Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Nils W., Schween

born in: Korbach, Germany

Oral examination: October 16th, 2024

Mathematical methods for modelling the transport of energetic particles in tenuous astrophysical plasmas

Referees: Dr. Brian Reville
Theorie astrophysikalischer Plasmen (TAP)
Max-Planck-Institut für Kernphysik

Prof. Dr. Cornelis P. Dullemond
Institut für Theoretische Astrophysik (ITA)
Universität Heidelberg

Supervisors: Dr. Brian Reville, Prof. John Kirk

Abstract

The acceleration and propagation of charged and energetic particles in tenuous and non-relativistically moving astrophysical plasmas is modelled with a variant of the Vlasov–Fokker–Planck equation. The distribution function of the particles is expanded in Cartesian tensors or spherical harmonics. The expansion leads to a system of partial differential equations (PDE) that determines the expansion coefficients. In this PhD thesis we derive new formulae to convert between the expansion coefficients of the Cartesian and spherical harmonic expansions, irrespective of the expansion order. These formulae are equally valid for the Cartesian and spherical multipole expansions of the electrostatic (or gravitational) potential. Moreover, we present a novel way to derive the system of PDEs that is based on operators that act in the Hilbert space of spherical harmonics and their representation matrices. The system of PDEs gained with the operators is a system of advection-reaction equations, that we numerically solve with the discontinuous Galerkin (dG) method. We test our implementation of the dG method to demonstrate that the numerical algorithm is robust. Applying it to simulate the acceleration of charged particles at a parallel shock wave shows that our implementation is suited to simulate astrophysical applications.

Zusammenfassung

Die Beschleunigung und Ausbreitung geladener und hochenergetischer Teilchen in nicht-relativistischen, astrophysikalischen Plasmen mit geringer Dichte wird mit einer Spielart der Vlasov–Fokker–Planck Gleichung modelliert. Die Verteilungsfunktion der Teilchen wird nach kartesichen Tensoren oder nach Kugelflächenfunktionen entwickelt. Die Entwicklungskoeffizienten werden durch ein System partieller Differentialgleichungen (PDG) bestimmt. In dieser Doktorarbeit werden neue Formeln zum umrechnen zwischen den Entwicklungskoeffizienten der Entwicklung nach kartesichen Tensoren und der nach Kugelflächenfunktionen hergeleitet, unabhängig davon bis zur welchen Ordnung entwickelt wurde. Die Formeln sind ebenfalls gültig für die kartesische und sphärische Multipolentwicklung des elektrostatischen Potentials (oder des Gravitationspotentials). Des Weiteren zeigen wir, dass sich das System der PDGs mit Hilfe von Operatoren, die im Hilbertraum der Kugelflächenfunktionen definiert sind, und ihren Darstellungsmatrizen herleiten lässt. In der Perspektive dieser neuen Methode zeigt sich, dass das System der PDGs ein System von Konvektions-Reaktions-Gleichungen ist. Dieses System lösen wir numerisch mit der diskontinuierlichen Galerkin (dG) Methode. Wir testen unsere Implementierung der dG Methode und demonstrieren, dass sie numerisch robust ist. Eine Simulation der Beschleunigung geladener Teilchen in einer parallelen Stoßwelle zeigt, dass unsere Implementierung für astrophysikalische Anwendungen geeignet ist.

Acknowledgements
I thank Dr. Brian Reville for his trust, open-mindedness and patience. Our on going dialogue about plasma and astrophysics and our joint computation sessions make me reflect my presumptions, inspire me and deepen my physical understanding of the models that we are fighting with. That is awesome. I am also very grateful to Prof. John Kirk who helps me with my computations, reads my manuscripts, answers many of my questions and engages me in enlightening discussions. I thank Prof. Dr. Cornelis P. Dullemond very much for being part of my thesis committee and for refereeing my PhD thesis.

The Astrophysical Plasma Theory group is a team and I am indebted to everyone. I thank Florian Schulze with whom I enjoy developing `Sapphire++`, Lucia Härer, Thibault Vieu and Cormac Larkin. They teach me astrophysics and create a unique atmosphere of friendliness and openness. I also thank the former APT members Dr. Michelle Tsirou whose love for astrophysics I admire, Dr. Makarim Bouyahiaoui, Dr. Zhiqiu Huang, Dr. Jieshuang Wang, Dr. Grigorios Katsoulakos, Dr. Gwenael Giacinti and Dr. Naveen Kumar. I also thank Yu Fung Wong who began his PhD with me and who will finish it with me.

I thank Prof. Jim Hinton whose Division offers endless opportunities for collaboration and exchange.

Scientifically I am very indebted to Dr. Philipp Gerstner, who introduced me to the finite element method, helped me with the mathematics and spent many hours discussing with me how to apply the discontinuous Galerkin method to the Vlasov–Fokker–Planck equation.

I thank Dr. Simon Sailer whose kindness helped me out of the many troughs I went through. I thank Leon Leuser, who is an exceptionally good friend and keeps on calling me, Dr. Ayla Nawaz whose friendship is priceless and Dr. Keno Riechers who accompanies me since my first encounter with physics at the University of Hamburg, supports me and shares my passion for rigour.

I do thank my wife Hella and my son Milan whose patience with my obligations was endless and whose support was incredible. I also thank Hella's parents Hans Pfeffer and, in particular, Stefanie Dilling who took care of Milan during my absence. I thank my mother Heidemarie Hesse-Schween and my father Heinrich Schween for their enduring love and support. I thank my brothers Jan and Lars Schween for the many hours we spent together.

To Hella, Milan and Johann

CONTENTS

1 Introduction **1**

2 Plasma models and their physical assumptions **9**
2.1 Kinetic description 9
2.1.1 Microscopic particle number density 10
2.1.2 Klimontovich equation 16
2.1.3 The Boltzmann and the Vlasov equation 18
2.1.4 BBGKY hierarchy 23
2.1.5 The Vlasov–Fokker–Planck equation 31
2.2 Magnetohydrodynamics 36
2.2.1 Intrinsic velocity 37
2.2.2 Intrinsic velocity moments 39
2.2.3 Fluid equations 41
2.2.4 Ideal MHD equations 45

3 Particle transport in tenuous astrophysical plasmas **49**
3.1 The Fermi acceleration process 51
3.1.1 Fermi's thoughts on the origin of the cosmic radiation . . 51
3.1.2 Particle acceleration at parallel shocks 56
3.2 The semi-relativistic VFP equation 67
3.2.1 Alfvén waves 68
3.2.2 Collision operator 71
3.2.3 Mixed coordinates 80
3.2.4 Transport equation 86
3.3 The cosmic-ray transport equation 87
3.3.1 Solution of the cosmic-ray transport equation 91

4 Cartesian tensor and spherical harmonic expansion **96**
4.1 Definition of the multipole moments 99
4.1.1 Definition of the Cartesian multipole moments 99
4.1.2 Definition of the spherical multipole moments 103
4.2 Efimov's ladder operator 105
4.2.1 Alternative definition of the multipole moments 105
4.2.2 Equivalence of the definitions 107

4.2.3 Preliminary basis transformation . . . 109
4.3 Basis transformation . . . 112
4.4 Inverse transformation . . . 114
4.5 Application to the VFP equation . . . 117

5 System of PDEs for the expansion coefficients 119
5.1 Operators and the system of PDEs . . . 119
5.1.1 The identity operator and the collision operator . . . 120
5.1.2 The angular momentum operator . . . 121
5.1.3 The direction operators . . . 123
5.1.4 Products of operators . . . 126
5.1.5 The complete system of equations . . . 128
5.2 Rotations in the spherical harmonic space . . . 129
5.2.1 The rotation operator and its matrix representation . . . 130
5.2.2 Rotated operators and their representation matrices . . . 131
5.2.3 Rotations of representation matrices . . . 132
5.3 A real system of equations . . . 133
5.3.1 Real spherical harmonics . . . 134
5.3.2 Turning the system of PDEs into a real system . . . 135
5.4 Eigenvectors and eigenvalues . . . 136
5.4.1 General statements . . . 137
5.4.2 Eigenvalues of the direction operators' matrices . . . 139
5.4.3 Sums of representation matrices . . . 141

6 Numerical solution of the system of PDEs 144
6.1 Discontinuous Galerkin . . . 145
6.1.1 Advection-reaction equation . . . 145
6.1.2 Discrete representation of the solution . . . 147
6.1.3 Numerical flux . . . 149
6.1.4 Numerical flux at the boundaries of the domain . . . 153
6.1.5 Time stepping method . . . 154
6.2 Tests and Simulations . . . 156
6.2.1 Convergence study . . . 157
6.2.2 Advection in a constant magnetic field . . . 162
6.2.3 Diffusive Shock Acceleration at a Parallel shock . . . 165

7 Concluding remarks 174

A Cartesian tensors and spherical harmonics 177
A.1 Notes on the definition of the Cartesian multipole moments . . . 177
A.1.1 The multipole moments are tensors . . . 177
A.1.2 The $\hat{P}$ operator . . . 177
A.1.3 $c_{l,k}$ coefficients . . . 179
A.2 Commutator of the ladder operators . . . 182

A.3 Direct derivation of the inverse transformation 182

B PDEs for the expansion coefficients **187**
B.1 Real representation matrices 187

C Numerical solution of the system of PDEs **191**
C.1 Definition of the real spherical harmonics 191
C.2 Higher order corrections 192

Chapter 1

Introduction

Hilbert spaces are ubiquitous in physics. Quantum Mechanics is probably the best known example. In this thesis we show that Hilbert spaces also play an important role in modelling the transport of charged particles in plasmas. Their vector space structure and their scalar product provide us with tools that we use to develop new methods that relate and simplify the relevant equations and deepen our theoretical understanding of them. Moreover, we demonstrate how these new insights can be used beneficially when solving the equations numerically.

But what are these equations, what are they about and why is it important to solve them? The equations and the methods we developed are located in a historical and current cluster of ideas, concepts and models that are centred about a more than one hundred years old discovery, namely the discovery of cosmic rays. Around 1900 Elster[1] and Geitel[2] noted that air was conductive and they deduced that the conductivity was due to the presence of positive and negative ions. The air was ionised by a radiation of unknown nature. In 1913, Rutherford concluded from experiments that the 'very weak activity [i.e. ionisation rate] actually is in all probability to be ascribed to the presence of traces of radioactive matter [...]'. An essentially correct explanation for the larger part of the observed ionisation rate, whence it follows that the ionising radiation are α-particles, electrons or γ-rays. (Hillas, 1972, Sec. 1.1)

However, one year earlier in 1912, Hess undertook seven balloon flights reaching an altitude of about five kilometres during which he measured a significant increase of the ionisation rate with altitude. This observation led him to conjecture that there exists an ionising radiation penetrating the atmosphere from above. In 1914 Kolhörster confirmed Hess' findings and extended them to even higher altitudes. In Table 1 the average difference between the ionisation rate at various heights and that at sea level is shown. The decrease in the

[1] Julius Elster, * 24.12.1854 Blankenburg (Harz), † 8.4.1920 Wolfenbüttel. E. studied natural sciences, with a focus on physics in Berlin and Heidelberg. He received his PhD in 1879 in Heidelberg. (Ottmer, 1959)

[2] Hans Geitel, * 16.7.1855 Braunschweig, † 15.8.1923 Wolfenbüttel. G. and J. Elster became friends in their childhood and stayed close friends until they died. As Dioscuri they studied under R. Bunsen and G. R. Kirchhoff in Heidelberg and the physicist G. H. Quincke in Berlin. (Weiser, 1964)

Table 1.1: Average difference between the ionisation rates at various heights and that at sea level. (Hillas, 1972, Tab. 1.1)

Altitude [km]	Difference [ions cm^{-3} s^{-1}]
1	−1.5
2	+1.2
3	+4.2
4	+8.8
5	+16.9
6	+28.7
7	+44.2
8	+61.3
9	+80.4

ionisation rate at one km is in agreement with the hypothesis that the observed ionisation at sea level is mainly due to radioactive elements, whereas the strong increase in the ionisation rate with altitude underpins Hess' conclusion that a part of the ionisation (also at sea level) is a consequence of an ionising radiation arriving from space. (Hillas, 1972, Sec. 1.2)

There were still physicists not convinced of the existence of a radiation coming from space, namely and famously Millikan. However, he and Cameron investigated the ionisation inside the water of snow-fed lakes in California, which were thought to be free from radioactive elements. Their electrometres showed an ionisation whose rate decreased with depth, which strongly indicates a radiation travelling downwards through the air with no local generation. It was Millikan who subsequently introduced the name 'cosmic rays' and we nowadays define a *cosmic ray* to be a charged particle that was accelerated somewhere in the universe and that hits the atmosphere of the earth.[3] (Hillas, 1972, Sec. 1.3)

Having established the existence of cosmic rays, two questions come to one's mind: Firstly, what are the constituents of the radiation, i.e. what is the nature of the radiation? Secondly, what is the origin of the cosmic rays? Question number one could be partly settled in the course of the previous century. Today it is known that cosmic rays are particles, and not, as previously believed, γ-rays producing secondary electrons via the Compton effect (Bothe & Kolhörster, 1929). The *primary cosmic rays*, i.e. the cosmic rays arriving at the earth's atmosphere before interacting with it, are composed of mainly protons (about 90 %), α-particles (about 9 %) and small fractions of heavier nuclei (Gaisser et al., 2016, Sec. 1.1). The question of the origin of cosmic rays can be interpreted twofold: Firstly, where do the cosmic rays come from and, secondly, how do they acquire such high energies? Both interpretations form the framework of

[3]This definition is not capturing all usages of the term 'cosmic ray'. It is very often used synonymously with charged particle. Others also include γ-rays and sometimes electrons/positrons are excluded.

our research and, thus, of this PhD thesis. The equations with which we work model the propagation and acceleration of charged particles, for example cosmic rays, and because cosmic rays are extremely relativistic particles, they do so in the vicinity of the most energetic sources of the universe. Hence, solving the equations is not only an attempt to answer both questions, but also a mean to learn about these sources.

In addition to this general motivation, we have a specific research question in mind that is best explained with the help of a plot of the cosmic-ray energy spectrum, see Fig. 1.1. The energy spectrum allows for a gross categorisation in terms of where the cosmic rays are coming from: At energies around 1 GeV/nucleon the overall cosmic ray flux is modulated by charged particles originating in the sun, particles with energies up to $\sim 3 \times 10^6$ GeV are assumed to be accelerated in the Galaxy and cosmic rays with energies larger than $\sim 10^9$ GeV are believed to be of extragalactic origin. The corresponding breaks in the energy spectrum are referred to as the 'knee' and the 'ankle' respectively. (Gaisser et al., 2016, Ch. 2)

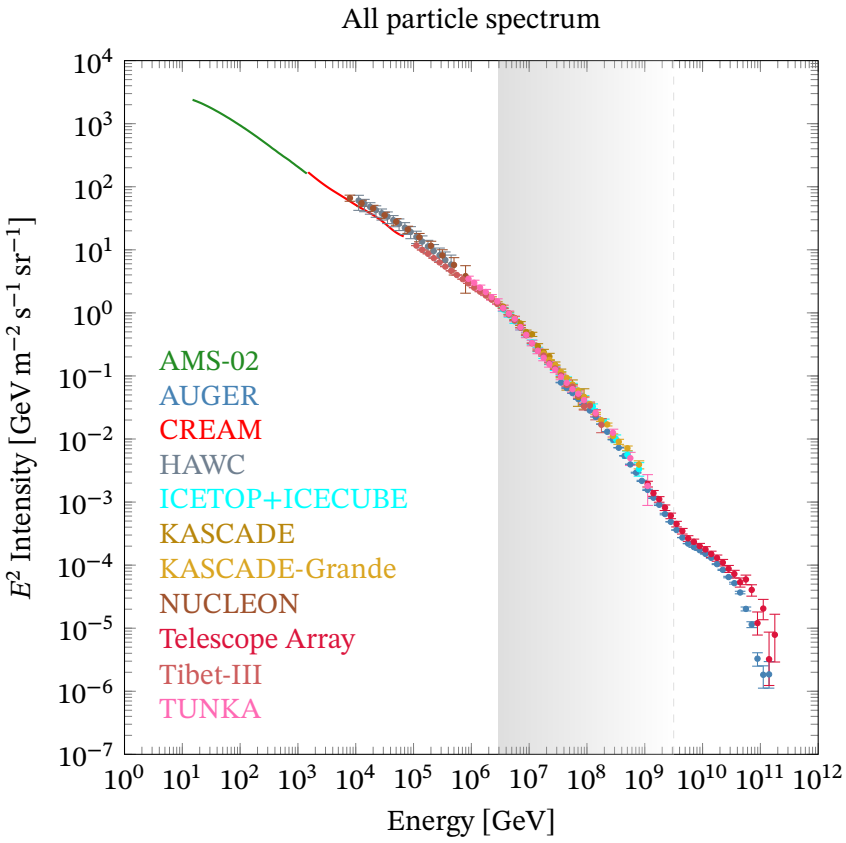

Figure 1.1: The all particle cosmic-ray spectrum. (Evoli, 2020)

Supernova remnants (SNR) were suggested early as possible acceleration

sites (Baade & Zwicky, 1934) and are now widely expected to contribute the bulk of the Galactic cosmic rays (Pasquale, 2013, Sec. 1.1.). The 'knee' could mark the beginning of an energy range that represents the maximum possible energies that particles can reach in our Galaxy, see the grey-shaded area in Fig. 1.1. We use *Hillas limit* (Hillas, 1984, cf. eq. 1) to estimate the maximum energy of a particle accelerated in a SNR shock wave, namely

$$E_{\max} \leq qU_S B_G R_{\mathrm{SNR}} \,.$$

If we assume that the larger part of the cosmic rays is accelerated when the radius of the SNR shock wave is about $R_{\mathrm{SNR}} \sim 10\,\mathrm{pc}$ and that $B_G \sim 0.1\,\mathrm{nT}$ and $U_s \sim 5000\,\mathrm{km\,s^{-1}}$ are typical values for the Galactic magnetic field and the speed of a SNR respectively (Blandford & Eichler, 1987, Sec. 2.5), then the estimate for the maximum energy gives

$$E_{\max} \leq 1.6 \times 10^5\,\mathrm{GeV}\,,$$

which is below the 'knee'. However, this estimate is now believed to be too low, because it is expected that the accelerated particles amplify the magnetic field around the shock wave (e.g. Gaisser et al., 2016, Sec. 12.5). This believe inspires the research question we have in mind, namely 'How does the non-linear interaction between the particles and the plasma, in which they are accelerated, influences the acceleration process and the maximum energy the particles can reach?' We couple a kinetic particle model, a variant of the Vlasov–Fokker–Planck (VFP) equation and a plasma model, ideal magnetohydrodynamics (MHD), to answer this question quantitatively. The focus of this PhD thesis is on the kinetic particle model, i.e. on the development of a robust numerical algorithm to solve the VFP equation.

The thesis consists of two parts: Chapter 2 and 3 provide the reader with concepts, definitions and equations that form the background in front of which we develop our methods. The actual research is presented in Chapters 3–6. Readers familiar with kinetic plasma models, MHD and particle transport in astrophysical plasmas may skip the first part and proceed directly with the second part.

The aim of Chapter 2 is to introduce the different plasmas models that we use throughout this thesis and the physical and mathematical assumptions underlying them. In its first section, we use non-equilibrium statistical mechanics to derive a kinetic plasma model, namely the Boltzmann equation. We adopt Klimontovich's approach and start with a microscopic particle density, i.e. a sum over delta distributions that determine the positions and velocities of all particles. The fact that the number of particles must be conserved leads to the Klimontovich equation. Because the actual positions and velocities of all particles are infeasible to track, we proceed as usual in statistical mechanics and presuppose a smooth probability density function that encodes the likelihood to find a specific particle with specific phase-space coordinates $\boldsymbol{x}$ and $\boldsymbol{v}$. We

compute the average of the Klimontovich equation with this probability density function and use that particles of the same plasma species are indistinguishable from each other to reduce the average to a partial differential equation (PDE) that evolves the one particle distribution function. It is this PDE that is called the Boltzmann equation and the interactions between the particles are symbolically represented by a collision term on its right-hand side. We review three different ways to derive explicit expressions for the collision term, namely correlation functions, scattering cross sections and, most importantly, the Fokker–Planck theory. In the second section of this chapter, we compute the moments of the Boltzmann equation to derive the MHD equations. We assume that there is no heat flow, i.e. all changes of state are adiabatic, to close the chain of moment equations. Additionally, we simplify Ohm's law and presume that the conductivity of the plasma is infinite, which yields the ideal MHD equations.

Chapter 3 has two purposes: Firstly, it presents the Fermi acceleration mechanism and, secondly, it introduces a particle transport equation that models the propagation and acceleration of relativistic particles in a background plasma that moves non-relativistically. We summarise Fermi's original paper and apply his arguments to the acceleration of particles at a parallel shock front. The idea is to provide the reader with a physically intelligible picture of the acceleration process, before we proceed with employing the Boltzmann equation to capture it more formally. Whereas in Chapter 2 the Boltzmann equation was used to model a plasma, we now utilise it to model the charged particles, which are accelerated, and their interaction with the background plasma. This means, in particular, that the Lorentz force that acts on the charged particles is not generated by themselves, but is due to the electromagnetic fields sustained by the background plasma. Moreover, the collision term quantifies the changes in the distribution of the charged particles, because of their interactions with the waves and the electromagnetic turbulence of the background plasma. We derive an explicit expression for the collision term using Fokker–Planck theory. The Lorentz force and the Fokker–Planck collision term allow us to change the name of the Boltzmann equation to Vlasov–Fokker–Planck equation. The collision term has a distinctly simple form in the rest frame of the waves that scatter the particles. Therefore, we use a mixed-coordinate frame, i.e. we use a Lorentz transformation to get the momentum coordinates in the rest frame of the background plasma, which is assumed to be an approximation to the rest frame of the waves, while leaving the spatial coordinates unchanged. We show how the VFP equation changes under such a transformation and we, subsequently, use that we are interested in background plasmas that move slowly in comparison to the speed of light to further simplify it. We call the resulting equation the semi-relativistic VFP equation, because it describes the transport of relativistic particles in a slowly moving plasma. It is the starting point of our research. At the end of this chapter, we apply it to the acceleration of particles at parallel shock. This yields the well-known cosmic-ray transport equation.

