Delta s}{2\Delta x} & -\frac{\Delta s}{2\Delta x} \\ \frac{\Delta s}{6\Delta x} & -\frac{\Delta s}{2\Delta x} \end{pmatrix}.$$

We have seen in the proof of Theorem 15.1 that we now need to check, that (a) A_{PN}^T yields an admissible discretization of the non-classical adjoint equation, and that (b) the derivative of A_{PN}^T yields a discretization of the continuous Gradient (13.3). The discrete adjoint equation without residual term

$$\lambda = A_{PN}^T \lambda$$

can be written line by line as

$$\begin{aligned} -\frac{\lambda_i^{(0),n+1} - \lambda_i^{(0),n}}{\Delta s} - \frac{\lambda_{i+1}^{(1),n+1} - \lambda_{i-1}^{(1),n+1}}{2\Delta x} - \frac{\lambda_{i+1}^{(0),n+1} - 2\lambda_i^{(0),n+1} + \lambda_{i-1}^{(0),n+1}}{2\Delta x} \\ + \Sigma_t^n \lambda_i^{(0),n} = c\Sigma_t^n \lambda_i^{(0),0}, \\ -\frac{\lambda_i^{(1),n+1} - \lambda_i^{(1),n}}{\Delta s} - \frac{1}{3}\frac{\lambda_{i+1}^{(0),n+1} - \lambda_{i-1}^{(0),n+1}}{2\Delta x} - \frac{\lambda_{i+1}^{(1),n+1} - 2\lambda_i^{(1),n+1} + \lambda_{i-1}^{(1),n+1}}{2\Delta x} \\ + \Sigma_t^n \lambda_i^{(1),n} = 0, \end{aligned}$$

where $\lambda^{(0)}$ and $\lambda^{(1)}$ denote the zeroth and first angular moment of the adjoint variable λ. This is the P_1-HLL discretization of the non-classical adjoint equation, where the collision source term is treated implicitly. The residual term can obviously be included and then shows up on the right-hand side.
Because of Assumption (3) the discretized cross section contributions are only

found in the submatrices $\bar{S}_n$ and $\bar{D}_n$, and it is easy to show that

$$\frac{\partial F}{\partial \Sigma_t^n} = \Psi^T \frac{\partial A_{PN}^T}{\partial \Sigma_t^n} \tilde{C}_2 \lambda = \frac{c}{2} \Delta s \sum_i \beta_i \Psi_i^{(0),n} \lambda_i^{(0),0} - \Delta s \sum_i \sum_{j=0}^{1} \frac{2j+1}{2} \beta_i \Psi_i^{(j),n} \lambda_i^{(j),n} = 0,$$

where the notation $E = \Psi^{(0)}$ and $F = \Psi^{(1)}$ was used, and $\tilde{C}_2$ contains the factors of the P_N reconstruction. This is the P_1-HLL discretization of Equation (13.3) without regularization term. At the beginnining and the end of the s-interval we have additional errors of the order $O(\Delta s)$, again. Because of the tridiagonal form of the P_N equations, these results hold for arbitrary N.

Hence, the following theorem holds:

Theorem 15.2. *(Consistency with discrete optimality system)*
Assuming (1) and (3), the solution of the first order optimality system of the P_N-HLL discretized minimization problem (15.1) and the solution of the P_N-HLL discretization of the continuous optimality system coincide up to an error of the order of Δs.

16 Discussion, conclusion, and outlook

In this work, we investigate a new and non-classical linear transport equation for the transport of particles in correlated background media. This work falls naturally into three parts, where we study the derivation of mathematical models, the design and analysis of numerical methods, and a parameter estimation method based on the preceding results. We give a quick summary of this work:

- In **Part I** we derive a non-classical transport equation that is capable of reproducing arbitrary path length distributions, in contrast to the classical theory. The rigorous mathematical derivation is based on recent results on the periodic Lorentz gas, and results in a time-dependent kinetic equation that has the distance s to the next collision as an additional independent variable. We discuss generalizations of the resulting equation, as well as the connection to another recently proposed non-classical steady state transport equation.
- The scope of **Part II** is the development and analysis of numerical methods for this non-classical steady state transport equation. A variation of a standard finite volume HLL method is developed for moment models of this equation. In a detailed numerical analysis we show that this scheme has a number of desirable properties. The naive source iteration method can be shown to become arbitrary slow in the scattering dominated regime, therefore we adapt an acceleration scheme for the classical equation based on the diffusion approximation for the non-classical equation.
- Finally, **Part III** is devoted to the solution of inverse problems based on the non-classical steady state transport equation. We formulate the inverse problem as an optimal control problem, and derive a first order optimality system using a Lagrange formalism. Then we adapt the numerical schemes developed in Part II to discretize this first order optimality system. We also show that these schemes are consistent with the underlying discrete optimality system. Numerical solutions of the optimality system can then be obtained using gradient based optimization methods.

In the following we present a detailed discussion and conclusion of each part in a separate section. Therefore, we first summarize our results and achievements, but then also point out open questions and give an outlook on future research in the corresponding topics.

16.1 Part I: Mathematical Modeling

In the first part of this work, we present a rigorous derivation of a time-dependent non-classical transport equation (Equation (4.11)), and establish a connection to the non-classical steady state equation recently introduced by Larsen [73] (Equation (6.4)). The derivation is based on recent results on the periodic Lorentz gas [53, 81], and holds in the following simplified setting:

- scatterers are fixed on a Cartesian grid,
- we only consider 2 space dimensions, and
- scattering is isotropic.

In this setting we show, that the distribution function of a microscopic particle game converges in the Boltzmann-Grad limit towards the solution of the non-classical transport Equation (4.11). The proof heavily relies on an analytic expression for the distribution of free path lengths in the limit, obtained by [21, 15, 81]. Nevertheless, this proof is not restricted to the above described simplified setting. This is discussed in detail in Section 4.3, where we suggest that this result holds in arbitrary space dimension, and for general homogeneous background media. A Monte-Carlo simulation of path length distributions in the Boltzmann-Grad limit in Chapter 5 underlines this. We find a number of correlated but homogeneous media that show path length distributions which clearly deviate from the classical exponential distribution. The connection to Larsen's equation, that we establish in Chapter 6, shows that the non-classical cross section $\Sigma_t(s)$ in this equation is inseparably linked to the underlying path length distribution, and can therefore be interpreted as a physical quantity. Finally, the formal extensions of the non-classical transport equation to more general situations, where inhomogeneities or external source terms are present, indicate that these equations can be applied to model a number of realistic physical situations, where path length distributions different from the classical exponential occur.

Outlook

We propose that future research regarding the mathematical modeling of non-classical transport is set to the following open questions and problems.