Numerically solving the VFP equation is computationally expensive, because

the single particle distribution function depends on t, $\boldsymbol{x}$ and $\boldsymbol{p}$, i.e. it depends on six variables plus time. A way to reduce the dimensionality of the problem is to use spherical coordinates in momentum space and to write the distribution function as a series of spherical harmonics. This separates out the angular variables and reduces the number of independent variables to four, namely $\boldsymbol{x}$ and p, where p is the magnitude of the momentum. The spherical harmonic expansion is the second chamber of the heart of this thesis, if we think of the semi-relativistic VFP equation as its first one. The spherical harmonics are a basis of the space of square-integrable functions whose domain is the sphere and whose codomain are the complex numbers, commonly denoted with $L^2(S^2)$, where S^2 is the unit sphere. Together with the usual scalar product this space is a Hilbert space.

In Chapter 4 we begin to exploit the vector space structure of the spherical harmonics. The spherical harmonic expansion is not the only possibility to separate out the angular dependence of the single particle distribution function. A second option is the Cartesian tensor expansion. How two convert between the coefficients of the two expansions for an arbitrary expansion order is an open question. However, we succeeded in using the vector space structure to derive formulae that convert between the expansion coefficients. The key idea is that both expansions represent the same function in $L^2(S^2)$, but they use a different basis to do so. The relation between the expansion coefficients can thus be represented as a basis transformation. That this is the case is not obvious. To show it, we take a detour and investigate an analogue situation, namely that of the multipole expansion which is used to approximate the potential of a charge (or mass) distribution. There also exist two variants of it: The expansion coefficients are either the (Cartesian) multipole moments, namely monopole, dipole, quadrupole etc. , or the spherical multipole moments and both again represent the same function in $L^2(S^2)$ and, hence, the relation between the two multipole moments is the same as the relation between the expansion coefficients of the Cartesian and spherical harmonic expansion of the distribution function. However, there exist explicit definitions of the Cartesian and spherical multipole moments that both use the spherical harmonics, though disguised as homogeneous and harmonic polynomials. We use these definitions to compute the basis transformation, i.e. we obtain formulae which allow us to systematically convert between the Cartesian and spherical multipole moments irrespective of the order of the expansion. These formulae and the way in which we derived them are, to the best of our knowledge, a novelty. Eventually, we apply these formulae to the expansions of the single particle distribution function. A highlight of this chapter are the two definitions of the Cartesian multipole moments: We introduce a new one that is based on the Kelvin transform and that allows us to identify the basis of $L^2(S^2)$ that is used for the Cartesian multipole expansion. The second definition uses Efimov's ladder operators, which serve us as a tool to explicitly compute the basis transformation.

Operators form the essence of Chapter 5. Using a spherical harmonic ex-

pansion implies that the direct computation of the distribution function of the charged particles is replaced with solving a system of PDEs that determines the expansion coefficients. We derive this system in a yet unknown way, namely we show that it is built of representation matrices of operators that act in the space of spherical harmonics. We start with plugging the spherical harmonic expansion into the semi-relativistic VFP equation and, thereafter, we take the scalar product of the resulting equation and use that spherical harmonics are orthonormal functions. The appearing expressions can be interpreted as elements of matrix representations of operators, for example, of the angular momentum operator known from quantum mechanics. We explicitly compute the matrix elements of all revealed operators. We then undertake a foray into group theory to illustrate that the computed matrix representations are related to each other via rotations. The rotation operators in the space of spherical harmonics are well-known in the quantum theory of angular momentum and their matrix representations are the Wigner-D matrices. To numerically solve the system of PDEs, we need to know the eigenvalues of some of the representation matrices. We prove that the eigenvalues of the relevant ones are the roots of the associated Legendre polynomials, where we use the fact that they are related via rotations which implies that they all have the same eigenvalues. We also address an issue that stems from using complex spherical harmonics to represent the distribution function which itself is real. As in the case of a Fourier series, the fact that the represented function is real implies that there exist simple relations between the expansion coefficients. This in turn implies that the system of PDEs determining the expansion coefficients contains redundant equations. They can be removed by the changing the basis of the space of spherical harmonics. We, therefore, multiply the system of PDEs with a basis transformation matrix, that turns the complex spherical harmonics into the real spherical harmonics.

In Chapter 6, the last chapter of this thesis, we numerically solve the system of PDEs. The way we have written it down in the previous chapter allows us to identify it as a system of advection-reaction equations. This suggests to use a numerical method that is able to represent the 'flow' of the variables, namely the finite volume (FV) or the discontinuous Galerkin (dG) method. The dG method is a finite element method (FEM). It has the advantage that the spatial accuracy of the numerical solution can be increased without the necessity to gather information from neighbouring cells. As a finite element method the analysis of its properties, e.g. if it is well-posed, relies on the structure of the Hilbert space in which the solution is expected to be found. We opt for the dG method and we begin this chapter with thoroughly explaining how to apply it to the system of PDEs, because it is the first time that the dG method is used in conjunction with a spherical harmonic expansion of the distribution function. An essential freedom of the dG method is the choice of the numerical flux. We choose an upwind flux, because it tightens the stability of the method. Computing an upwind flux normally means to solve a computationally expansive Riemann problem at each cell interface. However, the formulation of the system of PDEs

in terms of the representation matrices allows us to get the solution to the Riemann problem in the spatial directions for free. In the p-direction we do solve the corresponding eigenproblem. In the second part of this chapter we test our implementation[4] of the dG method in three different examples. The first example is a toy model with known analytical solution that we use to test the temporal and spatial accuracy of the numerical solution. We show that the error decreases as theoretically expected if we decrease the time step or increase the spatial resolution. The second example investigates what the effects of truncating the spherical harmonic expansion are. In the third example, we simulate the acceleration of particles at a parallel shock and illustrate that the implementation is able to accurately reproduce the analytical solution that we derived in Chapter 3.

[4] It's publicly available on https://sapphirepp.org

CHAPTER 2

PLASMA MODELS AND THEIR PHYSICAL ASSUMPTIONS

A plasma is a collection of charged particles whose density is high enough to ensure that its evolution is not predominantly due to individual encounters but due to the long-range Coulomb forces that all particles collectively and simultaneously exert. This entails that the plasmas' behaviour is collective (Krall & Trivelpiece, 1973, Ch. 1). In this chapter we present to the reader two different mathematical descriptions of plasmas and their underlying assumptions. First, a kinetic description embellished in the Boltzmann equation and based on non-equilibrium statistical mechanics, and secondly, a fluid description that is derived from the kinetic description and known as magnetohydrodynamics (MHD). We use the kinetic description to compute the propagation and the acceleration of charged particles, i.e. their transport, in a background plasma, which in turn is modelled using the MHD equations. Though, the charged particles, e.g. cosmic rays, are not dense enough to qualify as a plasma in the sense of the given definition. Nonetheless, their density is high enough for the statistical methods, which we introduce in the first part of this chapter, to apply.

2.1 KINETIC DESCRIPTION

We stress that a statistical treatment of plasmas is not merely a simplification, but a necessity stemming from the fact that it is impractical to know all the initial conditions of the ordinary differential equations, which would determine the particles' trajectories thus the evolution of the plasma. Since it is infeasible to know the velocities and positions of all particles at some time t_0, we rely on a theoretical framework which is capable to handle this lack of information, namely a set of tools collected under the label 'classical statistical mechanics'. The specification 'classical' highlights the first assumption that we make through out this work, i.e. we assume that the plasma density is low enough to neglect all quantum effects.[1] At the heart of statistical mechanics are probability density functions, which allow us to compute the probabilities that the plasma particles

[1] Only if the plasmas' density was so high that the average distance between particles, $\Delta x \approx n^{-1/3}$, was of the order of the de Broglie wavelength, quantum effects would become important.

are at a specific point in physical space and move in a given direction at a known speed.

We explicitly mention the physical intuition guiding the usage of probabilities: We do not know everything about the particles at hand, but we may know something, for example, the temperature of a set of particles in thermal equilibrium, which allows us to deduce, at least in principal, a set of *possible* particle positions and velocities in accord with the given temperature. Some particle configurations will be more likely than others. The lack of information is compensated with the allowance of many possibilities in agreement with the little we know. Another example for something we may know but which is not tight to some sort of equilibrium is that plasma particles do not tend to form clusters, i.e. their positions are uncorrelated.

A way to derive the Boltzmann equation is to use that the number of particles has to be conserved and Hamilton's equation of motion, see Thorne and Blandford (2017, Box 3.2). We take a different route and adopt Klimontovich's approach. This allows us to precisely define the concept of the one (or single) particle distribution function; the central quantity of this thesis.

2.1.1 MICROSCOPIC PARTICLE NUMBER DENSITY

A plasma may consist of N_c different components. For example, a fully ionised plasma of protons and electrons has two components. Furthermore, we denote with N_s the number of particles of component s. For each component, we define the *microscopic number density* of particles in phase space as

$$f_{s,\mu}(t, \boldsymbol{x}, \boldsymbol{p}) \coloneqq \sum_{i=1}^{N_s} \delta\left(\boldsymbol{x} - \boldsymbol{x}_{s,i}(t)\right) \delta\left(\boldsymbol{p} - \boldsymbol{p}_{s,i}(t)\right) , \tag{2.1}$$

where $\boldsymbol{x}_{s,i}$ is the position of the ith particle of component s in physical (or configuration) space and $\boldsymbol{p}_{s,i}$ its position in momentum space (Schram, 1991, eq. 3.1.1). We note that this definition is symbolic in character, because of the delta distributions. Though, the microscopic particle number density is well defined inside integral expressions, for example,

$$N_s = \int_V f_{s,\mu} \, \mathrm{d}^3x \, \mathrm{d}^3p \, ,$$

where N_s is the total number of particles of component s contained in the phase-space volume V.

We emphasise that the microscopic particle number density incorporates the trajectories of all particles of component s. However, as pointed out in the introduction to this section it is impossible to know all the initial conditions determining these trajectories. We instead presuppose the existence of a *probability density function*

$$D(t, \Xi_1, \dots, \Xi_{N_c}) =: D(t, \Xi) \, . \tag{2.2}$$

The components of the vector $\boldsymbol{\Xi}_s$ are the phase-space coordinates of the N_s particles of the plasma component s, namely $\boldsymbol{\Xi}_s = (\boldsymbol{\xi}_{s,1}, \dots, \boldsymbol{\xi}_{s,N_s})$ where $\boldsymbol{\xi}_{s,i} = (\boldsymbol{x}_{s,i}, \boldsymbol{p}_{s,i})$ are the coordinates of the ith particle of component s. We also defined the vector $\boldsymbol{\Xi} = (\boldsymbol{\Xi}_1, \dots, \boldsymbol{\Xi}_{N_c})$ that stacks together the coordinates of all plasma components. Moreover, we introduce the infinitesimal volumes

$$\begin{aligned} \mathrm{d}\Xi &\coloneqq \mathrm{d}\Xi_1 \cdots \mathrm{d}\Xi_{N_c}\,, \\ \mathrm{d}\Xi_s &\coloneqq \mathrm{d}\xi_{s,1} \dots \mathrm{d}\xi_{s,N_s} \quad \text{and} \\ \mathrm{d}\xi_{s,i} &\coloneqq \mathrm{d}^3x_{s,i}\,\mathrm{d}^3p_{s,i}\,. \end{aligned}$$

With these definitions at hand, we can say that $D(t, \boldsymbol{\Xi})\,\mathrm{d}\Xi$ is the probability to find particle 1 of component 1 at time t in the infinitesimal volume $\mathrm{d}\xi_{1,1}$ around $\boldsymbol{\xi}_{1,1}$, particle 2 of component 1 in $\mathrm{d}\xi_{1,2}$ around $\boldsymbol{\xi}_{1,2}$, etc. (Schram, 1991, p. 10). More generally, $D(t, \boldsymbol{\Xi})\,\mathrm{d}\Xi$ is the probability for the plasma particles to be at time t located in the volume $\mathrm{d}\Xi$ around $\boldsymbol{\Xi}$. We note that the probability density function (2.2) implicitly assumes that the positions in phase space of the particles of different plasma components are statistically dependent, in particular, $D(t, \boldsymbol{\Xi}) \neq \prod_s D_s(t, \boldsymbol{\Xi}_s)$.

An important property of the probability density function D is that its value does not change if two particles of the same component exchange their phase-space coordinates, i.e.

$$\begin{aligned} &D(t, \boldsymbol{\Xi}_1, \dots, \boldsymbol{\xi}_{s,1}, \dots, \boldsymbol{\xi}_{s,i}, \dots, \boldsymbol{\xi}_{s,j}, \dots, \boldsymbol{\xi}_{s,N_s}, \dots, \boldsymbol{\Xi}_{N_c}) \\ &\quad = D(t, \boldsymbol{\Xi}_1, \dots, \boldsymbol{\xi}_{s,1}, \dots, \boldsymbol{\xi}_{s,j}, \dots, \boldsymbol{\xi}_{s,i}, \dots, \boldsymbol{\xi}_{s,N_s}, \dots, \boldsymbol{\Xi}_{N_c})\,. \end{aligned} \tag{2.3}$$

The reason is that the particles are identical and their labels arbitrary, i.e. nothing prevents us from changing the index of the ith particle to be j and vice versa. We refer to this property of D as symmetry and say that D is *symmetric* with respect to its arguments (Schram, 1991, eq. 1.4.10).

This property, for example, allows us to uniquely define *reduced distribution functions* of the plasma component s, namely

$$\begin{aligned} f_s^{(k)}(t, \boldsymbol{\xi}_{s,1}, \dots, \boldsymbol{\xi}_{s,k}) &\coloneqq \frac{N_s!}{(N_s-k)!} \int_V D(t, \boldsymbol{\Xi})\,\mathrm{d}\Xi_1 \cdots \mathrm{d}\xi_{s,k+1} \cdots \mathrm{d}\xi_{s,N_s} \cdots \mathrm{d}\Xi_{N_c} \\ &\approx N_s^k \int_V D(t, \boldsymbol{\Xi})\,\mathrm{d}\Xi_1 \cdots \mathrm{d}\xi_{s,k+1} \cdots \mathrm{d}\xi_{s,N_s} \cdots \mathrm{d}\Xi_{N_c}\,, \end{aligned} \tag{2.4}$$

where $k \ll N_s$ and $f_s^{(k)}\,\mathrm{d}\xi_{s,1} \cdots \mathrm{d}\xi_{s,k}$ is the probability to find an *arbitrary* particle of the plasma component s in the infinitesimal volume $\mathrm{d}\xi_{s,1}$ around $\boldsymbol{\xi}_{s,1}$ *and* an arbitrary particle in $\mathrm{d}\xi_{s,2}$ around $\boldsymbol{\xi}_{s,2}$ etc. Without the factor $N_s!/(N_s-k)!$, it would be the probability to find particle 1 of component s, i.e. a specific particle, in $\mathrm{d}\xi_{s,1}$ around $\boldsymbol{\xi}_{s,1}$ and particle 2 in $\mathrm{d}\xi_{s,2}$ around $\boldsymbol{\xi}_{s,2}$ etc. The factor counts the possibilities to put k of the N_s particles at the k positions $\boldsymbol{\xi}_{s,1}, \dots, \boldsymbol{\xi}_{s,k}$ (Thorne & Blandford, 2017, eq. 22.60). Moreover, we note that we only use one integral

sign in eq. (2.4), although we integrate with respect to all explicitly appearing variables. We just implicitly assume that the volume of integration is the same for all them.

We now show how we obtain *macroscopic quantities* with the help of the probability density function D. When we say macroscopic quantities, we mean the average of functions, depending on the individual particles' positions in phase space, computed with $D(t, \boldsymbol{\Xi})$. The average of the microscopic particle number density serves us as an example. In a first step we have to redefine $f_{s,\mu}$, because it is a function of $t, \boldsymbol{x}, \boldsymbol{p}$ and not a function of the individual positions of the particles. In the context of averaging, we define it to be a *random variable* that depends on all possible particle positions of the plasma component s, i.e.

$$\begin{aligned} f_{s,\mu}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{x}_{s,1}, \boldsymbol{p}_{s,1}, \dots, \boldsymbol{x}_{s,N_s}, \boldsymbol{p}_{s,N_s}) &= f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\xi}_{s,1}, \dots, \boldsymbol{\xi}_{s,N_s}) \\ &= f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s) \coloneqq \sum_{i=1}^{N_s} \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}) \,, \end{aligned} \tag{2.5}$$

where $\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}) = \delta(\boldsymbol{x} - \boldsymbol{x}_{s,i})\delta(\boldsymbol{p} - \boldsymbol{p}_{s,i})$. This allows us to straightforwardly compute the average particle number density, namely

$$\begin{aligned} f_s(t, \boldsymbol{\xi}) \coloneqq \mathrm{E}\left[f_{s,\mu}\right] &\coloneqq \sum_{i=1}^{N_s} \int_V \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}) D(t, \boldsymbol{\Xi}) \, \mathrm{d}\boldsymbol{\Xi} \\ &= N_s \int_V \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,1}) D(t, \boldsymbol{\Xi}) \, \mathrm{d}\boldsymbol{\Xi} \\ &= \int_V \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,1}) f_s^{(1)}(t, \boldsymbol{\xi}_{s,1}) \, \mathrm{d}\xi_{s,1} = f_s^{(1)}(t, \boldsymbol{\xi}) \,. \end{aligned} \tag{2.6}$$

Notice that the only time dependent quantity is the probability density function D. In the second equation we exploited that D is symmetric in the sense of eq. (2.3). We call $f_s = f_s^{(1)}$ the *one* or *single particle distribution function* (Schram, 1991, eq. 3.1.2) and we recall that as a reduced distribution function $f_s^{(1)} \, \mathrm{d}^3x \, \mathrm{d}^3v$ is the likelihood for an arbitrary particle of component s to be found in the infinitesimal volume $\mathrm{d}\xi$ around $\boldsymbol{\xi}$. However, the one particle distribution function is of course also interpreted as the average number density. The relation between these two interpretations becomes clear if we consider a *homogeneous* gas, i.e. a gas with a constant number density. Assume that the gas occupies the volume V, then the probability for a particle, say particle 1, to be in $\mathrm{d}V$ around $\boldsymbol{x}$ is $\mathrm{d}V/V$. This implies that the probability density function is $p(\boldsymbol{x}) = 1/V$. If the gas consists of N particles, the probability to find an *arbitrary* particle in this volume element is $N \, \mathrm{d}V/V = Np(\boldsymbol{x}) \, \mathrm{d}V$. $Np(\boldsymbol{x})$ is the example's analogue of the one particle distribution function and, clearly, it is the average number density of the gas. This example highlights another aspect of the one particle distribution function, namely its normalisation. Integrating $Np(\boldsymbol{x})$ over V yields the number of particles. The same is true for the integral of $f_s^{(1)}(t, \boldsymbol{\xi})$.

Before we proceed with investigating the single particle distribution function, we present to the reader an alternative (re-)definition of the microscopic particle density. Instead of the definition given in eq. (2.5), we define $f_{s,\mu}$ to be a random variable that depends on the initial conditions of all particles of component s and that evolves deterministically with t. This results in

$$\begin{aligned} &f_{s,\mu}(t, \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{x}_{s,1,0}, \boldsymbol{p}_{s,1,0}, \dots, \boldsymbol{x}_{s,N_s,0}, \boldsymbol{p}_{s,N_s,0}) \\ &\quad = f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\xi}_{s,1,0}, \dots, \boldsymbol{\xi}_{s,N_s,0}) = f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) \coloneqq \sum_{i=1}^{N_s} \delta\left(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}(t, \boldsymbol{\xi}_{s,i,0})\right), \end{aligned} \tag{2.7}$$

where $\boldsymbol{\xi}_{s,i}(t, \boldsymbol{\xi}_{s,i,0})$ is the phase-space trajectory of the ith particle of component s. It depends on the random initial conditions $\boldsymbol{\xi}_{s,i,0} = (\boldsymbol{x}_{s,i,0}, \boldsymbol{p}_{s,i,0})$. The corresponding probability density is $D(\boldsymbol{\Xi}_0) \coloneqq D(0, \boldsymbol{\Xi})$.

The two definitions of the microscopic particle densities are related to each other. The first definition of the microscopic particle density, as given in eq. (2.5), uses random particle positions, whereas the second definition in eq. (2.7) is based on random initial conditions. If the random particle positions at $t = 0$ coincide with the random initial conditions, as implied by the definition $D(\boldsymbol{\Xi}_0) \coloneqq D(0, \boldsymbol{\Xi})$, then $D(t, \boldsymbol{\Xi})\,\mathrm{d}\Xi = D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0$, i.e. the the probability for the particles to be in the infinitesimal volume $\mathrm{d}\Xi$ around $\boldsymbol{\Xi}$ at time t equals the probability for the particles to have initial conditions in $\mathrm{d}\Xi_0$ around $\boldsymbol{\Xi}_0$. The reason is that the phase-space positions of the particles are determined by their equation of motions, i.e. $\boldsymbol{\Xi} = \boldsymbol{\Xi}(t, \boldsymbol{\Xi}_0)$.

This allows us to exchange the two definitions of the microscopic particle density when computing its average or higher order (central) moments. For example,

$$\begin{aligned} f_s(t, \boldsymbol{\xi}) = \mathrm{E}\left[f_{s,\mu}\right] &= \sum_{i=1}^{N_s} \int_V \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}) D(t, \boldsymbol{\Xi})\,\mathrm{d}\Xi \\ &= \sum_{i=1}^{N_s} \int_V \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}(t, \boldsymbol{\xi}_{s,i,0})) D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,. \end{aligned}$$

In general, whenever we define a microscopic function, say $A_\mu(t, \boldsymbol{\xi})$, which implicitly depends on the phase-space coordinates of the particles, we interpret them as random variables when computing macroscopic quantities, i.e when computing their average values. Hence, $A_\mu(t, \boldsymbol{\xi})$ becomes $A_\mu(\boldsymbol{\xi}, \boldsymbol{\Xi})$ or $A_\mu(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_0)$.

We return to the one particle distribution function. It is *the* quantity of our work, because we use it to model the distribution of charged and energetic particles whose positions and energies we would like to compute. If we say 'to model', we mean that we are looking for a partial differential equation whose solution describes how f_s evolves in phase space. Because derivatives are only defined for 'smooth' functions, we need to make sure that f_s is smooth. The average particle density will be smooth, if we choose the length and velocity

scales of our model such that the average number of particles in a small cell, small enough to be thought of as infinitesimal with respect to the chosen scales, is not subject to large *statistical* fluctuations. This means that the standard deviation of the average number of particles is small in comparison to the average. We derive a quantitative criterion ensuring that the described condition is fulfilled. To this end we compute the standard deviation of the average particle number of the plasma component s, cf. Schram (1991, Section 3.1).

The exact number of particles in the phase-space volume Δ is

$$N_{s,\Delta}(t) = \int_{\Delta} f_{s,\mu}(t, \boldsymbol{\xi}) \, \mathrm{d}\xi \, .$$

In agreement with our redefinitions of the microscopic particle density in the context of averaging, the average number of particles in Δ is

$$\mathrm{E}\left[N_{s,\Delta}\right](t) = \int_{\Delta} \int_{V} f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s) D(t, \boldsymbol{\Xi}) \, \mathrm{d}\Xi \, \mathrm{d}\xi = \int_{\Delta} f_s^{(1)}(t, \boldsymbol{\xi}) \, \mathrm{d}\xi \, ,$$

see eq. (2.6). The variance of the average number density is

$$\begin{aligned}
\mathrm{Var}(N_{s,\Delta}) &\coloneqq \mathrm{E}\left[\left(N_{s,\Delta} - \mathrm{E}\left[N_{s,\Delta}\right]\right)^2\right] = \mathrm{E}\left[N_{s,\Delta}^2\right] - \mathrm{E}\left[N_{s,\Delta}\right]^2 \\
&= \int_{\Delta} \int_{\Delta} \mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s) f_{s,\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_s)\right] \mathrm{d}\xi \, \mathrm{d}\xi' \\
&\quad - \int_{\Delta} \int_{\Delta} \mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s)\right] \mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_s)\right] \mathrm{d}\xi \, \mathrm{d}\xi' \\
&=: \int_{\Delta} \int_{\Delta} \mathrm{Cov}\left(f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s), f_{s,\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_s)\right) \mathrm{d}\xi \, \mathrm{d}\xi' \, .
\end{aligned}$$

The covariance $\mathrm{Cov}(f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s), f_{s,\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_s))$ quantifies how strongly deviations of the microscopic particle number density from the average particle number density at $\boldsymbol{\xi}$ are correlated with simultaneous deviations of the microscopic particle number density from its average at $\boldsymbol{\xi}'$. Note that we made the time dependence implicit in the above equation. And we will leave it implicit for the following equation to increase its readability.

We compute the covariance in two steps: We, firstly, evaluate

$$\begin{aligned}
&\mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi},\boldsymbol{\Xi}_s)f_{s,\mu}(\boldsymbol{\xi}',\boldsymbol{\Xi}_s)\right] \\
&\quad= \sum_{i=1}^{N}\sum_{j=1}^{N}\int_V \delta(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i})\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}_{s,j})D(\boldsymbol{\Xi})\,\mathrm{d}\boldsymbol{\Xi} \\
&\quad= \sum_{i=1}^{N_s}\int_V \delta(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i})\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}_{s,i})D(\boldsymbol{\Xi})\,\mathrm{d}\boldsymbol{\Xi} \\
&\qquad+ \sum_{i=1}^{N_s}\sum_{\substack{j=1\\ i\neq j}}^{N_s}\int_V \delta(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i})\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}_{s,j})D(\boldsymbol{\Xi})\,\mathrm{d}\boldsymbol{\Xi} \\
&\quad= N_s\int_V \delta(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,1})\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}_{s,1})D(\boldsymbol{\Xi})\,\mathrm{d}\boldsymbol{\Xi} \\
&\qquad+ N_s(N_s-1)\int_V \delta(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,1})\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}_{s,2})D(\boldsymbol{\Xi})\,\mathrm{d}\boldsymbol{\Xi} \\
&\quad= \delta(\boldsymbol{\xi}-\boldsymbol{\xi}')f_s^{(1)}(\boldsymbol{\xi}) + f_s^{(2)}(\boldsymbol{\xi},\boldsymbol{\xi}')\,,
\end{aligned} \tag{2.8}$$

where we in the last equation used that

$$\int_V \delta(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,1})\delta(\boldsymbol{\xi}'-\boldsymbol{\xi}_{s,1})f_s^{(1)}(\boldsymbol{\xi}_{s,1})\,\mathrm{d}\boldsymbol{\xi}_{s,1} = \delta(\boldsymbol{\xi}-\boldsymbol{\xi}')f_s^{(1)}(\boldsymbol{\xi})\,.$$

We note that $f_s^{(2)}(t,\boldsymbol{\xi},\boldsymbol{\xi}')$ is a reduced distribution function, see the definition in eq. (2.4), and, accordingly, $f_s^{(2)}\,\mathrm{d}\boldsymbol{\xi}\,\mathrm{d}\boldsymbol{\xi}'$ is the probability that an arbitrary particle of the plasma component s is in the infinitesimal volume $\mathrm{d}\boldsymbol{\xi}$ around $\boldsymbol{\xi}$ *and* another particle, no matter which, in $\mathrm{d}\boldsymbol{\xi}'$ around $\boldsymbol{\xi}'$. It is called the *pair distribution function*. It is common to write the pair distribution function as

$$f_s^{(2)}(t,\boldsymbol{\xi},\boldsymbol{\xi}') = f_s^{(1)}(t,\boldsymbol{\xi})f_s^{(1)}(t,\boldsymbol{\xi}') + g_{ss}(t,\boldsymbol{\xi},\boldsymbol{\xi}')\,. \tag{2.9}$$

The first term is proportional to the probability of finding two arbitrary particles at the respective positions assuming that the particles' positions are statistically independent, also called uncorrelated. The function $g_{ss}(t,\boldsymbol{\xi},\boldsymbol{\xi}')$ may be called the *two-point correlation function* and represents the effects on the 'uncorrelated' probability stemming from interactions between the particles. These interactions may drive the particles apart or cluster them, which leads to correlations in the particles positions (Schram, 1991; Thorne & Blandford, 2017, eq. 3.15 and eq. 22.65 respectively).

In the second step we plug the above result into our definition of the covariance. This yields

$$\begin{aligned}
&\mathrm{Cov}\left(f_{s,\mu}(\boldsymbol{\xi},\boldsymbol{\Xi}_s), f_{s,\mu}(\boldsymbol{\xi}',\boldsymbol{\Xi}_s)\right) \\
&\quad= \mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi},\boldsymbol{\Xi}_s)f_{s,\mu}(\boldsymbol{\xi}',\boldsymbol{\Xi}_s)\right] - \mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi},\boldsymbol{\Xi}_s)\right]\mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi}',\boldsymbol{\Xi}_s)\right] \\
&\quad= \delta(\boldsymbol{\xi}-\boldsymbol{\xi}')f_s^{(1)}(\boldsymbol{\xi}) + f_s^{(2)}(\boldsymbol{\xi},\boldsymbol{\xi}') - f_s^{(1)}(\boldsymbol{\xi})f_s^{(2)}(\boldsymbol{\xi}') \\
&\quad= \delta(\boldsymbol{\xi}-\boldsymbol{\xi}')f_s^{(1)}(\boldsymbol{\xi}) + g_{ss}(t,\boldsymbol{\xi},\boldsymbol{\xi}')\,.
\end{aligned} \tag{2.10}$$

If we assume that the positions of the particles of component s are uncorrelated, i.e. $g_{ss} = 0$, the variance of the average particle number in the phase-space volume Δ is

$$\mathrm{Var}(N_{s,\Delta}) = \int_\Delta \int_\Delta \delta(\boldsymbol{\xi} - \boldsymbol{\xi}') f_s^{(1)}(\boldsymbol{\xi}) \,\mathrm{d}\xi \,\mathrm{d}\xi' = \mathrm{E}\left[N_{s,\Delta}\right] .$$

This implies for the standard deviation $\sigma_\Delta := \mathrm{Var}(N_{s,\Delta})^{1/2}$ that

$$\frac{\sigma_\Delta}{\mathrm{E}\left[N_{s,\Delta}\right]} = \mathrm{E}\left[N_{s,\Delta}\right]^{-1/2} ,$$

i.e. the relative deviation from the average decreases with an increasing number of particles in the volume Δ.

We conclude that if we would like to work with a 'smooth' single particle distribution function, we have to assume length and velocity scales large enough such that an infinitesimal volume (with respect to these scales) includes sufficient particles. Sufficient means that the statistical variations of the values of f_s will be below a required accuracy.

2.1.2 KLIMONTOVICH EQUATION

The idea behind the Klimontovich equation is to start with the microscopic particle number density instead of the equations of motion of the individual particles. Klimontovich (1967, p. 48) writes that this is 'convenient' and he may mean that this approach allows one to obtain equations for macroscopic quantities, like the single particle distribution function, via averaging.

The formulation of the Klimontovich equation requires us to introduce the *microscopic particle flux* in phase space

$$\boldsymbol{J}_{s,\mu}(t, \boldsymbol{\xi}) := \sum_{i=0}^{N} \dot{\boldsymbol{\xi}}_{s,i} \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}(t)) , \tag{2.11}$$

where $\dot{\boldsymbol{\xi}}_{s,i} = (\boldsymbol{v}_{s,i}, \dot{\boldsymbol{p}}_{s,i})$.

We assume that no particles are created nor destroyed, i.e. the number of particles is conserved. This statement is equivalent to[2]

$$\int_\Delta f_{s,\mu}(t_2, \boldsymbol{\xi}) \,\mathrm{d}\xi - \int_\Delta f_{s,\mu}(t_1, \boldsymbol{\xi}) \,\mathrm{d}\xi + \int_{t_1}^{t_2} \int_{\partial\Delta} \boldsymbol{J}_{s,\mu}(t, \boldsymbol{\xi}) \cdot \mathrm{d}\boldsymbol{S} \,\mathrm{d}t = 0 . \tag{2.12}$$

[2]One may wonder why we integrated over t. The reason is that the number of particles does not change continuously and, thus,

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Delta f_{\mu,a}(t, \boldsymbol{\xi}) \,\mathrm{d}\xi$$

is not defined.

This is the *Klimontovich equation* in integral form. Δ is an arbitrary phase-space volume with surface $\partial\Delta$ and $\mathrm{d}\boldsymbol{S}$ is a corresponding outward pointing normal.

If the microscopic particle flux was known, the Klimontovich equation would determine the microscopic particle density. The microscopic particle flux could, in principle, be computed if all initial conditions were known, because the force acting on, say, the ith plasma particle of component s is the Lorentz force, namely

$$\dot{\boldsymbol{p}}_{s,i} = q\left(\boldsymbol{E}_\mu(t, \boldsymbol{x}_{s,i}) + \frac{\boldsymbol{p}_{s,i}}{m} \times \boldsymbol{B}_\mu(t, \boldsymbol{x}_{s,i})\right), \tag{2.13}$$

and we could integrate the particles' trajectories. Though, the source of the Lorentz force is the motion of the plasma particles, because they are charged. We introduced the symbols $\boldsymbol{E}_\mu$ and $\boldsymbol{B}_\mu$ to denote these *microscopic electromagnetic fields*. Notice that there is no subscript s; all plasma components contribute the electromagnetic fields. This implies that we also would have to solve Maxwell's equations to determine the microscopic particle flux, i.e.

$$\int_{\partial\Delta_x} \boldsymbol{E}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{S} = \frac{1}{\epsilon_0} \int_{\Delta_x} \rho_\mu(t, \boldsymbol{x})\, \mathrm{d}^3x \tag{2.14}$$

$$\int_{\partial\Delta_x} \boldsymbol{B}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{l} = \mu_0 \int_{\partial\Delta_x} \boldsymbol{j}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{S} + \frac{1}{c^2}\frac{\mathrm{d}}{\mathrm{d}t} \int_{\partial\Delta_x} \boldsymbol{E}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{S} \tag{2.15}$$

$$\int_{\partial\Delta_x} \boldsymbol{E}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{l} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_{\partial\Delta_x} \boldsymbol{B}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{S} \tag{2.16}$$

$$\int_{\partial\Delta_x} \boldsymbol{B}_\mu(t, \boldsymbol{x}) \cdot \mathrm{d}\boldsymbol{S} = 0\,, \tag{2.17}$$

where Δ_x is the configuration space volume of $\Delta = \Delta_x\Delta_p$. We use the integral equations, because the fields, when evaluated at the position of a specific particle, diverge and the positions of the particles are a null set and hence irrelevant for the value of the integral. The fact that the microscopic charge densities and current densities must be defined as

$$\rho_\mu(t, \boldsymbol{x}) \coloneqq \sum_{s=1}^{N_c} q_s \sum_{i=1}^{N_s} \delta(\boldsymbol{x} - \boldsymbol{x}_{s,i}(t)) = \sum_{s=1}^{N_c} q_s \int f_{s,\mu}(t, \boldsymbol{\xi})\, \mathrm{d}^3p \tag{2.18}$$

$$\boldsymbol{j}_\mu(t, \boldsymbol{x}) \coloneqq \sum_{s=1}^{N_c} q_s \sum_{i=1}^{N_s} \boldsymbol{v}_{s,i}\delta(\boldsymbol{x} - \boldsymbol{x}_{s,i}(t)) = \sum_{s=1}^{N_c} q_s \int \frac{\boldsymbol{p}_{s,i}}{m_s} f_{s,\mu}(t, \boldsymbol{\xi})\, \mathrm{d}^3p\,, \tag{2.19}$$

means that the Klimontovich equation (2.12) together with Maxwell's equations (2.14)–(2.17) and the definitions of the microscopic charge and current density form a closed set of equations that fully determines the state of the plasma. Essentially Klimontovich reformulated the original problem of computing all particle trajectories in terms of the microscopic particle densities. An actual solution to the above set of equations would still require to compute all trajectories and, thus, is infeasible. However, Klimontovich's idea is to use the

Klimontovich equation as a starting point to derive an analogue set of equations for the average microscopic particle number density, namely the single particle distribution function.

To average the Klimontovich equation (2.12), we proceed as in the previous section, i.e. we redefine the microscopic particle density and the microscopic particle flux to be random variables of the initial positions of the particles, see eq. (2.7). We subsequently multiply the Klimontovich equation with the probability density function $D(\boldsymbol{\Xi}_0)$ and integrate over a phase-space volume that includes all possible particle positions. This yields

$$\int_V \int_\Delta f_{s,\mu}(t_2, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\, \mathrm{d}\Xi_0\, \mathrm{d}\xi - \int_V \int_\Delta f_{s,\mu}(t_1, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\, \mathrm{d}\Xi_0\, \mathrm{d}\xi + \int_V \int_{t_1}^{t_2} \int_{\partial\Delta} \boldsymbol{J}_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_0) D(\boldsymbol{\Xi}_0)\, \mathrm{d}\Xi_0 \cdot \mathrm{d}\boldsymbol{S}\, \mathrm{d}t = 0\,.$$

We now divide by Δt and take the limit $\Delta t \to 0$ to get

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Delta \int_V f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\, \mathrm{d}\Xi_0\, \mathrm{d}\xi + \int_{\partial\Delta} \int_V \boldsymbol{J}_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_0) D(\boldsymbol{\Xi}_0)\, \mathrm{d}\Xi_0 \cdot \mathrm{d}\boldsymbol{S} = 0\,. \tag{2.20}$$

Notice that the t dependence appears in the arguments of the delta distributions and it might not be obvious that the integrals yield differentiable functions. Yet, that this is the case becomes clear if we use that $D(\boldsymbol{\Xi}_0)\, \mathrm{d}\Xi_0 = D(t, \boldsymbol{\Xi})\, \mathrm{d}\Xi$ together with the first redefinition of microscopic particle density, see eq. (2.5). The function $D(t, \boldsymbol{\Xi})$ is smooth in t and, hence, the integrals are as well.

A way to interpret equation (2.20) is that particle number conservation holds for all possible particle trajectories, i.e. eq. (2.12) holds; weighting the trajectories with the respective likelihood and adding them up leads to the conservation of the average particle number density.

2.1.3 THE BOLTZMANN AND THE VLASOV EQUATION

In this section we evaluate the integrals in eq. (2.20). We emphasise that the integrated equation liberates us from the necessity to know all the initial conditions determining the exact state of the plasma at all times. However, we need a way to determine the unknown probability density function D. This way leads to a chain of equations that gradually builds up the information contained in D and that if not interrupted matches the complexity of determining the trajectories of all particles (Krall & Trivelpiece, 1973, Sec. 7.4). It is possible to stop that chain with the formal introduction of a collision term. This yields the Boltzmann equation. Alternatively, a consideration of the parameters characterising a specific plasma, e.g. the interstellar medium, makes it also possible to stop at an

early link of the chain. The Vlasov equation, as we will see in the following paragraphs, results from the earliest possible interruption.

We now integrate the time derivative term in eq. (2.20). This yields

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Delta}\int_{V} f_{s,\mu}(t,\boldsymbol{\xi},\boldsymbol{\Xi}_{s,0})D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}\xi \\
&\quad= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Delta} N_s \int_{V} \delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,1}\right) D(t,\boldsymbol{\Xi})\,\mathrm{d}\Xi\,\mathrm{d}\xi \\
&\quad= \frac{\mathrm{d}}{\mathrm{d}t}\int_{\Delta} f_s(t,\boldsymbol{\xi})\,\mathrm{d}\xi\,,
\end{aligned} \tag{2.21}$$

where we used the symmetry of the probability density function as introduced in eq. (2.3) and the definition of the single particle distribution function, see eq. (2.6).

We proceed with the surface integral

$$\begin{aligned}
&\int_{\partial\Delta}\int_{V} \boldsymbol{J}_{s,\mu}(t,\boldsymbol{\xi},\boldsymbol{\Xi}_0)D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\cdot\mathrm{d}\boldsymbol{S} \\
&\quad= \int_{\partial\Delta}\int_{V}\sum_{i=1}^{N_s}\dot{\boldsymbol{\xi}}_{s,i}(t,\boldsymbol{\Xi}_0)\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}(t,\boldsymbol{\xi}_{s,i,0})\right)D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\cdot\mathrm{d}\boldsymbol{S} \\
&\quad= \int_{\partial\Delta}\int_{V}\sum_{i=1}^{N_s}\begin{pmatrix}\boldsymbol{v}_{s,i}\\ \dot{\boldsymbol{p}}_{s,i}\end{pmatrix}\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}(t,\boldsymbol{\xi}_{s,i,0})\right)D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\cdot\begin{pmatrix}\mathrm{d}\boldsymbol{S}_x\\ \mathrm{d}\boldsymbol{S}_p\end{pmatrix},
\end{aligned} \tag{2.22}$$

where $\mathrm{d}\boldsymbol{S}_x$ is an outward normal of the surface $\partial\Delta$ that belongs to configuration space and $\mathrm{d}\boldsymbol{S}_p$ is its momentum-space counterpart. Note that we suppressed the arguments of $\dot{\boldsymbol{\xi}}_{s,i}$ in the last line to increase the readability. We accordingly split the evaluation of the integral and begin with the part belonging to the configuration space, i.e.

$$\begin{aligned}
&\int_{\partial\Delta}\int_{V}\sum_{i=1}^{N_s}\boldsymbol{v}_{s,i}(t,\boldsymbol{\xi}_{s,i,0})\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}(t,\boldsymbol{\xi}_{s,i,0})\right)D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\cdot\mathrm{d}\boldsymbol{S}_x \\
&\quad= \int_{\partial\Delta} N_s\int_{V}\boldsymbol{v}_{s,1}\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,1}\right)D(t,\boldsymbol{\Xi})\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_x \\
&\quad= \int_{\partial\Delta}\frac{\boldsymbol{p}}{m}f_s(t,\boldsymbol{\xi})\cdot\mathrm{d}\boldsymbol{S}_x\,.
\end{aligned} \tag{2.23}$$

We note that the velocity of the ith particle of component s only depends on the initial conditions $\boldsymbol{\xi}_{s,i,0}$ and *not* on the initial conditions of the other particles. Moreover, $\boldsymbol{v}_{s,1} = \boldsymbol{p}_{s,1}/m$.