1. The rigorous derivation of a non-classical transport equation in this work relied on the periodic setting of scatterers, since we make use of recent analytic results in this setting. Extensions to general media are carefully discussed in this work, and underlined by a Monte-Carlo study in several special cases. But a rigorous derivation in general correlated media is still lacking. Regarding a solution of this problem, the recent results of Marklof and Strömbergsson [82, 83] are very promising. They extend the results from the periodic Lorentz gas to a number of more general crystalline structures.

2. The number of correlated background media that can be produced using the Metropolis Monte-Carlo method from Chapter 5 seems to be very limited. A better, hence more generally valid notion of spatial correlations in the background medium could be based on so-called Lévy point processes [58]. An analytic relation between the distribution of the point processes and the underlying path length distributions could help to answer the previous problem 1. Furthermore, the study of path length distributions in fractal media (deterministic or random) is an interesting and promising field that is not covered by this work.

3. The numerical investigation in Section 10.2 shows that the decay of the scalar flux towards the spatial boundary is determined by the underlying path length distribution. This cannot replace an analytic investigation of the properties of the solution $\Psi(s, x, \Omega)$. In particular, the question of how the solution decays in the $s-$ and $x-$variable separately stands out.

4. We rigorously derive a non-classical transport equation in Chapter 4 in an infinite domain, and without external source terms. The introduction of source terms and boundary conditions in Chapter 6 is based on formal arguments, and therefore needs further investigation. The same holds for the introduction of an additional energy dependence, or non-classical transport in inhomogeneous media. A rigorous theory that covers all these effects is still lacking.

5. The diffusion approximation (given by Equation (6.7)) is only valid as long as the second moment of the path length distribution is finite. In Remark 6.5 we refer to an upcoming paper [43], where a fractional diffusion approximation is derived for the case of an infinite second moment. But to establish a closer connection to the theory of anomalous transport, the mean square displacement should be investigated. In order to do this, the time-dependent non-classical Equation (6.1) seems to be more appropriate.

16.2 Part II: Numerical Methods

Because of the reduced complexity, we restrict the development of numerical methods to the non-classical steady state transport equation (Equation (6.4)). This equation contains three independent variables: the spatial variable x, the direction of flight Ω, and the path length variable s. The directional variable Ω can be discretized using a direct approach (discrete ordinates), or the method of moments. We mainly focus on the development of numerical methods for moment models of Equation (6.4) in this work.

The path length variable s can be interpreted as a pseudo time, hence we can apply standard finite volume schemes for the discretization in s and x for the moment models, that are introduced in Chapter 7. We follow the formalism of Harten, Lax, and van Leer [57] and derive a numerical scheme for the P_1, and the M_1 model in a diffusive scaling in Chapter 8. By a light modification of

the numerical diffusion term, we show that the resulting schemes preserve the correct diffusion limit on a discrete level, in contrast to standard HLL methods. Furthermore, we show that the schemes preserve the area of admissibility and realizability, for both an explicit and implicit pseudo time discretization. In the P_1 case it is shown that the scheme is L^2-stable under a mild CFL condition. This shows, that the numerical schemes are suited to solve the non-classical steady state transport equation in diffusive regimes, which is important in many applications such as transport in atmospheric clouds.

A coupling of the initial value and the full solution in Equation (6.4) strongly suggests to use an iterative solution method. We show that the naive source iteration method, which is equivalent to a Richardson iteration, becomes arbitrarily slow in the scattering dominated regime. A von Neumann analysis of the full equation, as well as of the moment models in Chapter 9 shows, that the contraction rate is determined by the scattering ratio c. Therefore, we introduce a Diffusion Synthetic Acceleration method for the non-classical steady state transport equation, also in Chapter 9. The residual in each iteration step is approximated by the diffusion approximation, and used to correct the current iterate. A von Neumann analysis for this method, again for the full equation and the corresponding moment models, shows that the contraction rate is significantly decreased, and depends on the underlying path length distribution. We calculate contraction rates for different algebraically decaying path length distributions, and numerical results show a significant speed up of the iteration method.

Outlook

We propose that future research regarding numerical methods for non-classical transport equations is set to the following open questions and problems.

1. We present all numerical methods and results in 1D slab geometry. In order to be applicable to realistic applications, all methods have to be extended to 2D and 3D. This task poses no other additional problems than a higher computational complexity, since non of the arguments or techniques that we use are exclusively restricted to the 1D case.

2. It is known that the Diffusion Synthetic Acceleration method suffers from numerical instabilities, especially in higher dimensions. Krylov subspace methods, on the other hand, are known to be more robust and stable, and they are already successfully used to solve the classical linear Boltzmann equation [1, 96]. These methods are also applicable for the non-classical transport equation. A first investigation of Krylov subspace methods for the non-classical transport equation, using discrete ordinates and a diamond difference discretization in space, is carried out succesfully in [93].

3. We consider the AP property of the HLL schemes in Chapter 8 only for the case when the second moment of the path length distribution exists.

Therefore, it is an open question if the schemes also preserve a fractional diffusion limit (as suggested in [43]) for the case $\langle s^2 \rangle = \infty$. Furthermore, it is an open question if Diffusion Synthetic Acceleration methods can be applied based on fractional diffusion approximations.

4. From a numerical point of view, the time-dependent non-classical Equation (6.1) involves two time variables : the actual time t, and the distance to collision s that serves as a pseudo time. Furthermore, the equation needs to be solved forward in t, and backward in s. Numerical schemes for this kind of equation need to be developed and analyzed.

5. All numerical schemes that we present in this work are of first order in space and pseudo time. The development of higher order numerical schemes should be considered for future work. In [12, 91] a space-time generalized Riemann problem (GRP) solver of an arbitrary order k with unrestricted time step was introduced. It provides a polynomial representation of the solution of order k with respect to the spatial variable x, as well as to the time variable. These S_N-GRP schemes seem to be ideal to solve the non-classical steady state transport equation, especially because of the unrestricted time step.

16.3 Part III: Parameter Estimation

We present a method to calculate the underlying path length distribution from measured data for the angular flux, based on the non-classical steady state transport equation. The $s-$dependent total cross section $\Sigma_t(s)$ can be expressed in terms of the path length distribution $p(s)$, and vice versa. Therefore, the parameter estimation problem can be formulated as an optimal control problem for cost functionals that depend on the angular flux, with constraints given by Equation (6.4) (see Chapter 11). We also consider parametrized path length distributions, where the parameter estimation problem is reduced to the reconstruction of a finite number of optimal parameters.

We follow a Lagrange formalism to express the gradient of the cost functionals in terms of the solution of Equation (6.4), and the adjoint variable. By doing so, we avoid the direct analytical differentiation of the cost functional. Therefore, we formally derive the adjoint equation of Equation (6.4) in Chapter 12. The first order optimality system that we obtain in Chapter 13 then yields the desired gradient. Equipped with this, we can use gradient based optimization methods to solve the optimal control problem.