The momentum space part of the integral is

$$\begin{aligned}
&\int_{\partial\Delta}\int_V \sum_{i=1}^{N_s} \dot{\boldsymbol{p}}_{s,i}(t,\Xi_0)\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}(t,\boldsymbol{\xi}_{s,i,0})\right) D(\Xi_0)\,\mathrm{d}\Xi_0\cdot\mathrm{d}\boldsymbol{S}_p \\
&\quad=\int_{\partial\Delta}\sum_{i=1}^{N_s}\int_V q_s\boldsymbol{E}_\mu(\boldsymbol{x}_{s,i},\Xi)\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}\right) D(t,\Xi)\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_p \qquad (2.24)\\
&\qquad+\int_{\partial\Delta}\sum_{i=1}^{N_s}\int_V q_s\frac{\boldsymbol{p}_{s,i}}{m}\times\boldsymbol{B}_\mu(\boldsymbol{x}_{s,i},\Xi)\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}\right) D(t,\Xi)\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_p\,,
\end{aligned}$$

where we used the microscopic Lorentz force, as presented in eq. (2.13), and again exploited that $D(\Xi_0)\,\mathrm{d}\Xi_0 = D(t,\Xi)\,\mathrm{d}\Xi$. Notice that the microscopic Lorentz force acting on the particle i of the plasma component s, i.e. $\dot{\boldsymbol{p}}_{s,i}$, depends on the positions of all particles and not only the positions of the particle itself. For the electric field term we get

$$\begin{aligned}
&\int_{\partial\Delta}\sum_{i=1}^{N_s}\int_V q_s\boldsymbol{E}_\mu(\boldsymbol{x}_{s,i},\Xi_0)\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,i}\right) D(t,\Xi)\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_p \\
&\quad=\int_{\partial\Delta} N_s\int_V q_s\boldsymbol{E}_\mu(\boldsymbol{x}_{s,1},\Xi)\delta\left(\boldsymbol{\xi}-\boldsymbol{\xi}_{s,1}\right) D(t,\Xi)\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_p \qquad (2.25)\\
&\quad=\int_{\partial\Delta}\int_V q_s\boldsymbol{E}_\mu(\boldsymbol{x},\Xi)f_{s,\mu}(\boldsymbol{\xi},\Xi_s)D(t,\Xi)\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_p\,,
\end{aligned}$$

where we took advantage of the fact that the electric field $\boldsymbol{E}_\mu$ must be symmetric, i.e. it is possible to exchange $\boldsymbol{\xi}_{s,i}$ and $\boldsymbol{\xi}_{s,j}$ without changing its value; again particles of the same plasma component are identical, see eq. (2.3). In the last line of equation (2.25), we 'rexpanded' the sum over i to recover the microscopic particle density and replaced $\boldsymbol{x}_{s,1}$ with $\boldsymbol{x}$. This was possible, because for an arbitrary function $g(y)$ it holds that

$$\int g(y)\delta(x-y)\,\mathrm{d}y = g(x) = \int g(x)\delta(x-y)\,\mathrm{d}y\,.$$

It is possible express the microscopic electric field and the microscopic particle number density in terms of their average values, namely

$$f_{s,\mu}(t,\boldsymbol{\xi},\Xi_{s,0}) = f_s(t,\boldsymbol{\xi}) + \delta f_{s,\mu}(t,\boldsymbol{\xi},\Xi_{s,0}) \qquad (2.26)$$

$$\boldsymbol{E}_\mu(t,\boldsymbol{x},\Xi_0) = \boldsymbol{E}(t,\boldsymbol{x}) + \delta\boldsymbol{E}_\mu(t,\boldsymbol{x},\Xi_0)\,, \qquad (2.27)$$

where we defined $\boldsymbol{E}(t,\boldsymbol{x}) \coloneqq \mathrm{E}\left[\boldsymbol{E}_\mu(t,\boldsymbol{x},\Xi_0)\right]$ to be the average (or macroscopic) electric field. We note that these equations also implicitly define the new random variables $\delta f_{s,\mu}$ and $\delta\boldsymbol{E}_{s,\mu}$ whose average must be zero, i.e. $\mathrm{E}\left[\delta\boldsymbol{E}_\mu(t,\boldsymbol{x},\Xi_0)\right] = \mathrm{E}\left[\delta f_{s,\mu}(t,\boldsymbol{\xi},\Xi_{s,0})\right] = 0$.

The new expressions for the microscopic density and the microscopic electric field turn the electric field term into

$$\begin{aligned}
&\int_{\partial\Delta}\int_{V} q_s \boldsymbol{E}_\mu(\boldsymbol{x},\Xi) f_{s,\mu}(\boldsymbol{\xi},\Xi_s) D(t,\Xi)\,\mathrm{d}\Xi\cdot\mathrm{d}\boldsymbol{S}_p \\
&\quad= \int_{\partial\Delta} q_s \left(\boldsymbol{E}(t,\boldsymbol{x}) f_s(t,\boldsymbol{\xi}) + \mathrm{E}\left[\delta\boldsymbol{E}_\mu(\boldsymbol{x},\Xi)\delta f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right]\right)\cdot\mathrm{d}\boldsymbol{S}_p \qquad (2.28)\\
&\quad= \int_{\partial\Delta} q_s \left(\boldsymbol{E}(t,\boldsymbol{x}) f_s(t,\boldsymbol{\xi}) + \mathrm{Cov}\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi) f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right)\right)\cdot\mathrm{d}\boldsymbol{S}_p\,,
\end{aligned}$$

where we used that the expectation value is linear, i.e. $\mathrm{E}[aX+Y] = a\mathrm{E}[X] + \mathrm{E}[Y]$ and that the expectation values of the random variables $\delta f_{s,\mu}$ and $\delta\boldsymbol{E}_\mu$ vanish. Furthermore, we note that

$$\begin{aligned}
\mathrm{E}\left[\delta\boldsymbol{E}_\mu(\boldsymbol{x},\Xi)\delta f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right] &= \mathrm{E}\left[(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi) - \boldsymbol{E}(t,\boldsymbol{x}))(f_{s,\mu}(\boldsymbol{\xi},\Xi_s) - f_s(t,\boldsymbol{\xi}))\right]\\
&= \mathrm{Cov}\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right)\,.
\end{aligned}$$

A completely analogue computation can be performed for the magnetic force term. It results in

$$\int_{\partial\Delta} q_s \frac{\boldsymbol{p}}{m}\times\left(\boldsymbol{B}(t,\boldsymbol{x}) f_s(t,\boldsymbol{\xi}) + \mathrm{Cov}\left(\boldsymbol{B}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right)\right)\cdot\mathrm{d}\boldsymbol{S}_p\,. \qquad (2.29)$$

We collect the results our computations, namely eqs. (2.21), (2.23), (2.28) and (2.29) to plug them into the averaged Klimontovich equation (2.20). This yields

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Delta} f_s(t,\boldsymbol{\xi})\,\mathrm{d}\xi + \int_{\partial\Delta}\frac{\boldsymbol{p}}{m} f_s(t,\boldsymbol{\xi})\cdot\mathrm{d}\boldsymbol{S}_x\\
&\quad+\int_{\partial\Delta} q_s\left(\boldsymbol{E}(t,\boldsymbol{x}) f_s(t,\boldsymbol{\xi}) + \frac{\boldsymbol{p}}{m}\times\boldsymbol{B}(t,\boldsymbol{x}) f_s(t,\boldsymbol{\xi})\right)\cdot\mathrm{d}\boldsymbol{S}_p\\
&\quad= -\int_{\partial\Delta} q_s\left(\mathrm{Cov}\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi) f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right) + \frac{\boldsymbol{p}}{m}\times\mathrm{Cov}\left(\boldsymbol{B}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right)\right)\cdot\mathrm{d}\boldsymbol{S}_p
\end{aligned}$$

Applying the divergence theorem and exploiting the fact that the phase-space volume Δ was arbitrary, which implies that the integrand has to vanish, leads to

$$\begin{aligned}
&\frac{\partial f_s(t,\boldsymbol{\xi})}{\partial t} + \boldsymbol{v}\cdot\nabla_x f_s(t,\boldsymbol{\xi}) + q_s\left(\boldsymbol{E}(t,\boldsymbol{x}) + \frac{\boldsymbol{p}}{m}\times\boldsymbol{B}(t,\boldsymbol{x})\right)\cdot\nabla_p f_s(t,\boldsymbol{\xi}) \qquad (2.30)\\
&\quad= -q_s\nabla_p\cdot\left(\mathrm{Cov}\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right) + \frac{\boldsymbol{p}}{m}\times\mathrm{Cov}\left(\boldsymbol{B}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right)\right)\,.
\end{aligned}$$

We note that $\nabla_p\cdot(\boldsymbol{p}/m\times\boldsymbol{B}) = 0$. We explore the meaning of the right-hand side of this equation in the next section. Yet, we already mention that the appearance of the covariance of the microscopic electromagnetic fields and particle number density means that it will depend on the interactions of the plasma particles,

because they determine if clusters of charges are formed or not. In this sense, we formally replace the right-hand side with a *collision term* to obtain the *Boltzmann equation*, namely

$$\frac{\partial f_s(t,\boldsymbol{\xi})}{\partial t}+\boldsymbol{v}\cdot\nabla_x f_s(t,\boldsymbol{\xi})+q_s\left(\boldsymbol{E}(t,\boldsymbol{x})+\frac{\boldsymbol{p}}{m}\times\boldsymbol{B}(t,\boldsymbol{x})\right)\cdot\nabla_p f_s(t,\boldsymbol{\xi})=\left(\frac{\delta f_s}{\delta t}\right)_c . \quad (2.31)$$

If we ignore correlations of particle positions due to mutual interactions of the plasma particles and, thus, set the right-hand side to zero, we get the *Vlasov equation*, i.e.

$$\frac{\partial f_s(t,\boldsymbol{\xi})}{\partial t}+\boldsymbol{v}\cdot\nabla_x f_s(t,\boldsymbol{\xi})+q_s\left(\boldsymbol{E}(t,\boldsymbol{x})+\frac{\boldsymbol{p}}{m}\times\boldsymbol{B}(t,\boldsymbol{x})\right)\cdot\nabla_p f_s(t,\boldsymbol{\xi})=0 . \quad (2.32)$$

Analogue to the microscopic electromagnetic fields, which are determined by the microscopic particle number densities of the plasma components, the source of the average $\boldsymbol{E}$- and $\boldsymbol{B}$-fields are the corresponding single particle distribution functions. Thus, if we like to get a close set of equations for the f_s, we need to compute the macroscopic electromagnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$. To this end we average Maxwell's equations (2.14)–(2.17) for the microscopic electromagnetic fields. We exemplify the necessary steps using the microscopic Gauss law, i.e.

$$\begin{aligned}\int_{\partial\Delta_x}\mathrm{E}\left[\boldsymbol{E}_\mu(t,\boldsymbol{x},\boldsymbol{\Xi}_0)\right]\mathrm{d}\boldsymbol{S}&=\int_{\partial\Delta_x}\boldsymbol{E}(t,\boldsymbol{x})\cdot\mathrm{d}\boldsymbol{S} \quad (2.33)\\ &=\frac{1}{\epsilon_0}\int_{\Delta_x}\mathrm{E}\left[\rho_\mu(t,\boldsymbol{x},\boldsymbol{\Xi}_0)\right]\mathrm{d}^3x=\frac{1}{\epsilon_0}\int_{\Delta_x}\rho(t,\boldsymbol{x})\,\mathrm{d}^3x ,\end{aligned}$$

where we used the definition of the macroscopic $\boldsymbol{E}$-field and defined the macroscopic charge density, namely

$$\begin{aligned}\rho(t,\boldsymbol{x})&\coloneqq\mathrm{E}\left[\rho_\mu(t,\boldsymbol{x},\boldsymbol{\Xi}_0)\right] \quad (2.34)\\ &=\sum_{s=1}^{N_c}q_s\int\!\!\int_V f_{s,\mu}(t,\boldsymbol{\xi},\boldsymbol{\Xi}_{s,0})D_0(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}^3p=\sum_{s=1}^{N_c}q_s\int f_s(t,\boldsymbol{\xi})\,\mathrm{d}^3p .\end{aligned}$$

Averaging the other three Maxwell equations works analogously, though, we need the average current density, i.e.

$$\begin{aligned}\boldsymbol{j}(t,\boldsymbol{x})&\coloneqq\mathrm{E}\left[\boldsymbol{j}_\mu(t,\boldsymbol{x},\boldsymbol{\Xi}_0)\right] \quad (2.35)\\ &=\sum_{s=1}^{N_c}q_s\int\!\!\int_V\frac{\boldsymbol{p}_{s,i}}{m_s}f_{s,\mu}(t,\boldsymbol{\xi},\boldsymbol{\Xi}_0)D_0(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}^3p=\sum_{s=1}^{N_c}q_s\int\frac{\boldsymbol{p}}{m_s}f_s(t,\boldsymbol{\xi})\,\mathrm{d}^3p .\end{aligned}$$

Because the single particle distribution functions are smooth, we can apply the

divergence theorem to get the differential Maxwell's equations, namely

$$\nabla \cdot \boldsymbol{E}(t,\boldsymbol{x}) = \frac{1}{\epsilon_0}\sum_{s=1}^{N_c} q_s \int f_s(t,\boldsymbol{\xi})\,\mathrm{d}^3 p \tag{2.36}$$

$$\nabla \times \boldsymbol{B}(t,\boldsymbol{x}) = \mu_0 \sum_{s=1}^{N_c} q_s \int \frac{\boldsymbol{p}}{m_s} f_s(t,\boldsymbol{\xi})\,\mathrm{d}^3 p + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \tag{2.37}$$

$$\nabla \times \boldsymbol{E}(t,\boldsymbol{x}) = -\frac{\partial \boldsymbol{B}(t,\boldsymbol{x})}{\partial t} \tag{2.38}$$

$$\nabla \cdot \boldsymbol{B}(t,\boldsymbol{x}) = 0\,. \tag{2.39}$$

2.1.4 BBGKY HIERARCHY

The objective of this section is to explore the physical meaning of the collision term in the Boltzmann equation. Therefore, we compute the covariance of the microscopic electromagnetic fields and the microscopic density of the plasma particles appearing on the right-hand side of eq. (2.30). If we make the assumption that the plasma particles are moving slow in comparison to the speed of light, we can ignore the contribution from the microscopic magnetic field $\boldsymbol{B}_\mu$. Our computations result in the hierarchy (or chain) of equations that we mentioned in the introduction to the previous section and whose earliest possible interruption yielded the Vlasov equation. We climb up the hierarchy by one level and show that the speculation about mutual interactions and their relations to correlations in the particles' position is the correct physical interpretation of the collision term. At the end, we discuss physical conditions that may justify a stop at a low level of the hierarchy.

We set about showing that we can ignore the microscopic magnetic field when computing the right-hand side of eq. (2.30), if we assume a plasma that is build of particles moving with velocities that are small in comparison to the speed of light. We show that if this is the case the Lorentz force reduces to the Coulomb force. In general, the electric field is

$$\boldsymbol{E}(t,\boldsymbol{x}) = -\nabla\phi(t,\boldsymbol{x}) - \frac{\mathrm{d}\boldsymbol{A}}{\mathrm{d}t}(t,\boldsymbol{x})\,,$$

where $\phi(t,\boldsymbol{x})$ is the scalar potential and $\boldsymbol{A}$ the vector potential, see Jackson (1998, eq. 6.9). The potentials of a moving charge are the Liénard–Wichert potentials and for the case of a particle that moves uniformly with velocity $\boldsymbol{v}_0$ they have the form

$$\phi(t,\boldsymbol{x}) = \frac{q}{4\pi\epsilon_0\,|\boldsymbol{x}-\boldsymbol{X}(t)|}\,\frac{1}{\sqrt{1-\frac{v_0^2}{c^2}\sin^2\alpha}}$$

$$\boldsymbol{A}(t,\boldsymbol{x}) = \frac{\boldsymbol{v}_0}{c^2}\phi(t,\boldsymbol{x})\,,$$

where $\boldsymbol{X}(t) = \boldsymbol{X}_0 + \boldsymbol{v}_0 t$ is the trajectory of the particle and α is the angle enclosed by the velocity and the distance between $\boldsymbol{x}$ and $\boldsymbol{X}(t)$, i.e $\cos\alpha = \boldsymbol{v}_0 \cdot (\boldsymbol{x} - \boldsymbol{X}(t))/v_0 \, |\boldsymbol{x} - \boldsymbol{X}(t)|$. (Nolting, 2013, eq. 4.505)

This implies that we can treat the scalar potential ϕ as a function of the distance between $\boldsymbol{x}$ and $\boldsymbol{X}(t)$. Abusing notation, we write $\phi(t, \boldsymbol{x}) = \phi(\boldsymbol{x} - \boldsymbol{X}(t))$. The time derivative of the scalar potential is

$$\frac{\mathrm{d}\phi(\boldsymbol{x} - \boldsymbol{X}(t))}{\mathrm{d}t} = -\nabla\phi((\boldsymbol{x} - \boldsymbol{X}(t))) \cdot \boldsymbol{v}_0 = -\nabla\phi(t, \boldsymbol{x}) \cdot \boldsymbol{v}_0 \,,$$

which implies that $\mathrm{d}\boldsymbol{A}/\,\mathrm{d}t = -\boldsymbol{v}_0/c^2 \nabla\phi \cdot \boldsymbol{v}_0$. A consequence of the latter implication is that the magnitude of the rotational part of the $\boldsymbol{E}$-field is smaller than the magnitude of $\nabla\phi$ by a factor of v_0^2/c^2, because

$$\frac{|\dot{\boldsymbol{A}}|}{|\nabla\phi|} = \frac{v_0}{c^2} \frac{|\nabla\phi \cdot \boldsymbol{v}_0|}{|\nabla\phi|} \leq \frac{v_0^2}{c^2} \,,$$

where we used the Cauchy–Schwarz inequality.

Moreover, the Taylor expansion of the scalar potential is

$$\phi(t, \boldsymbol{x}) = \frac{q}{4\pi\epsilon_0 \, |\boldsymbol{x} - \boldsymbol{X}(t)|} \left(1 + \frac{1}{2}\frac{v_0^2}{c^2} \sin^2\alpha + \frac{3}{8}\frac{v_0^4}{c^4} \sin^4\alpha + \mathcal{O}(v_0^6/c^6)\right) .$$

The previous two statements show that the potential of a charged particle in uniform motion is the Coulomb potential if second order 'corrections' in v_0/c are neglected. Thus, its $\boldsymbol{E}$-field is,

$$\boldsymbol{E}(t, \boldsymbol{x}) \approx -\frac{q}{4\pi\epsilon_0} \nabla \frac{1}{|\boldsymbol{x} - \boldsymbol{X}(t)|} = -\frac{q}{4\pi\epsilon_0} \int \nabla \frac{1}{|\boldsymbol{x} - \boldsymbol{x}'|} \delta(\boldsymbol{x}' - \boldsymbol{X}(t)) \, \mathrm{d}^3 x' \,. \quad (2.40)$$

The $\boldsymbol{B}$-field of a slowly moving charged particle is

$$\begin{aligned} \boldsymbol{B}_i(t, \boldsymbol{x}) = (\nabla \times \boldsymbol{A}(t, \boldsymbol{x}))_i &= 1/c^2 \epsilon_{ijk} v_0^k \frac{\partial\phi}{\partial x^j}(t, \boldsymbol{x}) \\ &\approx -\frac{q}{4\pi\epsilon_0 c^2} \epsilon_{ijk} v_0^k \frac{\partial}{\partial x^j} \frac{1}{|\boldsymbol{x} - \boldsymbol{X}(t)|} \\ &= \epsilon_{ijk} \frac{v_0^k}{c^2} E^j(t, \boldsymbol{x}) = \frac{1}{c^2} (\boldsymbol{v}_0 \times \boldsymbol{E}(t, \boldsymbol{x}))_i \,. \end{aligned}$$

An expression that we can also derive by starting in the rest frame of the moving charge, Lorentz boosting into the inertial frame of interest and keeping terms of the order v_0^2/c^2. The magnitude of the magnetic force due to this $\boldsymbol{B}$-field acting on another particle that moves with velocity $\boldsymbol{v}_1$, where $v_1 \approx v_0$, is

$$\frac{q}{c^2} \, |\boldsymbol{v}_1 \times (\boldsymbol{v}_0 \times \boldsymbol{E})| \leq \frac{q}{c^2} \, |\boldsymbol{v}_1| \, |\boldsymbol{v}_0 \times \boldsymbol{E}| \leq q \frac{v_1 v_0}{c^2} \, |\boldsymbol{E}| \in \mathcal{O}(v_0^2/c^2) \,.$$

Hence, relative to the Coulomb force, the magnetic force term is second order in v_0/c.

In summary, if we neglect terms that are elements of $\mathcal{O}(v_0^2/c^2)$ in the averaged Klimontovich equation (2.30), we get

$$\begin{aligned}\frac{\partial f_s(t,\boldsymbol{\xi})}{\partial t} + \boldsymbol{v}\cdot\nabla_x f_s(t,\boldsymbol{\xi}) + q_s\boldsymbol{E}(t,\boldsymbol{x})\cdot\nabla_p f_s(t,\boldsymbol{\xi}) \\ = -q_s\nabla_p\cdot\operatorname{Cov}\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right).\end{aligned} \tag{2.41}$$

Neglecting these terms is called the *electrostatic approximation* (Schram, 1991, p. 40) or the *Coulomb approximation* (Krall & Trivelpiece, 1973, Section 7.3). See in particular Klimontovich (1967, p. 51).