The HLL finite volume schemes are adapted for the S_N and P_N discretization of the non-classical equation and the correspondig adjoint equation in Chapter 14. We focus on simplified test functionals and 1D slab geometry cases to verify and test the numerical methods. Numerical results show that the gradient can be accurately calculated, and that parametrized and full cross sections can be reconstructed from perturbed data at a reasonable computational cost. Fur-

thermore, a numerical analysis in Chapter 15 shows that the discretizations of the adjoint equation are consistent with the underlying discrete optimization problem.

All this indicates, that the method proposed in this work is suited to solve realistic parameter estimation problems, for example for the transport of photons in atmospheric clouds. In contrast to recently used reconstruction methods [48], our method does not suffer from modeling inconsistencies, as the rigorous derivation of Equation (6.4) shows that any path length distribution can be reproduced by the non-classical transport equations.

Outlook

We propose that future research regarding parameter estimation based on the non-classical transport equation is set to the following open questions and problems.

1. The derivation of the first order optimality system in this work is purely formal. Therefore, an analytic well posedness result for the regularized inverse problem is lacking. This includes the development of a rigorous existence theory for the non-classical transport equation in L^2, and in suitable Banach spaces (such as L^1). For the reconstruction of total and scattering cross sections based on the classical transport equation a comparable analysis is carried out in [38], and we propose to transfer this analysis to the non-classical transport equation.

2. The use of a standard Tykhonov regularization is restricted to non-classical cross sections $\Sigma_t(s) \in L^2([0,\infty))$. This excludes the case of classical transport, but also cross sections from path length distributions that are log-normal. Therefore, a regularization needs to be developed and analyzed, that is able to cover all relevant types of non-classical cross sections.

3. Gradient based methods for the solution of the first order optimality system need to be investigated in detail. A comparison of different methods, and a study of the computational performance should be performed.

4. Parameter estimation based on the non-classical transport equation needs to be applied to realistic test cases in 2D and 3D. For the application to photon transport in atmospheric clouds suitable data is available (via I3RC [24]), and the analysis of the transmittance as an exemplary functional of the angular flux also aims at this application.

A Fundamentals of measure and probability theory

This Appendix is meant to provide the basic notations, definitions and results of measure and probability theory that are used in this work (mainly in Part I). All important terms and definitions are emphasized in bold letters. The notations and the style of presentation heavily rely on the books [40, 41] and [39, 47]. For the part on stochastic processes we also refer to [55] and [88]. We do not present any proofs in this Appendix. For that we refer to the aforementioned textbooks.

A.1 Measure theory

We focus on a short and simple introduction of the theory here. We only present what is used and needed within this piece of work.
Let Ω be an (arbitrary) set. A $\sigma-$algebra $\mathcal{F}(\Omega)$ is a system of subsets of Ω such that the empty set, the full set, and the union, the intersection of two elements and the complement of any element is again an element of this system. A **measure** μ on Ω is now a function $\mu:\ \mathcal{F}(\Omega) \longrightarrow [0,\infty)$ such that

$$\text{(a)}\ \ \mu(\emptyset) = 0, \quad \text{(b)}\ \ \mu(F) \geq 0 \ \text{ for all } F \in \mathcal{F}, \quad \text{(c)}\ \ \mu(\bigcup_{i\in\mathbb{N}} F_i) = \sum_{i\in\mathbb{N}} \mu(F_i),$$

for $\{F_i\} \subset \mathcal{F}$ and all F_i pairwise disjoint. In most cases that we consider, $\mathcal{F}(\Omega)$ are the Borel sets of Ω, which we denote by $\mathcal{B}(\Omega)$. A measure $\mathbb{P}$ on Ω that additionally satisfies $\mathbb{P}(\Omega) = 1$ is called a **probability measure**. We write

$$\mu(A) = \int_A \mathrm{d}\mu \ \text{ for } A \in \mathcal{F}.$$

Special cases of measures that are of importance within this work are the **Lebesgue measure**, for that we write $\mathrm{d}\mu = \mathrm{d}x$, and the **Dirac measure** $\mathrm{d}\mu = \delta(x-a)\,\mathrm{d}x$, that has the property

$$\int_A \delta(x-a)\,\mathrm{d}x = 1 \quad \text{if and only if} \quad a \in A.$$

Therefore, the Dirac measure is a probability measure, while the Lebesgue measure is in general not normalized, and may be unbounded. For each con-

tinuous and bounded function $f \in C_b(\Omega)$ we write

$$\mu(f) = \int_\Omega f(x)\,\mathrm{d}\mu$$

and we say that this is the **measure of the function** f. Oftentimes, a measures can be written as $\mathrm{d}\mu = p(x)\,\mathrm{d}x$, where p is a continuous function.
Let $\{P_\varepsilon : \varepsilon > 0\}$ be a family of and P be a single probability measures on $(\Omega, \mathcal{F})$. We say that P_ε **converges weakly** to P if

$$\int_\Omega f\,\mathrm{d}P_\varepsilon \longrightarrow \int_\Omega f\,\mathrm{d}P$$

for all $f \in C_b(\Omega)$, where this convergence is in the real numbers. We then write $P_\varepsilon \rightharpoonup P$. If the convergence only holds for all $f \in C_c(\Omega)$, that means for test functions that have compact support, we say that P_ε converges weakly*, and we write $P_\varepsilon \rightharpoonup^* P$.
If we allow for negative values of a measure μ, that means $\mu : \mathcal{F} \longrightarrow \mathbb{R}$, we call this a **signed measure**. The set of all signed measures obviously form a vector space. The set of all finite signed measures form a Banach space, which is usually denoted by $\mathcal{M}(\mathcal{B})$. This is the dual space of $C_b(\Omega)$, the space of continuous and bounded functions on Ω. Hence the following theorem applies.

Theorem A.1. *(Banach-Alaoglu)*
Let X be a Banach space and X^* its dual space. Then, the unit ball $B^* \in X^*$ is weakly compact. If X is separable, it is also sequentially compact in the weak* topology. Hence, every bounded subset of X^* has a weak* convergent subsequence.*

Note that $C_b(\Omega)$ is a separable Banach space for $\Omega \subset \mathbb{R}^n$, hence a bounded set of measures has a convergent subsequence.

A.2 Probability theory

In the following, we rely on an intuitive understanding of the term probability and randomness. Let $\mathcal{A}$ denote the set of all possible outcomes of a random experiment. With each outcome (or event) $a_i \in \mathcal{A}$ we associate a number $p_i \in [0, 1]$ such that $\sum_i p_i = 1$. Repeating the experiment a large number of times, p_i is the expected relative frequency of the event a_i. We say that "a_i has the **probability** p_i".