In this approximation it is possible to derive a more explicit expression for the collision term, i.e. we can compute the covariance of the microscopic electric field and the microscopic particle density. Using eq. (2.40) the microscopic Coulomb field becomes

$$\begin{aligned}\boldsymbol{E}_\mu(t,\boldsymbol{x}) &= \sum_{s'=1}^{N_c}\sum_{i=1}^{N_{s'}} -\frac{q_{s'}}{4\pi\epsilon_0}\int\nabla\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\delta\left(\boldsymbol{x}'-\boldsymbol{x}_{\boldsymbol{s}',\boldsymbol{i}}(t)\right)\mathrm{d}^3x' \\ &= \sum_{s'=1}^{N_c} -\frac{q_{s'}}{4\pi\epsilon_0}\int\nabla\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}f_{s',\mu}(t,\boldsymbol{\xi}')\,\mathrm{d}^3\xi'.\end{aligned} \tag{2.42}$$

The macroscopic(or average) $\boldsymbol{E}$-field then is

$$\begin{aligned}\boldsymbol{E}(t,\boldsymbol{x}) &= \sum_{s'=1}^{N_c} -\frac{q_{s'}}{4\pi\epsilon_0}\int\nabla\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\int_V f_{s',\mu}(\boldsymbol{\xi}',\Xi_s)D(t,\Xi)\,\mathrm{d}\Xi\,\mathrm{d}^3\xi' \\ &= \sum_{s'=1}^{N_c} -\frac{q_{s'}}{4\pi\epsilon_0}\int\nabla\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}f_{s'}(t,\boldsymbol{\xi}')\,\mathrm{d}^3\xi',\end{aligned} \tag{2.43}$$

where we used the definition of the single particle distribution function as presented in eq. (2.6). With eq. (2.42) and eq. (2.43) at hand, we evaluate the right-hand side of eq. (2.41), i.e.

$$\begin{aligned}&\operatorname{Cov}\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi), f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right) = \mathrm{E}\left[\left(\boldsymbol{E}_\mu(\boldsymbol{x},\Xi)-\boldsymbol{E}(t,\boldsymbol{x})\right)\delta f_{s,\mu}(\boldsymbol{\xi},\Xi)\right] \\ &= \int_V\left(\sum_{s'=1}^{N_c} -\frac{q_{s'}}{4\pi\epsilon_0}\int\nabla\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\left(f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})-f_{s'}(t,\boldsymbol{\xi}')\right)\delta f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\,\mathrm{d}^3\xi'\right)\mathrm{d}\Xi \\ &= \sum_{s'=1}^{N_c} -\frac{q_{s'}}{4\pi\epsilon_0}\int\nabla\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\mathrm{E}\left[\delta f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})\delta f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right]\mathrm{d}^3\xi'.\end{aligned} \tag{2.44}$$

Notice that

$$\begin{aligned}&\mathrm{E}\left[\delta f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})\delta f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right] \\ &\quad= \operatorname{Cov}\left(f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right) \\ &\quad= \mathrm{E}\left[f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right] - \mathrm{E}\left[f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})\right]\mathrm{E}\left[f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right],\end{aligned}$$

which shows, firstly, that the right-hand side of eq. (2.41) could be understood as a consequence of correlated (or co-varying) particle positions of the plasma

components and, secondly, that it depends on $\mathrm{E}\left[f_{s',\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_s) f_s(\boldsymbol{\xi}, \boldsymbol{\Xi}_s)\right]$. The latter expectation value in turn depends on the expectation value of the product of three microscopic particle densities and so forth. This is the coupled chain of equations that we mentioned and that determines the expectation values of the product of ever more microscopic particle densities. This chain is called the BBGKY hierarchy and is named after the physicists Nikolay Bogolyubov, Max Born, Herbert S. Green, John G. Kirkwood and Jacques Yvon. For a concise account of the BBGKY theory we refer the reader to Montgomery and Tidman (1964, Ch. 4).

We digress for a moment from our investigation of the physically meaning of the collision term, to derive the equation for $\mathrm{E}\left[f_{s',\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_s) f_s(\boldsymbol{\xi}, \boldsymbol{\Xi}_s)\right]$. We proceed completely analogue to the case of the single particle distribution function $f_s(t, \boldsymbol{\xi}) = \mathrm{E}\left[f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0})\right]$, i.e. we use that the number of particles must be conserved, see eq. (2.12). Of course, the number of particles of each plasma component is conserved separately. We can thus multiply the particle number conservation equation for component s with $\int_\Delta f_{s',\mu}(t_1, \boldsymbol{\xi}')\,\mathrm{d}\xi'$ and the one for the component s' with $\int_\Delta f_{s,\mu}(t_2, \boldsymbol{\xi})\,\mathrm{d}\xi$. We add the resulting equations up, average them, divide them by Δt and take the limit $\Delta t \to 0$. A careful handling of the involved integrals then yields

$$
\begin{aligned}
0 = \frac{\mathrm{d}}{\mathrm{d}t} &\int_\Delta \int_V f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}\xi'\,\mathrm{d}\xi \\
&+ \int_\Delta \frac{\boldsymbol{p}}{m} \cdot \nabla_x \int_V f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}\xi'\,\mathrm{d}\xi \\
&+ \int_\Delta \frac{\boldsymbol{p}'}{m} \cdot \nabla_{x'} \int_V f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}\xi'\,\mathrm{d}\xi \\
&+ \int_\Delta \nabla_p \cdot \int_V q_s \boldsymbol{E}_\mu(t, \boldsymbol{x}, \boldsymbol{\Xi}_0) f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}\xi'\,\mathrm{d}\xi \\
&+ \int_\Delta \nabla_{p'} \cdot \int_V q_{s'} \boldsymbol{E}_\mu(t, \boldsymbol{x}', \boldsymbol{\Xi}_0) f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0}) D(\boldsymbol{\Xi}_0)\,\mathrm{d}\Xi_0\,\mathrm{d}\xi'\,\mathrm{d}\xi\,.
\end{aligned}
$$

We plug into the above equation the expression for the microscopic electric field, see eq. (2.42), and convert it into a differential equation. This results in

$$
\begin{aligned}
&\left(\frac{\partial}{\partial t} + \frac{\boldsymbol{p}}{m} \cdot \nabla_x + \frac{\boldsymbol{p}'}{m} \cdot \nabla_{x'}\right) \mathrm{E}\left[f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi}, \boldsymbol{\Xi}_{s,0})\right] \\
&\qquad = \sum_{s''=1}^{N_c} \int \left(\frac{q_s q_{s''}}{4\pi\epsilon_0} \nabla_x \frac{1}{|\boldsymbol{x} - \boldsymbol{x}''|} \cdot \nabla_p + \frac{q_{s'} q_{s''}}{4\pi\epsilon_0} \nabla_{x'} \frac{1}{|\boldsymbol{x}' - \boldsymbol{x}''|} \cdot \nabla_{p'}\right) \\
&\qquad\qquad \times \mathrm{E}\left[f_{s''}(t, \boldsymbol{\xi}'', \boldsymbol{\Xi}_{s'',0}) f_{s',\mu}(t, \boldsymbol{\xi}', \boldsymbol{\Xi}_{s',0}) f_{s,\mu}(t, \boldsymbol{\xi})\right] \mathrm{d}\xi''\,.
\end{aligned}
\tag{2.45}
$$

Cf., for example, Krall and Trivelpiece (1973, eq. 7.3.5). This shows that the expectation value of the product of two microscopic particle densities depends on the expectation value of the product of three microscopic particle densities.

We stop here and do not show that the latter expectation value depends on the expectation value of the product of four microscopic particle densities. Though, we refer the reader to Schram (1991, eq. 3.5.10), who gives a formula, assuming the electrostatic approximation and a one component plasma, how the evolution of the expectation value of k microscopic particle number densities depends on the expectation value of $k+1$ densities.

Our overarching aim is to compute the single particle distribution function $f_s(t, \boldsymbol{\xi})$, i.e. we must be able to interrupt this coupled chain of equations. This brings us back to investigate the physical interpretation of the right-hand side of eq. (2.41), namely the meaning of $\mathrm{Cov}\big(f_{s',\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_{s'}), f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s)\big)$. We proceed with our computations and evaluate

$$
\begin{aligned}
&\mathrm{E}\left[f_{s',\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_{s'}) f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s)\right] \\
&\quad = \int_V f_{s',\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_{s'}) f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s) D(t, \boldsymbol{\Xi}) \,\mathrm{d}\boldsymbol{\Xi} \\
&\quad = \sum_{i=1}^{N_s} \sum_{j=1}^{N_{s'}} \int_V \delta(\boldsymbol{\xi}' - \boldsymbol{\xi}_{s',j}) \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,i}) D(t, \boldsymbol{\Xi}) \,\mathrm{d}\boldsymbol{\Xi} \\
&\quad = \begin{cases} N_s N_{s'} \displaystyle\int_V \delta(\boldsymbol{\xi}' - \boldsymbol{\xi}_{s',1}) \delta(\boldsymbol{\xi} - \boldsymbol{\xi}_{s,1}) D(t, \boldsymbol{\Xi}) \,\mathrm{d}\boldsymbol{\Xi} & \text{for } s' \neq s \\ f_s^{(2)}(t, \boldsymbol{\xi}, \boldsymbol{\xi}') + \delta(\boldsymbol{\xi} - \boldsymbol{\xi}') f_s(t, \boldsymbol{\xi}) & \text{for } s' = s \end{cases} \\
&\quad =: \begin{cases} f_{s's}^{(2)}(t, \boldsymbol{\xi}', \boldsymbol{\xi}) & \text{for } s' \neq s \\ f_s^{(2)}(t, \boldsymbol{\xi}, \boldsymbol{\xi}') + \delta(\boldsymbol{\xi} - \boldsymbol{\xi}') f_s(t, \boldsymbol{\xi}) & \text{for } s' = s \,. \end{cases}
\end{aligned}
$$

In the third equation we used the result obtained in equation (2.8) and in the last line we extended our definition of the reduced particle distribution function given in eq. (2.4) to multiple species. This means that $f_{s's}^{(2)}(t, \boldsymbol{\xi}, \boldsymbol{\xi}') \,\mathrm{d}\xi \,\mathrm{d}\xi'$ is the likelihood to find an arbitrary particle of component s in the infinitesimal volume $\mathrm{d}\xi$ around $\boldsymbol{\xi}$ and an arbitrary particle of component s' in $\mathrm{d}\xi'$ around $\boldsymbol{\xi}'$. We set $f_{ss}^{(2)} := f_s^{(2)}$ and write more compactly

$$
\mathrm{E}\left[f_{s',\mu}(\boldsymbol{\xi}', \boldsymbol{\Xi}_{s'}) f_{s,\mu}(\boldsymbol{\xi}, \boldsymbol{\Xi}_s)\right] = f_{s's}^{(2)}(t, \boldsymbol{\xi}, \boldsymbol{\xi}') + \delta_{s's} \delta(\boldsymbol{\xi} - \boldsymbol{\xi}') f_s(t, \boldsymbol{\xi}) \,. \tag{2.46}
$$

As before, we assume that the reduced distribution function $f_{s's}^{(2)}$ can be written as

$$
f_{s's}^{(2)}(t, \boldsymbol{\xi}, \boldsymbol{\xi}') = f_s(t, \boldsymbol{\xi}) f_{s'}(t, \boldsymbol{\xi}') + g_{s's}(t, \boldsymbol{\xi}, \boldsymbol{\xi}') \,,
$$

where the product $f_s f_{s'}$ is proportional to the probability of finding particles of the plasma component s and s' assuming statistically independence and $g_{ss'}$ is the two-point correlation function representing statistical dependence of the particles' positions due to mutual interactions. As we will see soon, we have to assume that $g_{ss'}$ is small, in a yet to specify way, and, thus, it can be considered a correction of the 'uncorrelated' probability.

We combine the results of our computations, namely we insert the expression for the expectation value of the product of the two microscopic number number densities, as presented in eq. (2.46), into eq. (2.44). The resulting expression is subsequently plugged into eq. (2.41), which yields

$$\begin{aligned}
&\frac{\partial f_s(t,\boldsymbol{\xi})}{\partial t} + \boldsymbol{v}\cdot\nabla_x f_s(t,\boldsymbol{\xi}) + q_s\boldsymbol{E}(t,\boldsymbol{x})\cdot\nabla_p f_s(t,\boldsymbol{\xi}) \\
&\quad= \nabla_p\cdot\sum_{s'=1}^{N_c}\frac{q_s q_{s'}}{4\pi\epsilon_0}\int\nabla_x\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\,\mathrm{E}\left[\delta f_{s',\mu}(\boldsymbol{\xi}',\Xi_{s'})\delta f_{s,\mu}(\boldsymbol{\xi},\Xi_s)\right]\mathrm{d}^3\xi' \\
&\quad= \nabla_p\cdot\sum_{s'=1}^{N_c}\frac{q_s q_{s'}}{4\pi\epsilon_0}\int\nabla_x\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\,(g_{s's}(t,\boldsymbol{\xi},\boldsymbol{\xi}') + \delta_{s's}\delta(\boldsymbol{\xi}-\boldsymbol{\xi}')f_s(t,\boldsymbol{\xi}))\,\mathrm{d}^3\xi' \\
&\quad= \sum_{s'=1}^{N_c}\frac{q_s q_{s'}}{4\pi\epsilon_0}\int\nabla_x\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\cdot\nabla_p g_{s's}(t,\boldsymbol{\xi},\boldsymbol{\xi}')\,\mathrm{d}^3\xi'\,. \qquad (2.47)
\end{aligned}$$

The last line follows, because

$$\int\nabla_x\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|}\delta(\boldsymbol{x}-\boldsymbol{x}')\,\mathrm{d}^3x' = 0\,.$$

Considering eq. (2.47), we find our speculation that the collision term represents the effects of correlated particle positions confirmed. In the electrostatic approximation, these correlations are due to to the mutual Coulomb interactions of the particles. Setting the right-hand side of equation (2.41) to zero, as we did when we derived the Vlasov equation (2.32), is thus equivalent to ignoring the two-point correlations of the particles' positions, i.e. to ignore inter-particle interactions.

A tenuous gas of *neutral* particles is an example for physical conditions that allow one to expect little correlations of the particles' positions. In such a gas the interaction radius, say r_0, will be less than the inter-particle spacing $n^{-1/3}$. This implies that the number of particles involved in mutual interactions, namely nr_0^3, is small. Since the $g_{s's}$ is a correction to the probability of mutually independent particles, due to mutual interactions among the particles, it is plausible to assume that $g_{s's}$ is proportional to the number of interacting particles, i.e. $g_{s's} \propto nr_0^3$. Furthermore, it is also reasonable to assume that three-point correlations $g_{s''s's}$ are less likely than interactions between two particles. Hence, it is possible to close the chain of coupled equations by setting the higher order correlation functions to zero. (Krall & Trivelpiece, 1973, Sec. 7.4)

For a plasma the situation is different, because the Coulomb force is a *long-range* force. However, the plasma particles mutually shield, or screen out, their respective Coulomb forces. The Coulomb potential of a test particle actually decays exponentially. The e-folding length λ_D is called the *Debye length* or the *Debye radius*. That this is the case can be made plausible with the following thought experiment: We consider a fully ionised and electrically neutral plasma consisting out of electrons and protons. Additionally, we assume that the plasma

is in thermal equilibrium, homogeneous and that it behaves like an ideal gas, i.e. the particles do not interact with each other. In this 'plasma', we immerse a positive charged particle and allow all other particles to interact with it. We expect that it attracts electrons and repels protons. As in a dielectric media the displacement of the electrons and protons introduces an electric field which weakens the electric field of the original particle.

It is possible to quantify this effect in terms of the potential $\Phi(x)$ of the test particle. Since the plasma particles are in thermal equilibrium we expect that they are distributed in phase space in agreement with the Boltzmann (or Gibbs) distribution, i.e.

$$p(x, v) \propto \exp\left(-E(x, v)/k_B T\right) ,$$

where $p(x, v)\,\mathrm{d}r\,\mathrm{d}v$ is the probability to find an electron (or proton) in the infinitesimal volume $\mathrm{d}r\,\mathrm{d}v$ around (x, v) and $E = \pm e\Phi(x) + 1/2mv^2$ is the energy of a proton (or an electron). Since we assumed that the plasma particles do not interact with each other, the potential energy is due to the interactions of the plasma particles with the charged particle that we immersed and located at the origin of the coordinate system. Hence, $\Phi(x)$ is the potential due to the additional charged particle *and* the charge density variation that it entails. This explains as well why $p(x, v)$ only depends on the magnitude of $\boldsymbol{x}$ and not on x, y and z, i.e. we expect the influence of the particle to be spherically symmetric.

We factor out the kinetic energy and integrate over the magnitude of the velocity, because we will compute the potential $\Phi(x)$ with Gauss' law and we, thus, need a charge density. This yields

$$p(x) \propto \exp\left(\mp e\Phi(x)/k_B T\right) .$$

Note that the signs in front of $\Phi(x)$ are exchanged, because of the minus sign in the exponent of the Boltzmann distribution. The ratio of the probability $p(x)\,\mathrm{d}x$ to the probability at $x \to \infty$, i.e. to the probability to find a particle in a volume around a specific point in configuration space that is far enough away from our immersed charge to be uninfluenced by it, yields an expression for the charge density. At $x \to \infty$ the potential of the immersed charge is zero and thus the probability to find particle there is constant (as expected for an ideal and homogeneous gas). We note that the probability $p(x)\,\mathrm{d}x$ must equal $n(x)\,\mathrm{d}^3x/N$, where $n(x)$ is the particle number density and N is the total number of particles. If we introduce the particle number density of protons (or electrons) at $x \to \infty$, namely $n_{0,p} = n_{0,e} = n_0 \coloneqq n(x \to \infty)$, then the probability density function far away from our test charge is $p(x \to \infty) = n_0/N$. The proton number density $n_{0,p}$ equals the electron number density $n_{0,e}$, because we assumed that the plasma is electrically neutral. The ratio is

$$\frac{p(x)}{p(x \to \infty)} = \frac{n(x)}{n_0} = \exp\left(\mp e\Phi(x)/k_B T\right) .$$

Hence the particle number density is $n(x) = n_0 \exp(\mp e\Phi(x)/k_B T)$. It can be interpreted as a deviation from the ideal gas density n_0 due to the influence of the immersed test particle.

We now compute the potential in a shell around the immersed charge. The shell is located at a distance x that is large enough to ensure that the average kinetic energy of the plasma particles, which is proportional to $k_B T$, is much greater than their potential energy $\Phi(x)$. In this shell $n(x) \approx n_0 (1 \mp e\Phi(x)/k_B T)$ and Gauss' law is

$$\begin{aligned} \Delta\Phi(x) &= -e(n_p(x) - n_e(x))/\epsilon_0 \\ &= \Phi(x)\left(\frac{n_{0,p}e^2}{k_B T\epsilon_0} + \frac{n_{0,e}e^2}{k_B T\epsilon_0}\right) \\ &=: \Phi(x)\left(\frac{1}{\lambda_{D,p}^2} + \frac{1}{\lambda_{D,e}^2}\right) =: \frac{\Phi(x)}{\lambda_D^2}\,. \end{aligned} \tag{2.48}$$

In the second line we distinguish between electrons and protons, although their number densities equal. We do so to motivate the expression for the Debye radius of a multi-component plasma. In the last line we define the reciprocal of the Debye radius for each component, namely for protons and electrons, and eventually for all components together. For later reference, we explicitly state that

$$\lambda_{D,s}^2 = \frac{k_B T\epsilon_0}{n_{0,s}e^2} \tag{2.49}$$

$$\frac{1}{\lambda_D^2} = \sum_s \frac{1}{\lambda_{D,s}^2}\,. \tag{2.50}$$

It is possible to solve the homogeneous differential equation (2.48) using spherical coordinates and the ansatz $\Phi(x) = c/x f(x)$, where c is a constant. Note that $f(x)$ modifies the potential of a point charge. This results in the equation

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} - f(x)/\lambda_D^2 = 0\,,$$

whose solution is $f(x) = \exp(-x/\lambda_D)$. The potential created by the immersed charge thus is $\Phi(x) = c/x \exp(-x/\lambda_D)$. The constant c can be determined by noting that the closer we get to the position of the immersed charge the less 'screening' occurs, i.e. $\exp(-x/\lambda_D) \to 1$. This implies that close enough to the immersed particle the potential reduces to the potential of a point charge and, hence, c must equal $e/4\pi\epsilon_0$. (Debye & Hückel, 1923, see in particular §3, eq. 14 for the computation of the constant)

The above thought experiment and its quantitative elaboration informs us that there is a characteristic length scale, namely the Debye radius, which limits microscopic particle interactions to the set of particles contained in a *Debye sphere* (a sphere of Debye radius). Particles further away do not participate,

because of the screening. It is clear, that the assumptions which went into the thought experiment, e.g. thermal equilibrium, do not hold for plasmas which are of interest to our research. The Debye radius, thus, needs to be understood as an approximate length scale. A detailed discussion of how to extend the concept of the Debye radius and its derivation to non-equilibrium plasmas can be found in Meyer-Vernet (1993, Sec. 2).

We return to our original line of thought concerning the closure of the chain of equations that appeared when we investigated the averaged Klimontovich equation in the electrostatic approximation and that is commonly referred to as the BBGKY hierarchy. We pointed out that for short-range interactions the two-point correlation function $g_{s's}$ is expected to be of the order $\mathcal{O}(nr_0^3)$ and since the interaction radius r_0 is small, it is plausible to assume that all higher correlation functions are even smaller and can thus be neglected closing the chain of equations. For a plasma the situation is quite contrary, i.e. the more plasma particles are contained in a Debye sphere, the smaller is the ratio of particles involved in microscopic particle interactions to the overall number of particles. For example, if we keep the temperature of the plasma fixed and increase the density, the interactions get more and more local, i.e. the Debye radius decreases. We define the *plasma parameter*

$$\varkappa \coloneqq 1/\lambda_D^3 n\,, \tag{2.51}$$

where n is the number density of the plasma. Following our considerations we expect the two-point correlation function $g_{s's}$ to be of the order $\mathcal{O}(\varkappa)$ and, in thermal equilibrium, it is possible to find an explicit expression for $g_{s's}$ which, in deed, is proportional to $\varkappa$, see Krall and Trivelpiece (1973, Sec. 3, eq. 2.3.6). The hope is that this carries over to non-equilibrium situations, despite the fact that it has not been demonstrated under which circumstances three-point correlation functions are negligible, see Krall and Trivelpiece (1973, Sec. 7.6) and Montgomery and Tidman (1964, p. 47). If this hope is fulfilled and if $\varkappa \ll 1$, we can interrupt the chain of equations and ignore higher order correlation functions.

2.1.5 THE VLASOV–FOKKER–PLANCK EQUATION

Assuming that it was enough to solve the first two equations of the BBGKY hierarchy, namely eq. (2.47) and eq. (2.45) with a right-hand side excluding the three-point correlation function, does not mean that this is straightforwardly done. In this section we thus explore two more options to compute the collision term. We briefly recapitulate Boltzmann's original approach, because it motivates a Fokker–Planck 'treatment' of the collision term, which we subsequently describe in more detail and which leads to the Vlasov–Fokker–Planck equation. We close this section with the presentation of a relativistic generalisation of the VFP equation.

Boltzmann's idea was to directly compute how the single particle distribution function changes due to particle interactions instead of using a two-point correlation function. For example, the form of the collision term in the centre-of-mass coordinate frame for a two-component plasma, e.g. a fully-ionised hydrogen plasma, is

$$\left(\frac{\delta f_s}{\delta t}\right)_c = \sum_{s'=1}^{2} \int \mathrm{d}^3 v_{s'} \, |\boldsymbol{v}_{s'} - \boldsymbol{v}| \int \mathrm{d}\Omega \frac{\mathrm{d}\sigma_{ss'}}{\mathrm{d}\Omega} \left[f_s(\hat{\boldsymbol{v}}) f_{s'}(\hat{\boldsymbol{v}}_{s'}) - f_s(\boldsymbol{v}) f_{s'}(\boldsymbol{v}_{s'}) \right] \quad (2.52)$$

where $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{v}}_{s'}$ are the velocities after the scattering event and the differential cross section

$$\frac{\mathrm{d}\sigma_{ss'}}{\mathrm{d}\Omega} = \left(\frac{e^2}{8\pi\epsilon_0 \left|\boldsymbol{v}_{s'} - \boldsymbol{v}\right|^2 \mu_{ss'}} \right)^2 \frac{1}{\sin^4(\theta/2)} \quad (2.53)$$

describes Rutherford scattering. $\mu_{ss'} = m_s m_{s'}/(m_s + m_{s'})$ is the reduced mass and θ is the scattering angle in the centre-of-mass coordinate system. (Montgomery & Tidman, 1964, in particular eq. 2.1 and eq. 2.2)

We will not discuss the details of the collision term, but we would like to highlight some aspects of the formula presented in eq. (2.52). The collision term can be thought of as counting the particles which are scattered into the range $(\boldsymbol{v}, \boldsymbol{v} + \mathrm{d}\boldsymbol{v})$ and out of this range. The positive part of the collision term corresponds to the former and the negative part to the latter. Since only products of two single particle distribution functions are considered, the collision term accounts for two-body interactions only. This is related to setting the three-point correlation functions to zero. Moreover, we can interpret the product of the single particle distribution functions as the likelihood that a particle with, say, velocity $\boldsymbol{v}$ interacts with a particle with velocity $\boldsymbol{v}_s$. This implies that we are also neglecting the two-point correlation functions, hence the particle's positions are uncorrelated before their mutual interaction. This assumption is called *molecular chaos*. Hence, we can vaguely think of the collision term as correcting for the effect of two-body interactions assuming that we can compute them without including there influence on the particles' positions. Another remark concerns the binary nature of the collisions: Since the Coulomb force is a long-range force many particles are simultaneously interacting and not only two particles. However, if the deflections contributing most to the integral in the collision term are small, simultaneous and random, they can be simply added together as though two-body sequential collisions were occurring.[3] (Shkarofsky et al., 1966, p. 16–17)

That small-angle scattering contributes most to the integral in eq. (2.52) becomes plausible when we investigate the length scales of the interactions. We define *large-angle scattering* to be a deflection by $\pi/2$ and compute the

[3] Montgomery and Tidman (1964, p. 22) disagree with this view. They state that two-body interactions are only correct for impact parameters b larger than the 90° impact parameter b_0 and smaller than the interparticle spacing d and not for all interactions in the Debye sphere.

corresponding impact parameter b_0 with the formula (see Mayer-Kuckuk, 2002, eq. 1.12)

$$b = \frac{e^2}{4\pi\epsilon_0\mu_{ss'}\,|\boldsymbol{v}_{s'} - \boldsymbol{v}|}\cot\frac{\theta}{2}\,. \tag{2.54}$$

We set $\theta = \pi/2$ and approximate the reduced mass with the electron mass, i.e. $\mu_{ss'} = m_e$. Additionally, we use the root mean square velocity of a plasma in thermal equilibrium to estimate the kinetic energy of the colliding particles to be

$$\mu_{ss'}\,|\boldsymbol{v}_{s'} - \boldsymbol{v}|^2 \approx m_e v_{\mathrm{rms}}^2 = 3k_B T\,.$$

All together, this yields an estimate for the 90° impact parameter

$$b_0 \coloneqq \frac{e^2}{12\pi\epsilon_0 k_B T} = \frac{1}{12\pi}\lambda_{D,e}\varkappa\,, \tag{2.55}$$

where $\lambda_{D,e}$ is the electron Debye radius given in eq. (2.49) and $\varkappa$ is the plasma parameter defined in eq. (2.51). We compare b_0 with the mean inter-particle spacing d, namely

$$d \coloneqq n_0^{-1/3} = \lambda_{D,e}\varkappa^{1/3}\,, \tag{2.56}$$

and find that for a plasma with $\varkappa \ll 1$ the interaction length scales are ordered as

$$b_0 \ll d \ll \lambda_{D,e}\,.$$

Hence, inside the distances for possible microscopic interactions, determined by the Debye radius, 90° deflections are rare, because the inter-particle distance is much greater than the impact parameter b_0. This in turn implies that small angle deflections will dominate the value of the Coulomb collision term given in eq. (2.52). (Montgomery & Tidman, 1964; Shkarofsky et al., 1966, p. 22 and Sec. 1-2.7 respectively)

It is actually possible to evaluate the integrals in the Coulomb collision term, if small-angle scattering is assumed and a minimal deflection angle $\theta_{\min}$, which ensures that the integral does not diverge, is adapted. The minimal deflection angle is a consequence of the Debye screening, i.e. particles with impact parameters larger than the Debye radius need not to be taken into account. Intricate algebraic manipulation lead to an equation of the following form

$$\frac{1}{Y}\left(\frac{\delta f_s}{\delta t}\right)_c = -\frac{\partial}{\partial v^i}\cdot\left(f_s\frac{\partial H_s}{\partial v^i}\right) + \frac{1}{2}\frac{\partial^2}{\partial v^i\,\partial v^j}\left(f_s\frac{\partial^2 G_s}{\partial v^i\,\partial v^j}\right), \tag{2.57}$$

where H_s and G_s are the Rosenbluth–MacDonald–Judd potentials and Y is constant. For details we refer the reader to Shkarofsky et al. (1966, Sec. 7.4) and Montgomery and Tidman (1964, eq. 2.26–2.28). The important point for us is that an equation of this form is called a *Fokker–Planck equation* and the quantity $Y\partial H_s/\partial v^i$ is the *coefficient of dynamical friction* or the *drift vector* and $Y\partial^2 G_s/\partial v^i\partial v^j$ the *dispersion coefficient* or the *diffusion tensor*. We remark that

H_s and G_s depend on f_s and, thus, eq. (2.57) is non-linear. The Fokker–Planck equation is, by definition, a *linear* equation, i.e. eq. (2.57) merely has the *form* of a Fokker–Planck equation.

Fokker–Planck equations appear, for example, when Brownian motion is studied, i.e. the *random* motion of a particle in a fluid whose size is such that it experiences many small deflections without changing its average velocity much. A charged particle in a Debye sphere constantly interacts with many particles causing its velocity to change randomly. This analogy, together with the fact that the collision term has the form of a Fokker–Planck equation, motivates to turn things around and to start with the assumption that the collision term can be computed with a Fokker–Planck equation.

To this end we assume that the probability that the charged particles' velocity $\boldsymbol{v}$ changes by $\Delta\boldsymbol{v}$ in Δt is $\psi(\boldsymbol{v}, \Delta\boldsymbol{v})\,\mathrm{d}^3\Delta v$ and we note that this probability is called a *transition probability*. The fact that ψ does not explicitly depend on time means that the transition probability only depends on the current positions of the plasma particles in phase space, i.e. on the current state of the system. Such a stochastic process is called a *Markov process* and forms an essential assumption when applying the Fokker–Planck formalism. With the probability density ψ at hand we can compute how the single particle distribution function evolves in Δt, i.e.

$$f_s(t + \Delta t, \boldsymbol{x}, \boldsymbol{v}) = \int f_s(t, \boldsymbol{x}, \boldsymbol{v} - \Delta\boldsymbol{v})\psi(\boldsymbol{v} - \Delta\boldsymbol{v}, \Delta\boldsymbol{v})\,\mathrm{d}^3\Delta v\,.$$

We now expand both sides in a Taylor series. The left-hand side is

$$f_s(t + \Delta t, \boldsymbol{x}, \boldsymbol{v}) = f_s(t, \boldsymbol{x}, \boldsymbol{v}) + \left(\frac{\delta f_s}{\delta t}\right)_c \Delta t + \mathcal{O}(\Delta t^2) \tag{2.58}$$

and the right-hand side equals

$$\begin{aligned}
&f_s(t, \boldsymbol{x}, \boldsymbol{v} - \Delta\boldsymbol{v})\psi(\boldsymbol{v} - \Delta\boldsymbol{v}) \\
&\quad = f_s(t, \boldsymbol{x}, \boldsymbol{v})\psi(\boldsymbol{v}, \Delta\boldsymbol{v}) - \left(\frac{\partial f_s}{\partial \boldsymbol{v}}\psi(\boldsymbol{v}, \Delta\boldsymbol{v}) + f_s(t, \boldsymbol{x}, \boldsymbol{v})\frac{\partial \psi}{\partial \boldsymbol{v}}\right) \cdot \Delta\boldsymbol{v} \\
&\qquad + \frac{1}{2}\left(\frac{\partial^2 f_s}{\partial v^i\,\partial v^j}\psi(\boldsymbol{v}, \Delta\boldsymbol{v}) + 2\frac{\partial f_s}{\partial v^i}\frac{\partial \psi}{\partial v^j} + f_s(t, \boldsymbol{x}, \boldsymbol{v})\frac{\partial^2 \psi}{\partial v^i\,\partial v^j}\right)\Delta v^i \Delta v^j + \mathcal{O}\left(|\Delta\boldsymbol{v}|^3\right) \\
&\quad = f_s(t, \boldsymbol{x}, \boldsymbol{v})\psi(\boldsymbol{v}, \Delta\boldsymbol{v}) - \frac{\partial}{\partial \boldsymbol{v}}\left(f_s(t, \boldsymbol{x}, \boldsymbol{v})\psi(\boldsymbol{v}, \Delta\boldsymbol{v})\right) \cdot \Delta\boldsymbol{v} \\
&\qquad + \frac{1}{2}\frac{\partial^2}{\partial v^i\,\partial v^j}\left(f_s(t, \boldsymbol{x}, \boldsymbol{v})\psi(\boldsymbol{v}, \Delta\boldsymbol{v})\right)\Delta v^i \Delta v^j + \mathcal{O}\left(|\Delta\boldsymbol{v}|^3\right).
\end{aligned} \tag{2.59}$$

We drop the remainders of the Taylor series and integrate the right-hand side over $\mathrm{d}^3\Delta v$. This results in

$$\left(\frac{\delta f_s}{\delta t}\right)_c \Delta t = -\frac{\partial}{\partial \boldsymbol{v}}\left(f_s(t, \boldsymbol{x}, \boldsymbol{v})\mathrm{E}\left[\Delta\boldsymbol{v}\right]\right) + \frac{1}{2}\frac{\partial^2}{\partial v^i\,\partial v^j}\left(f(t, \boldsymbol{x}, \boldsymbol{v})\mathrm{E}\left[\Delta v^i \Delta v^j\right]\right)$$

where we used that the transition probability is normalised, namely

$$\int \psi(\boldsymbol{v}, \Delta\boldsymbol{v})\, \mathrm{d}^3\Delta v = 1\,.$$

We define the components of the drift vector and the diffusion tensor to be

$$F^i \coloneqq \mathrm{E}\left[\Delta v^i\right]/\Delta t \tag{2.60}$$
$$D^{ij} \coloneqq \mathrm{E}\left[\Delta v^i \Delta v^j\right]/2\Delta t\,. \tag{2.61}$$

We note that in general the transition probability ψ may depend on $\boldsymbol{x}$ implying a dependence of the drift vector and the diffusion tensor on the position in configuration space. With these definitions at hand we obtain the Fokker–Planck equation, i.e.

$$\left(\frac{\delta f_s}{\delta t}\right)_c = -\frac{\partial}{\partial \boldsymbol{v}}\left(f_s(t,\boldsymbol{x},\boldsymbol{v})\boldsymbol{F}(\boldsymbol{x},\boldsymbol{v})\right) + \frac{\partial^2}{\partial v^i\, \partial v^j}\left(f_s(t,\boldsymbol{x},\boldsymbol{v})D^{ij}(\boldsymbol{x},\boldsymbol{v})\right)\,. \tag{2.62}$$

We emphasise that the computation of the drift vector and the diffusion tensor requires an explicit expression for the transition probability ψ. An example is provided in Sec. 3.2.2 in which we exemplify the computation of the diffusion tensor.

If we replace the collision term in eq. (2.31) with the Fokker–Planck expression of eq. (2.62), we get the *Vlasov–Fokker–Planck equation*, i.e.

$$\frac{\partial f_s(t,\boldsymbol{\xi})}{\partial t} + \boldsymbol{v}\cdot\nabla_x f_s(t,\boldsymbol{\xi}) + q_s\left(\boldsymbol{E}(t,\boldsymbol{x}) + \boldsymbol{B}^{\mathrm{ext}}(t,\boldsymbol{x})\right)\cdot\nabla_p f_s(t,\boldsymbol{\xi}) = \left(\frac{\delta f_s}{\delta t}\right)_c\,. \tag{2.63}$$

Note we added an external **B**-field. 'Ext' means that it is not created by the plasma particles, but by other processes like dynamos in stars or coils in fusion devices. This implies that it is independent of the plasma particles' positions and, thus, it can be taken out of the integral when averaging the Klimontovich equation to derive the Vlasov equation.

Because we are interested in relativistic particles, we end this section with discussing a relativistic generalisation of the VFP equation. If the plasma is relativistic the Coulomb approximation breaks down, i.e. the magnetic field generated by the particles contributes as much as the $\boldsymbol{E}$-field to the Lorentz force. Since the derivation of the BBGKY hierarchy was done in the Coulomb approximation it is unclear if a similar chain of equations arises if relativistic plasmas are considered and thus if the mutual particle interactions can still be represented with a Fokker–Planck collision term. We did not find any literature generalising the BBGKY hierarchy. However, we use the VFP equation to model charged and energetic particles that interact with a background plasma. This means that the interactions of interest to us are not mutual interactions of particles, but interactions of the charged and energetic particles with MHD waves propagating in the background plasma. If it is possible to model the

changes these interactions cause using a Fokker–Planck collision term, then the relativistic generalisation of the VFP equation, which we present now, is applicable to the distribution function of the energetic and charged particles.

The relativistic Vlasov–Fokker–Planck equation we have in mind is

$$\frac{p^\mu}{m}\frac{\partial \mathcal{F}_s}{\partial x^\mu} + \frac{q}{m}F_\mu{}^\nu p_\nu \frac{\partial \mathcal{F}_s}{\partial p_\mu} = \left(\frac{\delta \mathcal{F}_s}{\delta \tau}\right)_c , \tag{2.64}$$

where $\mathrm{d}N = \mathcal{F}_s(x^\mu, p_\mu)\,\mathrm{d}^4x\,\mathrm{d}^4p$ is the average number of particles of component s in the eight dimensional phase space parameterised by the variables (x^μ, p_μ). $F^{\mu\nu}$ is the electromagnetic field tensor and

$$qF_\mu{}^\nu p_\nu = q\gamma\,(\boldsymbol{E}\cdot\boldsymbol{v}, -[\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}])$$

is the Lorentz force, see Achterberg and Norman (2018, eq. 24). The relativistic Vlasov–Fokker–Planck equation can be derived exactly as the Vlasov equation, i.e. starting with the conservation of particles and averaging it of all possible particle positions in the eight dimensional phase space. This is done with a probability density function $\mathcal{D}(t, \dots, \xi^\mu_{s,1}, \dots, \xi^\mu_{s,N_s}, \dots)$ with $\xi^\mu_{s,i} = (x^\mu_{s,i}, p^\mu_{s,i})$. For details we refer the reader to Klimontovich (1967, p.53–55).

We emphasise that four dimensions in momentum space is one dimension too much for the plasmas of interest to our research. The reason is that for non-virtual massive particles the relativistic energy-momentum relation holds, i.e. $E^2 = p^2c^2 + m^2c^4$. Klimontovich (1967) directly addresses this with a different definition of $\mathcal{F}_s$, we follow Achterberg and Norman (2018) and remedy this by setting

$$\mathcal{F}_s(x^\mu, p_\mu) = 2mH(E)f(t, \boldsymbol{x}, \boldsymbol{p})\delta(p^\mu p_\mu - m^2c^2)\,. \tag{2.65}$$

The unit step function $H(E)$ enforces that the energies of the particles is positive and the evaluation of the argument of the delta distribution shows us that the relativistic energy-momentum relation is imposed, namely $\delta(p^\mu p_\mu - m^2c^2) = \delta\left(E^2/c^2 - (p^2 + m^2c^2)\right)$.

So far we reduced the statistical description of a plasma to a set of differential equations that determines the single particle distribution functions of its components. Under specific plasma conditions, yet to be named, it is possible to treat the plasma as a fluid and, hence, to work with a conservation of mass, a conservation of momentum and a conservation of energy equation. This will be explored in the next part of this chapter.

2.2 MAGNETOHYDRODYNAMICS

If collisions of plasma particles are frequent enough, it is possible to distinguish between their random thermal motion and the motion of their centre-of-mass. This distinction is at the core of a fluid model. Thermal effects will show up as

static pressure, thermal energy etc. The bulk will be characterised by a velocity and a density. (Shkarofsky et al., 1966, p. 52)

In this part of chapter, we restrict ourselves to a two component plasma, namely a plasma consisting of ionized hydrogen and electrons. Moreover, we assume that the velocity of the particles is much less than the speed of light, i.e. $\boldsymbol{p} \approx m\boldsymbol{v}$. In the context of non-relativistic plasmas it is therefore common to work with a single particle distribution function that depends on $\boldsymbol{v}$ instead of $\boldsymbol{p}$, see for example eq. (3.10) in Thorne and Blandford (2017) which is

$$\mathrm{d}N = f_s(t, \boldsymbol{x}, \boldsymbol{p})\,\mathrm{d}^3p = f_s(t, \boldsymbol{x}, m\boldsymbol{v})m^3\,\mathrm{d}^3v \coloneqq \mathfrak{f}_s(t, \boldsymbol{x}, \boldsymbol{v})\,\mathrm{d}^3v\,.$$

We also restrain ourselves to a *neutral* plasma, i.e. the number density of protons equals the number density of electrons $n_p = n_e$.

2.2.1 INTRINSIC VELOCITY

A key concept for distinguishing the motion of the centre-of-mass of the plasma particles from their random thermal motions is the intrinsic velocity. We follow Shkarofsky et al. (1966, Sect. 2.3) and start with a row of definitions: First, we define the velocity average of an arbitrary function $\eta(\boldsymbol{v})$ to be

$$\bar{\eta}_s(t, \boldsymbol{x}) \coloneqq \frac{1}{n_s(t, \boldsymbol{x})} \int \mathfrak{f}_s(t, \boldsymbol{x}, \boldsymbol{v})\eta(\boldsymbol{v})\,\mathrm{d}^3v\,, \tag{2.66}$$

secondly, we define the lth *velocity moment* of the single particle distribution function as

$$\mathfrak{V}_s^{i_1\cdots i_l}(t, \boldsymbol{x}) \coloneqq \int v^{i_1}\cdots v^{i_l}\mathfrak{f}_s(t, \boldsymbol{x}, \boldsymbol{v})\,\mathrm{d}^3v\,. \tag{2.67}$$

We note that the number density n_s is the zeroth moment of $\mathfrak{f}_s$. Furthermore, we define a *fluid element* to be the number of particles in the infinitesimal volume d^3x at $\boldsymbol{x}$ and specify that when we refer to the centre-of-mass of the plasma particles we mean the centre-of-masses of the fluid elements. The total momentum of a fluid element is the first velocity moment of the single particle distribution function times m, i.e.

$$m_s n_s(t, \boldsymbol{x})\bar{\boldsymbol{v}}_s = m_s \int \boldsymbol{v}\mathfrak{f}_s(t, \boldsymbol{x}, \boldsymbol{v})\,\mathrm{d}^3v\,,$$

where $m\bar{\boldsymbol{v}}_s$ is the average momentum of a particle in the fluid element located at $\boldsymbol{x}$, cf. eq. (2.66). Third, we define the *bulk velocity*

$$\boldsymbol{U}(t, \boldsymbol{x}) \coloneqq \frac{m_e n_e(t, \boldsymbol{x})\bar{\boldsymbol{v}}_e + m_p n_p(t, \boldsymbol{x})\bar{\boldsymbol{v}}_p}{n_e m_e + n_p m_p}\,. \tag{2.68}$$

Notice that this is the centre-of-mass velocity of the fluid elements of the two component plasma that we are considering. Fourth, the relation between the

motion of the bulk of the particles and their individual motions can now be quantified in terms of the *intrinsic* (or *peculiar*) *velocity* that is defined as

$$\boldsymbol{w}(t,\boldsymbol{x},\boldsymbol{v}) \coloneqq \boldsymbol{v} - \boldsymbol{U}(t,\boldsymbol{x}) \tag{2.69}$$

and depends on $t, \boldsymbol{x}$ and the velocity $\boldsymbol{v}$. Fifth, we also introduce the *density of a fluid element*, i.e.

$$\rho(t,\boldsymbol{x}) \coloneqq m_e n_e(t,\boldsymbol{x}) + m_p n_p(t,\boldsymbol{x})\,. \tag{2.70}$$

As pointed out in the introduction to this section, we are interested in separating the effects of the motion of the bulk of the particles from their 'thermal' motion. This requires us to change the velocity dependence of the single particle distribution function from $\boldsymbol{v}_s$ to $\boldsymbol{w}_s$. This change of variables modifies the VFP equation (2.63). In Shkarofsky et al. (1966, eq. 2-41a - c) we find that the time derivative becomes

$$\frac{\partial}{\partial t}\mathfrak{f}_s(t,\boldsymbol{x},\boldsymbol{w}(t,\boldsymbol{x},\boldsymbol{v})) = \frac{\partial w^i}{\partial t}\frac{\partial \mathfrak{f}_s}{\partial w^i} = -\frac{\partial U^i}{\partial t}\frac{\partial \mathfrak{f}_s}{\partial w^i}$$

and the adapted spatial derivatives are

$$\frac{\partial}{\partial x^i}\mathfrak{f}_s(t,\boldsymbol{x},\boldsymbol{w}(t,\boldsymbol{x}),\boldsymbol{v}) = \frac{\partial w^j}{\partial x^i}\frac{\partial \mathfrak{f}_s}{\partial w^j} = -\frac{\partial U^j}{\partial x^i}\frac{\partial \mathfrak{f}_s}{\partial w^j}\,.$$