Probability spaces and random variables

This concept can be formalized and extended to a continuous set of possible events as follows. A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

Ω	is the set of events,
$\mathcal{F}$	is a $\sigma-$algebra of subsets of Ω, and
$\mathbb{P}$	is a probability measure on Ω.

Here, Ω can be an arbitrary set, and $\mathcal{F} = \mathcal{F}(\Omega)$ is a $\sigma-$algebra as defined above. Given a subset $A \subset \Omega$, $\mathbb{P}(A)$ is then the probability, that the outcome of our random experiment is an element of A. Note that, if $\Omega \subset \mathbb{R}^n$, any single event has the probability 0, since it is a null set.
On a probability space we can define a **random variable**. That is a function $X : \Omega \longrightarrow \mathbb{R}$ such that

$$\{\omega \in \Omega : \ X(\Omega) \leq x\} \in \mathcal{F} \quad \text{for all } x \in \mathbb{R}.$$

Note that the image of a random variable does not need to be the entire real line but might be any subset of $\mathbb{R}$. Then, for a random variable X we define its **distribution function** as the function $F : \mathbb{R} \longrightarrow [0,1]$ with $F(x) = \mathbb{P}(X \leq x)$. Again, we also allow distribution functions to be defined on any subset of $\mathbb{R}$. Sometimes, we also use the term distribution function for

$$\tilde{F} : \mathbb{R} \longrightarrow [0,1], \quad \tilde{F}(x) = \mathbb{P}(X \geq x).$$

Obviously, it holds that $\tilde{F}(x) = 1 - F(x)$. The random variable X is called **continuous**, if the underlying distribution function can be expressed as

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u \quad \text{for all } x \in \mathbb{R}$$

for a measurable function $f : \mathbb{R} \longrightarrow [0, \infty)$, which is the **probability density (function)** of X. If f is smooth enough, then we have $F'(x) = f(x)$.

These concepts naturally extend to arbitrary subsets of $\mathbb{R}^n$. In that case we speak of a **random vector** instead of a random variable. A random vector X has a **joint distribution function** $F_X : \mathbb{R}^n \longrightarrow [0,1]$ such that $F_X(x) = \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}^n$. Note that the inequality is to be understood componentwise in this case.

Example 1. In Chapter 4 we treat the distance to the next collision in a homogenized stochastic medium as a random variable. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in this situation is defined by the following objects. The set of possible events is $\Omega = [0, \infty)$, and $\mathcal{F}(\Omega)$ is the set of all Borel sets of $[0, \infty)$. The probability measure $\mathbb{P}$ is characterized by the probability density. Consequently, the random variable under consideration is simply the identity $X(\omega) = \omega$ and has the distribution function

$$g(s) = \mathbb{P}(X(s) \leq s) = \int_0^s p(\tau)\,\mathrm{d}\tau = 1 - \int_s^\infty p(\tau)\,\mathrm{d}\tau,$$

which is a slight variation of the definition given above. Hence, we observe here that $g'(s) = -p(s)$.

Expectations and conditional expectations

The **expectation** of a random variable is the average of the possible values of X. For a continuous random variable X with density function f_X this is given by

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x,$$

if this integral exists. If X and $g(X)$ are continuous random variables, then the expectation of $g(X)$ is

$$\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x.$$

A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ **converges in distribution** to a random variable X, if the probability densities f_{X_n} converge weakly to f_X (see definition above). We then write $X_n \xrightarrow{D} X$. Therefore, convergence in distribution holds if and only if

$$\mathbf{E}[f(X_n)] \longrightarrow \mathbf{E}[f(X)] \qquad \text{for every } f \in C_b(\Omega).$$

Let us now consider a pair of random variables X and Y with distribution functions F_X and F_Y. They are called **independent** if and only if for their joint distribution function $F(x,y)$

$$F(x,y) = F_X(x)F_Y(y) \quad \text{for all } \; x,y \in \mathbb{R}$$

holds. This is equivalent to

$$f(x,y) = f_X(x)f_Y(y) \quad \text{for all } \; x,y \in \mathbb{R},$$

where $f(x,y)$ is the joint probability density of X and Y. Also, for a sufficiently smooth function $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ we have that

$$\mathbf{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Note that this implies that the expectation is a linear operator on random variables, i.e $\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y]$ for all $a, b \in \mathbb{R}$. The **conditional expectation** of Y given X is defined as

$$\mathbf{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\,\mathrm{d}y.$$

Here, $f_{Y|X}(y \mid x)$ is the conditional density function defined as

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)}.$$

Hence, for two independent random variables X and Y we have $f_{Y|X}(y \mid x) = f_Y(y)$. The conditional distribution function is defined in analogy. The conditional expectation has the important property, that $\mathbf{E}[\mathbf{E}[Y \mid X]] = \mathbf{E}[Y]$. This means especially that

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} \mathbf{E}[Y \mid X] f_X(x) \, \mathrm{d}x.$$

Stochastic processes

A **stochastic process** X is a family $\{X_t : t \in T\}$ of random variable indexed by some set T. Within this work, we always set $T = [0, \infty)$ and we think of t as being the time. That especially means, that we only consider continuous-time processes. The set of possible outcomes of X is called **phase space** in this context. A special and very important case are the so-called **Markov processes**, that are also called **Markov chains**. A process X satisfies the **Markov property**, if

$$\mathbb{P}(X_{t_n} = j \mid X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}) = \mathbb{P}(X_{t_n} = j \mid X_{t_{n-1}} = i_{n-1})$$

for all $j, i_1, \ldots, i_{n-1} \in \Omega$ and any sequence $t_1 < t_2 < \cdots < t_n$ of times. That means, that the current state of the process at time t_n only depends on the previous time t_{n-1} and is therefore independent of the others. We also say that the process has **no memory**. Let $X_0 = x$ be the initial state of the stochastic process. Then the entire set $X_t(x)$ is called a **trajectory** of X.

Example 2. We turn again to Chapter 4, where we describe the transport of a particle in a homogenized stochastic medium. The Markov process there is vector valued, i.e it has three components (X_t, Ω_t, S_t), where X_t is the position in space, Ω_t the direction of flight and S_t the distance to the next collision. The initial setting (X_0, Ω_0, S_0) can be treated deterministic (given by a fixed triple (x, Ω, s)) or stochastic (in this case the initial setting is given by a probability density $\phi_0(x, \Omega, s)$). Each component is a function of the two random variables ω and τ, where ω indicates the new direction of flight and τ the distance to collision after a collision event. These two random variables are independent of each other and are also independent of the past collision events. Furthermore, they do not depend on time. Hence, this process is Markovian in the phase space $\mathbb{R}^2 \times S^1 \times [0, \infty)$, since its evolution in time only depends on the last collision event, and the evolution between two collision events is deterministic. It is a well-known fact, that the reduced process (X_t, Ω_t) is Markovian in the phase-space $\mathbb{R}^2 \times S^1$ if and only if the distribution function of τ is an exponential. Furthermore, the process X_t alone is not Markovian. The method of extending the phase space in order to gain the Markov property of a stochastic process is sometimes referred to as **Markovian extension** or **Markovianization** [88].