The velocity derivatives do not change, because

$$\frac{\partial}{\partial v^i}\mathfrak{f}_s(t,\boldsymbol{x},\boldsymbol{w}(t,\boldsymbol{x},\boldsymbol{v})) = \frac{\partial w^j}{\partial v^i}\frac{\partial \mathfrak{f}_s}{\partial w^j} = \delta_i^j\frac{\partial \mathfrak{f}_s}{\partial w^j} = \frac{\partial}{\partial w^i}\mathfrak{f}_s\,.$$

We also must change $\boldsymbol{v}$ to $\boldsymbol{w}+\boldsymbol{U}$ and the Lorentz force becomes

$$\boldsymbol{F}_L' = q_s\left(\boldsymbol{E}' - \boldsymbol{U}\times\boldsymbol{B}' + (\boldsymbol{w}+\boldsymbol{U})\times\boldsymbol{B}'\right) = q_s\left(\boldsymbol{E}' + \boldsymbol{w}\times\boldsymbol{B}\right)\,,$$

where the primed quantities are defined in the centre-of-mass frame of the fluid element. We note that in the limit of an infinite speed of light, i.e. $c\to\infty$, $\boldsymbol{B}' = \boldsymbol{B}$. [4] The transformed VFP equation is

$$\frac{\mathrm{D}\mathfrak{f}_s}{\mathrm{D}t} + \boldsymbol{w}\cdot\nabla_x\mathfrak{f}_s + \left(\frac{q_s}{m_s}\left(\boldsymbol{E}' + \boldsymbol{w}\times\boldsymbol{B}\right) - \frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t} - \boldsymbol{J}_U\boldsymbol{w}\right)\cdot\nabla_w\mathfrak{f}_s = \left(\frac{\delta\mathfrak{f}_s}{\delta t}\right)_c\,, \tag{2.71}$$

where $\mathrm{d}D/\,\mathrm{d}Dt \coloneqq \partial/\partial t + \boldsymbol{U}\cdot\nabla_x$ is the material derivative, $(\boldsymbol{J}_U)_{ij} \coloneqq \partial U^i/\partial x^j$ is the Jacobian matrix of the fluid velocity $\boldsymbol{U}$ and the factor $1/m_s$ in front of the Lorentz

[4] In Shkarofsky et al. (1966) and other textbooks only the velocity $\boldsymbol{v}_s$ is changed, the electromagnetic fields are not transformed. This gives $\boldsymbol{E}+\boldsymbol{U}\times\boldsymbol{B}+\boldsymbol{w}\times\boldsymbol{B}$. This is equivalent to our transformation of the Lorentz force, since $\boldsymbol{E}' = \boldsymbol{E}+\boldsymbol{U}\times\boldsymbol{B}+\mathcal{O}((U/c)^2)$. We choose to transform the electromagnetic fields, because changing coordinates in velocity space also requires a change of the force. For example, let $p'^\mu = A^\mu_\nu p^\nu$ be a constant transformation, i.e. a A^μ_ν is independent of $\boldsymbol{x}$ and t, then the transformation of Newtons second's law is $\mathrm{d}p'^\mu/\,\mathrm{d}\tau = \mathrm{d}A^\mu_\nu p^\nu/\,\mathrm{d}\tau = A^\mu_\nu K^\nu = K'^\mu$, where K^μ is a four-force. We also refer the reader to eq. (28) in Achterberg and Norman (2018) which shows the transformation of the Lorentz force.

force appears because we are working with a velocity dependent single particle distribution function, see Shkarofsky et al. (1966, p. 60). The $\boldsymbol{U}$-dependent terms in front of the velocity derivative ∇_w appeared, because the definition of the intrinsic velocity can be understood as a coordinate transformation into a non-inertial frame, see eq. (2.69). A transformation into a non-inertial frame leads to a fictitious force, namely

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{p}'}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} m\boldsymbol{w} &= m\left(\frac{\partial \boldsymbol{w}}{\partial t} + \frac{\partial x^j}{\partial t}\frac{\partial \boldsymbol{w}}{\partial x^j}\right) \\ &= m\left(-\frac{\partial \boldsymbol{U}}{\partial t} - v^j \frac{\partial \boldsymbol{U}}{\partial x^j}\right) = -m\left(\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t} + \boldsymbol{J}_U \boldsymbol{w}\right) =: \boldsymbol{F}'_f\,,\end{aligned}$$

where we investigated the momentum of a *free* a particle, namely $\mathrm{d}\boldsymbol{p}/\mathrm{d}t = 0$, to isolate the fictitious force. Note that the fluid velocity dependent terms in front of ∇_w are exactly $\boldsymbol{F}'_f/m_s$.

2.2.2 INTRINSIC VELOCITY MOMENTS

The next step towards a fluid description of our plasma is to compute the *intrinsic* velocity moments of eq. (2.71). We show in the next section that its zeroth intrinsic velocity moment yields the mass conservation equation, its first moment the momentum conservation equation and its second moment the energy conservation equation. In this section, we derive an equation for the lth intrinsic velocity moment. This equation demonstrates that the lth intrinsic velocity moment of $\mathfrak{f}_s$ does depend on its $(l+1)$th moment; again resulting in a chain of equations, that needs to be closed with appropriate physical assumptions about the plasma to be described.

We start and multiply eq. (2.71) with $m_s w^{i_1} \cdots w^{i_l}$ and integrate over $\mathrm{d}^3 w$. This yields for the time derivative term

$$\int \frac{\mathrm{D}\mathfrak{f}_s}{\mathrm{D}t} m_s w^{i_1} \cdots w^{i_l}\, \mathrm{d}^3 w = \frac{\mathrm{D}}{\mathrm{D}t}\left(\frac{1}{n_s}\int n_s m_s w^{i_1} \cdots w^{i_l} \mathfrak{f}_s\, \mathrm{d}^3 w\right) =: \frac{\mathrm{D}}{\mathrm{D}t} \mathcal{W}_s^{i_1 \cdots i_l}\,,$$

where we applied Leibniz integral rule to pull the total time derivative out of the integral. Moreover, we extended the definition of the velocity average, as presented in eq. (2.66), to the intrinsic velocity $\boldsymbol{w}$ and, based on the definition of the velocity moments in eq. (2.67), we also defined $\mathcal{W}_s^{i_1 \cdots i_l} := n_s m_s w^{i_1} \cdots w^{i_l}$. Analogously, the integral of the spatial advection term is

$$\int m_s w^{i_1} \cdots w^{i_l} w^j \frac{\partial \mathfrak{f}_s}{\partial x^j}\, \mathrm{d}^3 w_s = \frac{\partial}{\partial x^j}\int m_s w^{i_1} \cdots w^{i_l} w^j \mathfrak{f}_s\, \mathrm{d}^3 w_s = \frac{\partial}{\partial x^j} \mathcal{W}_s^{j i_1 \cdots i_l}\,.$$

The integral of the Lorentz force term is

$$\begin{aligned}
&\int q_s w^{i_1} \cdots w^{i_l} (\boldsymbol{E}' + \boldsymbol{w} \times \boldsymbol{B}) \cdot \nabla_w \mathfrak{f}_s \, \mathrm{d}^3 w \\
&\quad = \int_{\partial V} q_s \left(w^{i_1} \cdots w^{i_l} (\boldsymbol{E}' + \boldsymbol{w} \times \boldsymbol{B}) \mathfrak{f}_s \right) \cdot \mathrm{d}\boldsymbol{S} \\
&\qquad - \int q_s \nabla_w \cdot \left(w^{i_1} \cdots w^{i_l} (\boldsymbol{E}' + \boldsymbol{w} \times \boldsymbol{B}) \right) \mathfrak{f}_s \, \mathrm{d}^3 w \\
&\quad = -q_s E'^k \int \frac{\partial}{\partial w^k} \left(w^{i_1} \cdots w^{i_l} \right) \mathfrak{f}_s \, \mathrm{d}^3 w \\
&\qquad - q_s \epsilon^k{}_{mn} B^n \int \frac{\partial}{\partial w^k} \left(w^{i_1} \cdots w^{i_l} w^m \right) \mathfrak{f}_s \, \mathrm{d}^3 w \,.
\end{aligned}$$

Integration by parts yields the surface integral in the second line. It disappears, because the single particle distribution function $\mathfrak{f}_s$ is zero at infinity. We note that raising or lowering an index of the Levi–Civita symbol does not change its sign, i.e. $\epsilon^k{}_{lm} = \epsilon_{klm}$, because we are in Euclidean space and the metric is δ^i_j. The evaluation of the integrals gives

$$\begin{aligned}
-q_s E'^k \int \frac{\partial}{\partial w^k} \left(w^{i_1} \cdots w^{i_l} \right) \mathfrak{f}_s \, \mathrm{d}^3 w &= -q_s E'^k \int \sum_{m=1}^{l} \delta^{i_m}_k \prod_{\substack{n=1 \\ n \neq m}} w^{i_n} \mathfrak{f}_s \, \mathrm{d}^3 w \\
&= -n_s q_s \sum_{m=1}^{l} E'^{i_m} \overline{\prod_{\substack{n=1 \\ n \neq m}} w^{i_n}} \\
&= -\frac{q_s}{m_s} l \bar{\mathcal{W}}_s^{(i_1 \cdots i_{l-1}} E'^{i_l)} \,.
\end{aligned}$$

Note the 'unusual' usage of the sum and the product symbol in the second line, i.e. we sum over or multiply different indices and *not* different values of indices. The notation used in the definition of the tensor in the last line is taken from eq. (1.7a) of Thorne (1980), namely

$$S_{ab(cde)} \coloneqq \frac{1}{3!} \left(S_{abcde} + S_{abced} + S_{abdce} + S_{abdec} + S_{abecd} + S_{abedc} \right) . \tag{2.72}$$

$\bar{\mathcal{W}}_s^{(i_1 \cdots i_{l-1}} E'^{i_l)}$ should hence be read as the 'symmetrisation' of a tensor with components $\bar{\mathcal{W}}_s^{i_1 \cdots i_{l-1}} E'^{i_l}$. The equivalence

$$l \bar{\mathcal{W}}_s^{(i_1 \cdots i_{l-1}} E'^{i_l)} = n_s m_s \sum_{m=1}^{l} E'^{i_m} \overline{\prod_{\substack{n=1 \\ n \neq m}}^{l} w^{i_n}}$$

is a consequence of the symmetry of $\bar{\mathcal{W}}_s^{i_1 \cdots i_{l-1}}$. The symmetry implies that there are $(l-1)!$ copies of each term on the left-hand side of the equation and, by

definition (2.72), the left-hand side is also divided by $l!$. In total each term is multiplied with $(l-1)!/l! = 1/l$. In a completely analogue manner, the magnetic force term is turned into

$$-\frac{q_s}{m_s} l \bar{\mathcal{W}}_s^{m(i_1\cdots i_{l-1}} \epsilon^{i_l)}{}_{mn} B^n \,,$$

where the index m is outside the parentheses, because contracting with the Levi–Civita symbol yields zero, i.e. $\epsilon^m{}_{mn} = 0$.

We repeat the above computations for the fictitious force terms and get

$$-\int m_s w^{i_1} \cdots w^{i_l} \frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t} \cdot \nabla_w \mathfrak{f}_s \, \mathrm{d}^3 w = l \bar{\mathcal{W}}_s^{(i_1\cdots i_{l-1}} \mathrm{D}U^{i_l)}/\mathrm{D}t \,.$$

It has the same structure as the $\boldsymbol{E}'$-field term, but with $\boldsymbol{E}'$ replaced by $\mathrm{D}\boldsymbol{U}/\mathrm{D}t$. Note that the convective derivative $\mathrm{D}/\mathrm{D}t$ only acts on the fluid velocity $\boldsymbol{U}$. Moreover,

$$-\int m_s w^{i_1} \cdots w^{i_l} \boldsymbol{J}_U \boldsymbol{w} \cdot \nabla_v \mathfrak{f}_s \, \mathrm{d}^3 w = l \bar{\mathcal{W}}_s^{j(i_1\cdots i_{l-1}} \partial U^{i_l)}/\partial x^j + \nabla_x \cdot \boldsymbol{U} \bar{\mathcal{W}}_s^{i_1\cdots i_l} \,.$$

To handle the collision term on the right-hand side, we define

$$\frac{1}{n_s} \int n_s m_s w^{i_1} \cdots w^{i_l} \frac{\delta \mathfrak{f}_s}{\delta t} \, \mathrm{d}^3 w =: \frac{\delta}{\delta t} \bar{\mathcal{W}}_s^{i_1\cdots i_l} \,,$$

which symbolically represents the changes of the velocity moment caused by the mutual interactions of the plasma particles.

Putting all the terms together results in the chain of equations that determines the intrinsic velocity moments of $\mathfrak{f}_s$, namely

$$\begin{aligned} \frac{\mathrm{D}}{\mathrm{D}t} \bar{\mathcal{W}}_s^{i_1\cdots i_l} + \frac{\partial}{\partial x^j} \bar{\mathcal{W}}_s^{j i_1\cdots i_l} - \frac{q_s}{m_s} \Big(l \bar{\mathcal{W}}_s^{(i_1\cdots i_{l-1}} E'^{i_l)} + l \bar{\mathcal{W}}_s^{m(i_1\cdots i_{l-1}} \epsilon^{i_l)}{}_{mn} B^n \Big) \\ + l \bar{\mathcal{W}}_s^{(i_1\cdots i_{l-1}} \mathrm{D}U^{i_l)}/\mathrm{D}t + l \bar{\mathcal{W}}_s^{j(i_1\cdots i_{l-1}} \partial U^{i_l)}/\partial x^j + \nabla_x \cdot \boldsymbol{U} \bar{\mathcal{W}}_s^{i_1\cdots i_l} \\ = \frac{\delta}{\delta t} \bar{\mathcal{W}}_s^{i_1\cdots i_l} \,, \end{aligned} \tag{2.73}$$

compare with Shkarofsky et al. (1966, eq. 9-4a). We emphasise that the lth intrinsic velocity moment depends on the $(l+1)$th intrinsic velocity moment via the second term of eq. (2.73).

2.2.3 FLUID EQUATIONS

In this section we use eq. (2.73) to compute the fluid equations (mass, momentum and energy conservation) of the fully ionized hydrogen plasma that we are currently considering, i.e. we compute the intrinsic velocity moments for $l = 0, 1$ and 2. We do so for the electron and proton component and directly add them up. We begin with defining the following symbols for the average intrinsic

velocity moments and derived quantities (Shkarofsky et al., 1966, eq. 9.5a-e and eq. 9.12-13):

$$
\begin{aligned}
(\boldsymbol{\pi}_s)^i &\coloneqq n_s m_s u_s^i \coloneqq n_s m_s \overline{w^i} = \bar{\mathcal{W}}_s^i && \text{(average intrinsic momentum/velocity)}\\
\pi_s^{ij} &\coloneqq n_s m_s \overline{w^i w^j} = \bar{\mathcal{W}}_s^{ij} && \text{(intrinsic pressure tensor)}\\
\Pi^{ij} &\coloneqq \sum_s \pi_s^{ij} && \text{(total pressure tensor)}\\
p_s &\coloneqq \pi_{s,i}^i/3 = n_s m_s \overline{w^2}/3 && \text{(intrinsic or static pressure)}\\
P &\coloneqq \sum_s p_s = \Pi^i{}_i/3 && \text{(total static pressure)}\\
q_s^{ijk} &\coloneqq n_s m_s \overline{w^i w^j w^k} = \bar{\mathcal{W}}_s^{ijk} && \text{(intrinsic heat flow tensor)}\\
Q^{ijk} &\coloneqq \sum_s q_s^{ijk} && \text{(total heat flow tensor)}\\
(\boldsymbol{q}_s)^i &\coloneqq q_{s,j}^{ij}/2 = n_s m_s \overline{w^2 w^i}/2 && \text{(intrinsic heat flow vector)}\\
(\boldsymbol{Q})^i &\coloneqq \sum_s (\boldsymbol{q}_s)^i = Q^{ij}{}_j/2 && \text{(total heat flow vector)}
\end{aligned}
\tag{2.74}
$$

In the course of the following computations we will repeatedly use that the sum of the intrinsic momenta is zero, i.e.

$$\sum_s \boldsymbol{\pi}_s = 0\,.$$

The reason is that the bulk velocity, relative to which $\boldsymbol{w}$ is defined, is the centre-of-mass velocity of the fluid elements.[5]

For the $l = 0$ case, we get

$$\frac{\mathrm{D}}{\mathrm{D}t}\left(n_e m_e + n_p m_p\right) + \frac{\partial}{\partial x^i}\left(\pi_e^i + \pi_p^i\right) + \nabla_x \cdot \boldsymbol{U}(n_e m_e + n_p m_p) = \frac{\delta n_e m_e}{\delta t} + \frac{\delta n_p m_p}{\delta t}\,,$$

Employing the definition of the plasma density, see eq. (2.70), and taking into account that the intrinsic momenta sum to zero and that no electrons or protons are created in collisions, the above equation becomes a mass conservation equation, namely

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \nabla_x \cdot \boldsymbol{U}\rho = \frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho \boldsymbol{U}) = 0\,, \tag{2.75}$$

see Shkarofsky et al. (1966, eq. 9-16).

[5] Explicitly computing

$$\boldsymbol{\pi}_s = \frac{1}{n_s}\int n_s m_s \boldsymbol{w} \mathfrak{f}_s \,\mathrm{d}^3 w = m_s n_s \bar{\boldsymbol{v}}_s - n_s m_s \boldsymbol{U}(t, \boldsymbol{x})\,,$$

where we used eq. (2.66) and that $\mathrm{d}^3 w = \mathrm{d}^3 v$ to get $\bar{\boldsymbol{v}}_s$, and summing $\boldsymbol{\pi}_e$ and $\boldsymbol{\pi}_p$ gives the claimed result, see the definition of $\boldsymbol{U}$ presented in eq. (2.68).

For the $l = 1$ case, we get

$$\frac{\partial}{\partial x^j}\left(\pi_e^{ji} + \pi_p^{ji}\right) - e\left(n_p - n_e\right)E'^i - e\epsilon^i{}_{mn}\left(n_p u_p^m - n_e u_e^m\right)B^n + \left(n_e m_e + n_p m_p\right)\frac{\mathrm{D}U^i}{\mathrm{D}t} = \frac{\delta(\boldsymbol{\pi}_e)^i}{\delta t} + \frac{\delta(\boldsymbol{\pi}_p)^i}{\delta t}\,.$$

We again exploited that the sum of the intrinsic momenta is zero. Furthermore, the assumption that the plasma is neutral, i.e. $n_e = n_p$, implies that the second term vanishes. With the introduction of the definition of the *intrinsic current density*, namely

$$\boldsymbol{J}' \coloneqq en_p\boldsymbol{u}_p - en_e\boldsymbol{u}_e\,, \tag{2.76}$$

the equation for the first velocity moment becomes the momentum conservation equation of the plasma, i.e.

$$\rho\frac{\mathrm{D}U^i}{\mathrm{D}t} + \frac{\partial}{\partial x^j}\Pi^{ji} - \epsilon^i{}_{mn}J'^m B^n = 0\,, \tag{2.77}$$

where we employed the definition of the total pressure tensor Π^{ij} (Shkarofsky et al., 1966, eq. 9-17). The right-hand side equals zero, because we presuppose that the change of the electron momenta due to collisions is balanced by a corresponding change of the proton momenta, i.e. $\delta\boldsymbol{\pi}_e/\delta t = -\,\delta\boldsymbol{\pi}_p/\delta t$. Before proceeding with the second velocity moment, we remark that the intrinsic current density equals the current density defined in the laboratory frame[6] in the approximations we made, namely the neutrality of the plasma and that the speed of light is infinite (Thorne & Blandford, 2017, eq. 19.3d). This becomes clear in the light of the following computation:[7]

$$\boldsymbol{J}' = en_p\left(\bar{\boldsymbol{v}}_p - \boldsymbol{U}\right) - en_e\left(\bar{\boldsymbol{v}}_e - \boldsymbol{U}\right) = en_p\bar{\boldsymbol{v}}_p - en_e\bar{\boldsymbol{v}}_e - e\boldsymbol{U}\left(n_p - n_e\right) =: \boldsymbol{J}\,. \tag{2.78}$$

For the $l = 2$ case, we get

$$\begin{aligned}\frac{\mathrm{D}}{\mathrm{D}t}\Pi^{ij} + \frac{\partial}{\partial x^k}Q^{kij} - \left(j^i E'^j + J'^j E'^i\right) \\ - \frac{e}{m_p}\left(\pi_p^{mi}\epsilon^j{}_{mn} + \pi_p^{mj}\epsilon^i{}_{mn}\right)B^n + \frac{e}{m_e}\left(\pi_e^{mi}\epsilon^j{}_{mn} + \pi_e^{mj}\epsilon^i{}_{mn}\right)B^n \\ + \Pi^{ki}\frac{\partial U^j}{\partial x^k} + \Pi^{kj}\frac{\partial U^i}{\partial x^k} + \nabla_x\cdot\boldsymbol{U}\Pi^{ij} = \frac{\delta}{\delta t}\pi_e^{ij} + \frac{\delta}{\delta t}\pi_p^{ij}\,.\end{aligned} \tag{2.79}$$

We simplify the above equation with the assumption that the electrons and protons are scattered frequently enough[8] to keep their single particle distribution

[6] We call the inertial frame in which we originally defined $t, \boldsymbol{x}$ and $\boldsymbol{v}$ the laboratory frame.