Let $X_t(x), t \geq 0$ be a Markov process, where $X_t(x)$ denotes that $X_0 = x$. We

define the corresponding **Kolmogorov operator** P_t by

$$(P_t\phi)(x) = \mathbf{E}[\phi(X_t(x))],$$

for every bounded and continuous $\phi \in C_b(\mathbb{R})$. For all $t, s \geq 0$, P_t has the semigroup property, meaning

$$P_{t+s}\phi = P_t P_s \phi.$$

It is also obvious, that P_t is a linear operator, and $(P_0\phi)(x) = \phi(x)$ holds, hence, $P_0 = \mathrm{Id}$. Consider the function $u(x,t) = (P_t\phi)(x)$. Then $u(x,t)$ satisfies the initial value problem

$$\partial_t u(x,t) = \mathcal{L}u(x,t), \quad u(x,0) = \phi(x), \tag{A.1}$$

where $\mathcal{L}$ is another operator, which is called the infinitesimal generator of P_t. We do not want to go into the details of semigroup theory here. Note only, that $\mathcal{L}$ is in general a differential operator acting on a suitable (function) space. Equation (A.1) is called the **backwards Kolmogorov equation**. It governs the evolution of an **observable**

$$u(x,t) = \mathbf{E}[\phi(X_t(x))] = \mathbf{E}[\phi(X_t)|X_0 = x].$$

We can define the corresponding adjoint semigroup P_t^* which acts on probability measures. The operators P_t and P_t^* are adjoint in the following L^2-sense:

$$\int P_t\phi(x)\,\mathrm{d}\mu(x) = \int \phi(x)\,\mathrm{d}(P_t^*\mu)(x).$$

Define $\mu_t := P_t^*\mu$. The equation satisfied by μ_t is

$$\partial_t \mu_t = \mathcal{L}^*\mu_t, \quad \mu_0 = \mu.$$

This equation is called the **forward Kolmogorov** or **Fokker-Planck** equation of X. Assuming that μ_t can be written as $\mu_t = p(y,t)\,\mathrm{d}y$ and $\rho(y,0) = p_0(y)$ the equation reads

$$\partial_t p(y,t) = \mathcal{L}^* p(y,t), \quad p(y,0) = p_0(y).$$

If the initial state $X_0 = x$ is deterministic, the initial condition becomes $p_0(y) = \delta(y-x)$. The solution of the forward Kolmogorov equation describes the **evolution of the probability density** of X_t or the **evolution of states**. In other words, μ_t is the law of X_t. Coming back to the backward Kolmogorov equation, the solutions are connected as

$$u(x,t) = \mathbf{E}[\phi(X_t)|X_0 = x] = \int \phi(x)p(t,y;x)\,\mathrm{d}y.$$

Furthermore, due to the adjoint nature of the equations, we have

$$\int \phi \, \mathrm{d}\mu_t = \int P_t \phi \, \mathrm{d}\mu_0.$$

Therefore, we can write

$$\int \phi(x) p(t,x) \, \mathrm{d}x = \int u(x,t) p_0(x) \, \mathrm{d}x = \int \mathbf{E}[\phi(X_t(x))] p_0(x) \, \mathrm{d}x. \tag{A.2}$$

This can also be seen as a stochastic version of the characteristics formula.

Example 3. We revisit Chapter 4 for a last time. The stochastic process (X_t, Ω_t, S_t) has the Markov property, as we have seen above. Let $\phi = \phi(x, \Omega, s)$ be an observable of the process. The corresponding Kolmogorov operator is then

$$P_t \phi(x, \Omega, s) = \Phi(t, x, \Omega, s) := \mathbf{E}[\phi((X_t, \Omega_t, S_t)(x, \Omega, s))].$$

From the explicit knowledge of the trajectories of the process we can then show, that $\Phi(t, x, \Omega, s)$ satisfies the non-classical equation

$$\begin{aligned}
&\partial_t \Phi(t, x, \Omega, s) - \Omega \nabla_x \Phi(t, x, \Omega, s) + \partial_s \Phi(t, x, \Omega, s) = 0, \\
&\Phi(t, x, \Omega, 0) = \frac{1}{2\pi} \iint p(s) \Phi(t, x, \Omega, s) \, \mathrm{d}\Omega \, \mathrm{d}s, \\
&\Phi(0, x, \Omega, s) = \phi(x, \Omega, s).
\end{aligned}$$

Hence, the corresponding infinitesimal operator $\mathcal{L}$ is given by

$$\mathcal{L}\phi = \Omega \nabla_x \phi - \partial_s \phi + \delta(s) \frac{1}{2\pi} \iint p(s) \phi \, \mathrm{d}\Omega \, \mathrm{d}s.$$

Since this is a linear partial differential equation, the adjoint operator can be calculated by multiplying with a test function and partially integrating. Let the density of the process at time $t \geq 0$ be $\mu_t(x, \Omega, s)$. The adjoint operator $\mathcal{L}^*$ then acts on μ_t by

$$\mathcal{L}^* \mu_t(x, \Omega, s) = \partial_s \mu_t(x, \Omega, s) - \Omega \nabla_x \mu_t(x, \Omega, s) + p(s) \int_{S^1} \mu_t(x, \Omega, 0) \, \mathrm{d}\Omega.$$

Consequently, the corresponding forward Kolmogorov or Fokker-Planck equation of the Markov process (X_t, Ω_t, S_t) is given by

$$\partial_t \mu_t(x, \Omega, s) + \Omega \nabla_x \mu_t(x, \Omega, s) - \partial_s \mu_t(x, \Omega, s) = p(s) \int_{S^1} \mu_t(x, \Omega, 0) \, \mathrm{d}\Omega,$$

with initial density $\mu_0(x, \Omega, s)$. Equation (A.2) then reads

$$\iiint_{\mathbb{R}^2 \times S^1 \times \mathbb{R}_+} \mathbf{E}[\phi((X_t, \Omega_t, S_t)(x, \Omega, s))]\mu_0(x, \Omega, s)\, \mathrm{d}s\, \mathrm{d}\Omega\, \mathrm{d}x = \iiint_{\mathbb{R}^2 \times S^1 \times \mathbb{R}_+} \phi(x, \Omega, s)\mu_t(x, \Omega, s)\, \mathrm{d}s\, \mathrm{d}\Omega\, \mathrm{d}x.$$

Bibliography

[1] M.L. Adams, E.W. Larsen, *Fast iterative methods for discrete-ordinates particle transport calculations*, Progr. in Nucl. En., Vol. 40, No. 1, pp. 3-159 (2002).

[2] I.S. Anderson, R. McGreevy, H.Z. Bilheux, Neutron Imaging and Applications, Springer (2009).

[3] D. Aregba-Driollet, M. Briani, R. Natalini, *Time Asymptotic high order Schemes for dissipative BGK hyperbolic Systems*, Preprint, arxiv:1207.6279v1 (2012).

[4] R. Barnard, M. Frank, M. Herty, *Optimal radiotherapy treatment planning using minimum entropy models*, Appl. Math. Comput., Vol. 219, pp. 2668-2679 (2012).

[5] R. Barnard, M. Frank, M. Herty, *Treatment Planning Optimization for Radiotherapy*, Proc. Appl. Math. Mech., Vol. 13, pp. 339-340 (2013).

[6] J.K. Barstow, *From spectra to atmospheres: Solving the underconstrained retrieval problem for exoplanets*, Proc. Int. Astr Un., Vol. 8, pp. 275-276 (2013).

[7] P. Barthelemy, J. Bertolotti, D.S. Wiersma, *A Lévy flight for light*, Nature, Vol. 453, pp. 495-498 (2008).

[8] R. Beals, J. Szmigielski, *Meijer G-Functions: A Gentle Introduction*, Notices of the American Mathematical Society, Vol. 60, No. 7, pp. 866-872 (2013).

[9] R. Bellman, Numerical inversion of the Laplace transform: applications to biology, economics, engineering, and physics, Elsevier (1966)

[10] E. Bernard, E. Cagliotti, F. Golse, *Homogenization of the Linear Boltzmann Equation in a Domain with a Periodic Distribution of Holes*, SIAM J. Math. Analysis, pp. 2082-2113 (2010).

[11] C. Berthon, P. Charrier, B. Dubroca, *An HLLC scheme to solve the M1 model of radiative transfer in two space dimensions*, J. Sci. Comput., Vol. 31, No. 3, pp. 347-389 (2007).

[12] C. Berthon, C. Sarazin, R. Turpault, *Space-time GRP solvers of order k for linear advection with unrestricted time step*, J. Sci. Comp., Vol. 55, No. 2, pp. 268-308 (2013).

[13] C. Berthon, R. Turpault, *Asymptotic preserving HLL schemes*, Num. Meth. Part. Diff. Eq., Vol. 27, No. 6, pp. 1396-1422 (2011).

[14] J. Bertolotti, K. Vynck, L. Pattelli, P. Barthelemy, S. Lepri, D.S. Wiersma, *Engineering Disorder in Superdiffusive Lévy Glasses*, Advanced Functional Materials, Vol. 20, No. 6, pp. 965-968 (2010).

[15] F. P. Boca, A. Zaharescu, *The distribution of the free path lengths in the periodic two-dimensional Lorentz gas in the small-scatterer limit*, Commun. Math. Phys., Vol. 269, pp. 425-471 (2007).

[16] C. F. Bohren, E. E. Clothiaux, Fundamentals of Atmospheric Radiation, Wiley (2006).

[17] C.G. Broyden, *The Convergence of a Class of Double-rank Minimization Algorithms*, J. Inst. Maths. Applics., Vol. 6, pp. 76-90 (1970).

[18] C. Buet, S. Cordier, *An asymptotic preserving scheme for hydrodynamics radiative transfer models*, Num. Math, Vol. 108, No. 2, pp. 199-221 (2007).

[19] M. Burger, M. Hinze, R. Pinnau, *Optimization models for semiconductor dopant profiling*, in Transport Phenomena and Kinetic Theory, C. Cercignani, E. Gabetta (Eds.), Birkhäuser, pp. 90-116 (2007).

[20] V.A. Bykovskii, A.V. Ustinov, *The statistics of particle trajectories in the inhomogeneous Sinai problem for a two-dimensional lattice*, Izv. Ran. Ser. Mat., Vol. 73, pp. 17-36 (2009).

[21] E. Caglioti, F. Golse, *On the distribution of free-path lengths for the periodic Lorentz gas III*, Comm. Math. Phys., Vol. 236, pp. 199-221 (2003).

[22] E. Caglioti, F. Golse, *The Boltzmann-Grad limit of the periodic Lorentz gas in two space dimensions*, C. R. Acad. Sci. Paris, Vol. 346, pp. 477-482 (2008).

[23] E. Caglioti, F. Golse, *On the Boltzmann-Grad limit for the Two-Dimensional Periodic Lorentz Gas*, J. Statist. Phys., Vol. 141, pp. 264-317 (2010).

[24] R. F. Cahalan et al., *The I3RC - Bringing Together the Most Advanced Radiative Transfer Tools for Cloudy Atmospheres*, Bulletin of the American Meteorological Society, Vol. 86, No. 9, pp. 1275 - 1293 (2005).

[25] C. Cercignani, The Boltzmann Equation and its applications, Springer (1988).

[26] D. Chandler, Introduction to Modern Statistical Mechanics, Oxford University Press (1987).

[27] S. Chandrasekhar, *On the radiative equilibrium of a stellar atmosphere*, II. Astrophys. J., Vol. 100, pp. 76-86 (1943).

[28] S. Chandrasekhar, Radiative Transfer, Oxford University Press (1950).

[29] A. M. Cohen, Numerical methods for Laplace transform inversion, Springer (2007).

[30] A. Davis. A. Marshak, *Lévy kinetics in slab geometry: Scaling of transmission probability*, in Fractal Frontiers, M. M. Novak and T. G. Dewey (Eds.), World Scientific, Singapore, pp. 63-72 (1997).

[31] A.B. Davis, A. Marshak, K.P. Pfeilsticker, *Anomalous/Lévy Photon Diffusion Theory: Toward a New Parameterization of Shortwave Transport in Cloudy Columns*, Ninth ARM Science Team Meeting Proceedings, San Antonio, Texas, March (1999).

[32] A.B. Davis, A. Marshak, *Photon propagation in heterogeneous optical media with spatial correlations: enhanced mean-free-paths and wider-than-exponential free-path distributions*, J. Quant. Spec. Rad. Trans., Vol. 84, pp. 3-34 (2004).

[33] B. Davison, J.B. Sykes, Neutron Transport Theory, Oxford University Press (1957).

[34] N. Devaney, E. Thiebaut, *Inverse problem approach to the detection of exoplanets in multi-wavelength data*, Conference Paper, Adaptive Optics: Methods, Analysis and Applications, July (2011).

[35] B. Dubroca, J.-L. Feugeas, *Étude théorique et numérique d'une hiérarchie de modèles aux moments pour le transfert radiatif*, C. R. Acad. Paris, Vol. 329, No. 10, pp. 915-920 (1999).

[36] H. S. Dumans, L. Dumas, F. Golse, *Remarks on the Notion of Mean Free Path for a Periodic Array of Spherical Obstacles*, Journal of Statistical Physics, Vol. 87, No. 3-4, pp. 943-950 (1997).

[37] H. Egger, M. Schlottbom, *An L^p theory for stationary radiative transfer*, Applicable Analysis, Vol. 93, No. 6, pp. 1283-1296 (2014).

[38] H. Egger, M. Schlottbom, *Numerical methods for parameter identification in stationary radiative transfer*, Comput. Optim. Appl. (2014).

[39] J. Elstrodt, Maß- und Integrationstheorie, Springer (2009).

[40] W. Feller, An Introduction to Probability Theory and its Applications, Vol. 1, Wiley (1965).

[41] W. Feller, An Introduction to Probability Theory and its Applications, Vol. 2, Wiley (1971).

[42] R. Fletcher, *A New Approach to Variable Metric Algorithms*, Computer Journal, Vol. 13, pp. 317-332 (1970).

[43] M. Frank, *Fractional diffusion limit of non-classical transport*, in preparation (2015).

[44] M. Frank, T. Goudon, *On a generalized Boltzmann equation for non-classical particle transport*, Kinet. Relat. Models, Vol. 3, pp. 395-407 (2010).

[45] M. Frank, M. Herty, M.' Schäfer, *Optimal Treatment Planning in Radiotherapy Based On Boltzmann Transport Calculations*, Math. Mod. Meth. Appl. Sci., Vol. 18, pp. 573-592 (2008).

[46] M. Frank, M. Herty, A.N. Sandjo, *Optimal Radiotherapy Treatment Planning Governed by Kinetic Equations*, Math. Mod. Meth. Appl. Sci., Vol. 20, pp. 661-678 (2010).

[47] B. Fristedt, L. Gray, A Modern Approach to Probability Theory, Birkhäuser (1997).

[48] O. Funk, K. Pfeilsticker, *Photon path length distributions for cloudy skies - oxygen A-Band measurements and model calculations*, Ann. Geophys., Vol. 21, pp. 615-626 (2003).

[49] G. Gallavotti, *Rigorous Theory of the Boltzmann equation in the Lorentz gas*, Nota interna no. 358, Istituto di Fisica, Univ, di Roma, Available as preprint mp-arc-93-304 (1972).

[50] D. Goldfarb, *A Family of Variable Metric Updates Derived by Variational Means*, Mathematics of Computing, Vol. 24, pp. 23-26 (1970).

[51] F. Golse, S. Jin, C. D. Levermore, *The convergence of numerical transfer schemes in diffusive regimes I: The discrete ordinate method*, SIAM J. Numer. Anal., Vol. 36, pp. 1333-1369 (1999).

[52] F. Golse, *On the periodic Lorentz gas in the Boltzmann-Grad scaling*, Ann. Fac. des Sci. Toulouse, Vol. 17, pp. 735-749 (2008).

[53] F. Golse, *Recent Results on the Periodic Lorentz Gas*, Book Chapter in Nonlinear Partial Differential Equations, Springer (2012).

[54] L. Gosse, G. Toscani, *Asymptotic-preserving & well-balanced schemes for radiative transfer and the Rosseland approximation*, J. Num. Math., Vol. 98, No. 2, pp. 223-250 (2004).

[55] G. Grimmet, D. Stirzaker, Probability and Random Processes, Oxford University Press (2004).

[56] C. W. Groth, A. R. Akhmerov, and C. W. J. Beenakker, *Transmission probability through a Lévy glass and comparison with a Lévy walk*, Phys. Rev. E, Vol. 85, 021138 (2012).

[57] A. Harten, P. Lax, B. van Leer, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Review, Vol. 25, No. 1, pp. 35 – 61 (1983).

[58] G. Hellmund, M. Prokesova, and E. B. Vedel Jensen, *Lévy-based Cox point processes*, Adv. in Appl. Probab., Vo. 40, No. 3, pp. 603-629 (2008).

[59] M. Hinze, R. Pinnau, M. Ulbrich, S. Ulbrich, Optimization with PDE constraints, Springer (2009).

[60] K. Ivanova, T.P. Ackerman, *Multifractal characterization of liquid water in clouds*, Phys. Rev. E, Vol. 59, 27781782 (1999).

[61] E. T. Jaynes, *Information Theory and Statistical Mechanics*, Phys. Rev., Vol. 106, pp. 620-630 (1957).

[62] S. Jin, *Efficient Asymptotic-Preserving (AP) Schemes For Some Multiscale Kinetic Equations*, SIAM J. Sci. Comp, Vol. 21, pp. 441-454 (1999).

[63] R. Klages, G. Radons, I.M. Sokolov, Anomalous Transport, Wiley (2008).

[64] A. Klose, A. Hielscher, *Optical tomography using the time-independent equation of radiative transfer Part 2: inverse model*, J. Quant. Spec. Rad. Trans., Vol. 72, No. 5, pp. 715-732 (2002).

[65] A. Klose, U. Netz, J. Beuthan, A. Hielscher, *Optical tomography using the time-independent equation of radiative transfer Part 1: forward model*, J. Quant. Spec. Rad. Trans., Vol. 72, No. 5, pp. 691-713 (2002).

[66] A.B. Kostinski, *On the Spatial Distribution of Cloud Particles*, J. Atmos. Sci., Vol. 57, No. 7, pp. 901-915 (2000).

[67] A.B. Kostinski, *On the Extinction of Radiation by a Homogeneous but Spatially Correlated Random Medium*, J. Optical Society of America, Vol. 18, No. 8, pp. 1929-1933 (2001).

[68] A.B. Kostinski, *On the extinction of radiation by a homogeneous but spatially correlated random medium: reply to comment*, J. Optical Society of America,Vol. 19, No. 12, pp. 2521-2525 (2002).

[69] A.B. Kostinski, R.A. Shaw, *Scale-dependent droplet clustering in turbulent clouds*, J. Fluid Mech., Vol. 434, pp. 389-398 (2001).

[70] A.B. Kostinski, R.A. Shaw, *Possible implications of droplet clustering for radiative transfer in clouds*, Extended abstract in the Proceedings of 11th Conference on Cloud Physics, Ogden, UT, June (2002).

[71] K. Krycki, C. Berthon, M. Frank, R. Turpault, *Asymptotic preserving numerical schemes for a non-classical radiation transport model for atmospheric clouds*, Math. Meth. Appl. Sci., Vol. 36, No. 16 pp. 2101-2116 (2013).

[72] B. Lapeyre, E. Pardoux and R. Sentis, Introduction to Monte-Carlo Methods for Transport and Diffusion Equations, Oxford Texts in Applied and Engineering Mathematics, Oxford University Press (2003)

[73] E.W. Larsen, *A generalized Boltzmann equation for "non-classical" particle transport*, Joint Int. Meeting on Math. Comp. & Supercomp. in Nucl. App., Monterey, California, on CD-ROM, Am. Nuc. Soc., LaGrange Park, IL (2007).

[74] E. W. Larsen, J. E. Morel, W. F. Miller Jr., *Asymptotic Solutions of Numerical Transport Problems in Optically Thick, Diffusive Regimes*, J. of Comp. Phys, Vol. 69, pp. 283-324 (1987).

[75] E. W. Larsen, J. E. Morel, J. M. McGhee, Asymptotic Derivation of the Multigroup P_1 and Simplified P_N Equations with Anisotropic Scattering, Nucl. Sci. Eng., Vol. 123, pp. 328-342 (1996).

[76] E. W. Larsen, R. Vasques, *A generalized linear Boltzmann equation for non-classical particle transport*, J. Quant. Spec. Rad. Trans., Vol. 112, No. 4, pp. 619-631 (2011).

[77] A. Leitao, P.A. Markowich, J.P. Zubelli, *Inverse problems for semiconductors: models and methods*, in: Transport Phenomena and Kinetic Theory, Eds.: C. Cercignani, E. Gabetta, Birkhäuser, pp. 117-149 (2007).

[78] C. D. Levermore, *Moment closure hierarchies for kinetic theories*, J. Statist. Phys., Vol. 83, pp. 1021-1065 (1996).

[79] H.A. Lorentz, *The motion of electrons in metallic bodies I*, Proc. Roy. Acad. Amst., Vol. 7, pp. 438-453 (1905).

[80] S. Lovejoy, D. Schertzer, *Multifractals, universality classes and satellite and radar measurements of cloud and rain fields*, J. Geophys. Res., Vol. 95, pp. 2021-2034 (1990).

[81] J. Marklof, A. Strömbergsson, *The Boltzmann-Grad Limit of the Periodic Lorentz Gas*, Annals of Mathematics, Vol. 174, pp. 225-298 (2011).

[82] J. Marklof, A. Strömbergsson, *Free Path Lengths in Quasicrystals*, Comm. Math. Phys., Vol. 330, No. 2, pp. 723-755 (2014).

[83] J. Marklof, A. Strömbergsson, *Generalized linear Boltzmann equations for particle transport in polycrystals*, preprint (2015).

[84] A. Marshak, A. Davis, W.J. Wiscombe, R.F. Cahalan, *Scale-invariance of liquid water distributions in marine stratocumulus, Part 2: Multifractal properties and intermittency issues*, J. Atmos. Sci., Vol. 54, 14231744 (1997).

[85] A. Marshak, A. Davis, 3D Radiative Transfer in Cloudy Atmospheres, Springer (2005).

[86] N. Metropolis et al., *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys., Vol. 21, p. 1087 (1953).

[87] G. N. Minerbo, *Maximum entropy Eddington factors*, J. Quant. Spectrosc. Radiat. Transf., Vol. 20, pp. 541-545 (1978).

[88] G.A. Pavliotis, Applied Stochastic Processes, Lecture Notes, Imperial College London, UK, January (2011).

[89] K. Pfeilsticker, *First geometrical path length probability density function derivation of the skylight from high-resolution oxygen A-band spectroscopy: 2. Derivation of the Lévy index for the skylight transmitted by midlatitude clouds*, J. of Geophys. Res., Vol. 104, pp. 4101-4116 (1999).

[90] J.C. de los Reyes, Numerical PDE-Constrained Optimization, Springer (2015).

[91] C. Sarazin, Méthodes numériques pour des systèmes hyperboliques avec terme source provenant de physiques complexes autour du rayonnement, PhD Thesis, Université de Nantes (2013).

[92] M. Schäfer, M. Frank, M. Herty, *Optimal treatment planning in radiotherapy based on Boltzmann transport calculations*, Proc. Appl. Math. Mech., Vol. 7, pp. 2627-2628 (2007).

[93] A. T. Schmitt, Fast iterative methods for non-classical particle transport, Master's Thesis, RWTH Aachen University (2015).

[94] T. Scholl, K. Pfeilsticker, A.B. Davis et al., *Path length distributions for solar photons under cloudy skies: Comparison of measured first and second moments with predictions from classical and anomalous diffusion theories*, J. Geophys. Res., Vol. 111, No. D12, pp. 2156-2202 (2006).

[95] V. Scipolo, *Scattered Neutron Tomography Based on a Neutron Transport Problem*, PhD Thesis, Texas A & M University (2004).

[96] M. Seaid, A. Klar, *Efficient Preconditioning of Linear Systems Arising from the Discretization of Radiative Transfer Equation*, Challenges in Scientific Computing, Berlin 2002, Lecture Notes in Computational Science and Engineering, 35, pp. 211-236 (2003).

[97] D.F. Shanno, *Conditioning of Quasi-Newton Methods for Function Minimization*, Mathematics of Computing, Vol. 24, pp. 647-656 (1970).

[98] R.A. Shaw, A.B. Kostinski, M.L. Larsen, *Towards quantifying droplet clustering in clouds*, Q. J. R. Meteorol. Soc., Vol. 128, pp. 1043-1057 (2002).

[99] Y.G. Sinai, *Dynamical systems with elastic reflections*, Russ. Math. Surv., Vol. 25, pp. 137-189 (1970).

[100] H. Spohn, Large Scale Dynamics of Interacting Particles, Springer (1991).

[101] T. Svensson, K. Vynck, *Holey Random Walks: Optics of Heterogeneous Turbid Composites*, Phys. Rev. E, Vol. 87, 022120 (2013).

[102] F. Tröltzsch, Optimal Control of Partial Differential Equations, AMS (2005).

[103] R. Vasques, Anisotropic Diffusion of Neutral Particles in Stochastic Media, PhD Thesis, University of Michigan, Ann Arbor (2009).

[104] R. Vasques, K. Krycki, *Solving the 1D non-classical transport equation in random periodic media*, Proceedings of the Joint International Conference on Mathematics and Computation, Nashville (TN), April (2015).

[105] Z. Wu, Advances in Inverse Transport Methods and Applications to Neutron Tomography, PhD Thesis, Texas A & M University (2010).