[7] A Lorentz transformation of the four-current density shows that $\boldsymbol{J}' = \boldsymbol{J} + \mathcal{O}((U/c)^2)$, i.e. the equation is correct up to second-order corrections in U/c.

[8] The vagueness of this statement is intended. If the assumption applies depends on the timescales of interest.

functions isotropic in the rest frames of the fluid elements, i.e. $\mathfrak{f}_s(t, \boldsymbol{x}, w)$ does not depend on the angles θ and φ. This has two implications: Firstly,

$$\pi_s^{ij}(t, \boldsymbol{x}) = m_s \int w^i w^j \mathfrak{f}_s(t, \boldsymbol{x}, w) w^2 \, \mathrm{d}w \, \mathrm{d}\Omega = 0 \quad \text{for } i \neq j \,,$$

i.e. the off-diagonal elements of the intrinsic pressure tensor are zero and, secondly, its diagonal elements are equal, namely $\pi_s^{11} = \pi_s^{22} = \pi_s^{33}$. This motivates to contract eq. (2.79), i.e.

$$\frac{\mathrm{D}}{\mathrm{D}t} \Pi^i{}_i + \frac{\partial}{\partial x^k} Q^{ki}{}_i - 2 J'^i E'_i + 2 \Pi^k{}_i \frac{\partial U^i}{\partial x^k} + \nabla_x \cdot \boldsymbol{U} \Pi^i{}_i = \frac{\delta}{\delta t} \pi^i_{e,i} + \frac{\delta}{\delta t} \pi^i_{p,i} \,,$$

where we exploited that the contraction of a symmetric tensor, namely π_s^{mi}, with an anti-symmetric tensor, ϵ_{imn}, gives zero. Dividing the equation by two and employing the definitions given in eq. (2.74), results in

$$\frac{\partial}{\partial t} \left(\frac{3}{2} P \right) + \nabla_x \cdot \left(\boldsymbol{Q} + \frac{3}{2} P \boldsymbol{U} \right) + P \nabla_x \cdot \boldsymbol{U} - \boldsymbol{J}' \cdot \boldsymbol{E}' = 0 \,. \tag{2.80}$$

We define in analogy to a monoatomic ideal gas the *thermal* or *internal energy* of the plasma to be $3/2P$. Hence, eq. (2.80) can be interpreted as an energy conservation equation. This also explains why we set the right-hand side to zero: the energy transfer between electrons and ions due to collisions must balance. (Shkarofsky et al., 1966, eq. 9-19)

We obtain an alternative form of the energy conservation equation by noting that eq. (2.80) is equivalent to

$$\frac{\mathrm{D}P}{\mathrm{D}t} + \frac{5}{3} P \nabla_x \cdot \boldsymbol{U} = -\frac{2}{3} \nabla_x \cdot \boldsymbol{Q} + \frac{2}{3} \boldsymbol{J}' \cdot \boldsymbol{E}' \,,$$

and by writing its left-hand side as

$$\frac{\mathrm{D}P}{\mathrm{D}t} - \frac{5}{3} \frac{P}{\rho} \frac{\mathrm{D}\rho}{\mathrm{D}t} = \frac{\mathrm{D}P}{\mathrm{D}t} + P \rho^{5/3} \frac{\mathrm{D}\rho^{-5/3}}{\mathrm{D}t} = \rho^{5/3} \frac{\mathrm{D}}{\mathrm{D}t} \left(P \rho^{-5/3} \right) \,,$$

where we used the mass conservation equation (2.75), see Shkarofsky et al. (1966, eq. 9-21b). If we now assume that there is no heat flow involved in changing the pressure and, moreover, that the conductivity of the plasma is infinite, i.e. $\boldsymbol{Q} = 0$ and $\boldsymbol{E}' = 0$ respectively, the energy conservation equation is turned into the *adiabatic equation of state*, that is

$$\frac{\mathrm{D}}{\mathrm{D}t} \left(P \rho^{-5/3} \right) = 0 \,. \tag{2.81}$$

2.2.4 IDEAL MHD EQUATIONS

The fluid equations are insufficient to determine the state of the plasma, because they contain electromagnetic fields that exist due to the plasma's motion. A closed system of equations is obtained with Maxwell's equations. The objective of this section is to demonstrate that it is possible to eliminate the $\boldsymbol{E}$-field from Maxwell's equation. The result are the magnetohydrodynamic (MHD) equations. The key step is to derive an expression of the $\boldsymbol{E}$-field in terms of the current density $\boldsymbol{J}$, the so-called *generalised Ohm's law*. Furthermore, we show that the assumption of an infinitely conductive plasma leads to a major simplification of the generalised Ohm's law and, consequently, of the MHD equations. The simplified equations are refereed to as ideal MHD equations.

In a first step towards the generalised Ohm's law, we multiply the zeroth velocity-moment equations of the protons and electrons, i.e. the $l = 0$ specialisation of eq. (2.73), with $\pm e/m_s$ and add them together. This yields

$$\frac{\partial}{\partial t}\left(en_p - en_e\right) + \nabla_x \cdot \boldsymbol{J}' = 0\,, \tag{2.82}$$

which is the *charge conservation equation*. Note that its right-hand side vanishes, because no particles are created or destroyed in collisions. Since we assume that the plasma is neutral the time derivative vanishes.

In a second step, we repeat the computation using the first velocity-moment equations, namely the momentum conservation equation of the protons and electrons. The result is

$$\begin{aligned}
&\frac{\mathrm{D}}{\mathrm{D}t}J'^i + e\frac{\partial}{\partial x^j}\left(\frac{\pi_p^{ji}}{m_p} - \frac{\pi_e^{ji}}{m_e}\right) - e^2\left(\frac{n_p}{m_p} + \frac{n_e}{m_e}\right)E'^i \\
&\quad - e^2\left(\frac{n_p}{m_p}u_p^m + \frac{n_e}{m_e}u_e^m\right)\epsilon^i{}_{mn}B^n + (en_p - en_e)\frac{\mathrm{D}U^i}{\mathrm{D}t} + J'^j\frac{\partial U^i}{\partial x^j} \\
&\quad + \nabla_x \cdot \boldsymbol{U}J'^i = \frac{e}{m_p}\frac{\delta(\boldsymbol{\pi}_p)^i}{\delta t} - \frac{e}{m_e}\frac{\delta(\boldsymbol{\pi}_e)^i}{\delta t}
\end{aligned}$$

For the case of a neutral plasma the above equation can be rearranged to the generalised Ohm's law, i.e.

$$\begin{aligned}
&\frac{\partial J'^i}{\partial t} + \nabla_x \cdot \left(\boldsymbol{U}J'^i + \boldsymbol{J}'U^i\right) + e\frac{\partial}{\partial x^i}\left(\frac{p_p}{m_p} - \frac{p_e}{m_e}\right) \\
&\quad - \frac{ne^2}{\mu_{pe}}E'^i - e\frac{m_e - m_p}{m_p m_e}\epsilon^i{}_{mn}J'^m B^n = \frac{e}{\mu_{pe}}(\Delta\boldsymbol{\pi}_{pe})^i\,,
\end{aligned} \tag{2.83}$$

where we used the conservation of charge equation (2.82), the definition of the reduced mass μ_{pe}, defined $n \coloneqq n_p = n_e$ and exploited that the momentum transfer between protons and electrons must balance each other, i.e. $\Delta\boldsymbol{\pi}_{pe} \coloneqq \delta\boldsymbol{\pi}_p/\delta t = -\,\delta\boldsymbol{\pi}_e/\delta t$. Moreover, the assumption that the particle distribution

functions of the protons and electrons are isotropic in the rest frames of the fluid elements implies that the ion and electron pressure are a scalar. We note that the form of the magnetic force term is due to the fact that

$$\begin{aligned}\frac{n_p}{m_p}\boldsymbol{u}_p + \frac{n_e}{m_e}\boldsymbol{u}_e &= \frac{1}{m_p m_e}\left(m_e n_p \boldsymbol{u}_p - \left(\boldsymbol{\pi}_p + \boldsymbol{\pi}_e\right) + m_p n_e \boldsymbol{u}_e\right) \\ &= \frac{m_e - m_p}{m_p m_e}\left(n_p \boldsymbol{u}_p - n_e \boldsymbol{u}_e\right),\end{aligned}$$

where we once more used that the sum of the intrinsic momenta vanishes. We note that a generalised Ohm's law for a non-neutral plasma is presented in Shkarofsky et al. (1966, p. 430–431).

The mass conservation equation (2.75), the momentum conservation equation (2.77) and the energy conservation equation (2.80) together with Maxwell's equation form a complete set of equations. In principle, the $\boldsymbol{E}$-field can be replaced with the current density via the generalised Ohm's law and, as stated in the introduction to this section, we refer to the resulting set of equations as *magnetohydrodynamic (MHD) equations* or, more compactly, as *magnetohydrodynamics*.

Further simplifications of this set are possible depending on the parameters of the plasma that should be modelled. In the rest of this text we adopt for the background plasma, in which the charged an energetic particles are accelerated, a simplification that bears the name *ideal MHD*. In the ideal MHD approximation it is assumed that the heat flow does not contribute to the energy change of the plasma and that the generalised Ohm's law reduces to

$$\boldsymbol{E}' = \boldsymbol{E} + \boldsymbol{U} \times \boldsymbol{B} = 0\,, \tag{2.84}$$

which describes an infinitely conducting plasma and implies that the energy conservation equation can be replaced with the adiabatic equation of state given in eq. (2.81). The discussion of the validity of the assumptions going into this simplifications is beyond the scope of this thesis. However, we quote the necessary conditions on the plasma parameters. To this end we once more transform the generalised Ohm's law for a neutral plasma given in eq. (2.83), namely we neglect terms that are multiplied with $m_e/m_p \ll 1$ and replace μ_{pe} with m_e. We also introduce a scattering frequency to parameterise the ion-electron collisions, i.e. $(e/m_e)(\Delta\boldsymbol{\pi}_{pe})^i := -\nu_{pe}\boldsymbol{J}'$. This results in

$$\begin{aligned}-\frac{1}{\nu_{pe}}\frac{\partial J'^i}{\partial t} - \frac{1}{\nu_{pe}}\nabla_x \cdot \left(\boldsymbol{U}J'^i + \boldsymbol{J}'U^i\right) \\ + \frac{e}{m_e\nu_{pe}}\frac{\partial}{\partial x^i}p_e + \frac{ne^2}{\nu_{pe}m_e}E'^i - \frac{e}{\nu_{pe}m_e}\epsilon^i{}_{mn}J'^m B^n = J'^i\,.\end{aligned} \tag{2.85}$$

We note that in steady-state and in the *static limit*, namely $\boldsymbol{B} = 0$ and $p_e = \text{const.}$, the plasma does not flow and the above equation reduces to the 'standard' Ohm's

law, usually defined in the rest frame of a conductor, i.e.

$$J'^i = \frac{ne^2}{\nu_{pe} m_e} E'^i \coloneqq \sigma E'^i \,.$$

where we defined the *conductivity* σ.[9]

The reduction of eq. (2.85) to the form used in ideal MHD, namely eq. (2.84), requires that all terms are much smaller than $\boldsymbol{E}'$. After diving eq. (2.85) by σ, the conditions are

$$\frac{m_e}{ne^2}\frac{\partial \boldsymbol{J}'}{\partial t} \text{ can be neglected if } \frac{L^2\omega_{p_e}^2}{c^2} \gg 1 \tag{2.86}$$

$$\nabla_x p_e \quad \text{can be neglected if } \frac{LU_0\omega_{ce}}{k_B T_e/m_e} \gg 1 \tag{2.87}$$

$$\frac{\boldsymbol{J}' \times \boldsymbol{B}}{ne} \text{ can be neglected if } \frac{L\omega_{p_e}^2}{\omega_{ce}}\frac{U_0}{c^2} \gg 1 \tag{2.88}$$

$$\frac{\boldsymbol{J}'}{\sigma} \quad \text{can be neglected if } R_M \coloneqq \frac{LU_0}{\sigma\mu_0} \gg 1\,, \tag{2.89}$$

where L is a characteristic length and U_0 is a characteristic velocity. $\omega_{p_e} \coloneqq (ne^2/m_e\epsilon_0)^{1/2}$ is the *electron plasma frequency*, $\omega_{ce} \coloneqq eB/m_e$ is the *electron cyclotron frequency* and R_M is the *magnetic Reynolds number*. The factor $k_B T/m_e = v_{e,\mathrm{rms}}^2/3$ is one third of the root mean square velocity of the electron distribution and can be considered to represent a typical velocity of an electron. (Krall & Trivelpiece, 1973, Sec. 3.6)

At the end of this chapter we use the ideal MHD approximation in conjunction with Maxwell's equation to derive an equation that evolves the $\boldsymbol{B}$-field and to express the current density $\boldsymbol{J}$, the $\boldsymbol{E}$-field and the charge density ρ_q as functions of $\boldsymbol{B}$. We use the magnetic field as the *primary variable* (Thorne & Blandford, 2017, Sec. 19.2). We derive an equation for the evolution of the $\boldsymbol{B}$-field by taking the curl of the reduced generalised Ohm's law, as presented in eq. (2.84), and we, subsequently, plug the result into the Maxwell equation $\nabla_x \times \boldsymbol{E} = -\partial \boldsymbol{B}/\partial t$. This yields

$$\frac{\partial \boldsymbol{B}}{\partial t} = \nabla_x \times (\mathbf{U} \times \boldsymbol{B}) \,. \tag{2.90}$$

This equation is called the *induction equation* for an infinitely conducting plasma, see Thorne and Blandford (2017, eq. 19.6).

For the electron-proton plasma studied, Ampère's law (2.37) is

$$\begin{aligned}
\nabla_x \times \boldsymbol{B}(t,\boldsymbol{x}) &= \mu_0 \sum_{s=1}^{2} q_s \int \frac{\boldsymbol{p}}{m_s} f_s(t,\boldsymbol{\xi})\, \mathrm{d}^3 p + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \\
&= \mu_0 \left(e n_p \bar{\boldsymbol{v}}_p - e n_e \bar{\boldsymbol{v}}_e\right) + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \\
&= \mu_0 \boldsymbol{J} + \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} \,,
\end{aligned}$$

[9] We note that in our laboratory frame the standard Ohm's law is $\boldsymbol{J} = \sigma\,(\boldsymbol{E} + \boldsymbol{U} \times \boldsymbol{B})$.

where we used the expression for the current density $\boldsymbol{J}$ in the laboratory frame given in eq. (2.78). The displacement current is of order $E/(c^2T) \sim UB/(c^2T) \sim (U^2/c^2)(B/L)$, where $T \sim L/U$ is the time scale on which the $\boldsymbol{E}$-field varies and L is the corresponding length scale. Moreover, the ideal MHD approximation implies that $E \sim UB$. Since the curl of $\mathbf{B}$ is of order B/L, we drop the displacement current in agreement with the non-relativistic approximation that we made (Thorne & Blandford, 2017, cf. eq. 19.5a). The current density thus is

$$\boldsymbol{J} = \frac{1}{\mu_0} \nabla_x \times \boldsymbol{B}(t, \boldsymbol{x}) \,. \tag{2.91}$$

Eventually, the charge density can be computed with Gauss' law (2.36) in conjunction with the reduced Ohm's law (2.84), i.e.

$$\rho_q = -\epsilon_0 \nabla_x \cdot (\boldsymbol{U} \times \mathbf{B}) \tag{2.92}$$

see Thorne and Blandford (2017, eq. 19.5c).

This completes the set of ideal MHD equations. We summarise them for later reference (cf. Shkarofsky et al., 1966, eq. 9-40 – 9-44):

(a) The mass conservation equation (2.75)

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} + \nabla_x \cdot \boldsymbol{U}\rho = \frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho \boldsymbol{U}) = 0 \,, \tag{2.93}$$

(b) the momentum conservation equation (2.77)

$$\rho \frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t} + \nabla_x P - \boldsymbol{J} \times \boldsymbol{B} = 0 \,, \tag{2.94}$$

(c) the adiabatic equation of state

$$\frac{\mathrm{D}}{\mathrm{D}t} \left(P \rho^{-5/3} \right) = 0 \,, \tag{2.95}$$

(d) and Maxwell's equations

$$\begin{aligned} \frac{\partial \boldsymbol{B}}{\partial t} &= \nabla_x \times (\mathbf{U} \times \boldsymbol{B}) \\ \boldsymbol{J} &= \frac{1}{\mu_0} \nabla_x \times \boldsymbol{B}(t, \boldsymbol{x}) \,. \end{aligned} \tag{2.96}$$

CHAPTER 3

PARTICLE TRANSPORT IN TENUOUS ASTROPHYSICAL PLASMAS

In the last chapter we derived different equations to model plasmas. Most importantly, the VFP equation and the ideal MHD equations. In this chapter, we apply them to model the transport, i.e. the propagation and acceleration, of energetic and charged particles in astrophysical environments.

The MHD quantities, e.g. the bulk (or fluid) velocity $\boldsymbol{U}$, are used to describe a plasma in which the energetic and charged particles are transported. We call this fully ionised and electrically neutral plasma the background plasma. We assume that the ion and electron distribution functions are isotropic, that its conductivity is infinite and that the plasma flows with velocities that are small compared to the speed of light. This ensures that the ideal MHD approximation is valid. In the context of this chapter, the single particle distribution function f is the average number density of the energetic and charged particles in phase space. The VFP equation is used to evolve f in the electromagnetic fields of the background plasma. The collision term $(\delta f/\delta t)_c$ models the interactions of the particles with excitations of the background $\boldsymbol{E}$- and $\boldsymbol{B}$-fields. These excitations may be plasma waves propagating in the background plasma and/or MHD turbulence.

In this chapter we pursue two different, yet related, goals: First, to develop a physical intuition for the transport process that is at the core of our research, namely the *Fermi acceleration* process, and, second to formally adapt the VFP equation to the scenario described in the previous paragraph. In particular, it is necessary to obtain an explicit expression for the collision term. The main reason that the Fermi acceleration process is the focus of our research is that the energy spectrum of Galactic cosmic rays and the electron energy spectrum of many non-thermal sources is an inverse power-law, i.e.

$$\frac{\mathrm{d}N}{\mathrm{d}E} \propto E^{-\alpha},$$

where the *spectral index* α ranges between 2 and 3. The fact that the observed spectral indices are about the same in very distinct sources identifies collisionless shock waves as candidate locations for particle acceleration, because these shock

waves are ubiquitous in astrophysical environments. A supernova remnant shock is a prototypical example. (Longair, 1994, p. 344)

The application of Fermi's arguments to a collisionless shock wave yields a power law in rough agreement with the observed spectral indices. Because the bulk of Galactic cosmic rays is expected to be accelerated at SNR shocks and because it is likely that the Fermi acceleration mechanism operates there, we need equations that model it accurately if our aim is to better model the energy range representing the maximum energies that particles can reach in our Galaxy, namely the energy range starting at the 'knee' of the cosmic-ray energy spectrum, see Fig. 1.1.

Hence, the first part of this chapter starts with a reconstruction of the original derivation of the Fermi acceleration mechanism as described in Fermi (1949). Subsequently, we apply Fermi's arguments to the acceleration of particles at a parallel shock, in doing so we use computations and arguments that we took from Kirk (1994) and Bell (1978).

In the second part of this chapter we adapt the fully relativistic VFP equation (2.64) such that it models the transport of relativistic particles in a non-relativistically flowing background plasma. Therefore, we follow Blandford and Eichler (1987) in embedding Fermi's ideas into the Fokker–Planck formalism, i.e. we derive an explicit expression for the collision operator that models the particle-wave interactions. The collision operator has a simple form in the rest frame of the waves that scatter the particles. This motivates the introduction of a mixed-coordinate system, namely a coordinate system in which the configuration space coordinates $\boldsymbol{x}$ are defined in an inertial laboratory frame and the momentum space coordinates in the non-inertial rest frame of the background plasma. The rest frame of the background plasma is an approximation to the rest frame of the waves. We proceed as Achterberg and Norman (2018) and use a Lorentz transformation to obtain the momentum variables in the rest frame of the background plasma. Because the plasma's rest frame is non-inertial, the VFP equation must contain a fictitious force term, cf. our computations done at the end of Sec. 2.2.1. In a next step, we take advantage of the assumption that the velocity of the background plasma is small compared to the speed of light and remove terms of the order $\mathcal{O}(U/c)$ from the VFP equation. This yields the main result of this chapter, namely the semi-relativistic VFP equation.

In the last part of this chapter, we apply the adapted VFP equation to model the acceleration of particles at a parallel shock. Under certain physical conditions the distribution of energetic and charged particles can be assumed to be *almost* isotropic. If these conditions are fulfilled it is possible to use the diffusion approximation. They lead to an equation dubbed the cosmic-ray transport equation, which is an advection-diffusion equation. We compute its steady-state solution as done in Drury (1983) and Kirk (1994).

3.1 THE FERMI ACCELERATION PROCESS

This section begins with a presentation of Fermi's original work on the origin of the cosmic radiation. It contains the essential arguments and computations that are nowadays used to explain the acceleration of particles in astrophysical plasmas. In particular, Fermi collects in a physically intuitive manner the necessary 'ingredients' to derive the power law energy spectrum of the cosmic rays. Though, the original physical scenario he envisioned failed to explain the later observed fact that the spectral indices are about the same in many different sources. We apply his arguments to the parallel shock scenario and show that this remedies the issue. Today, processes like the one originally described by Fermi are called Fermi acceleration processes of type II, whereas processes analogue to the acceleration at shock waves are referred to as type I processes.

3.1.1 FERMI'S THOUGHTS ON THE ORIGIN OF THE COSMIC RADIATION

Fermi (1949) argued in favour of the hypothesis that cosmic rays are accelerated in interstellar space. He assumed that interstellar space was filled with a very dilute, ionised and magnetised hydrogen gas that is sparsely interspersed with moving clouds. These clouds moved and thereby stirred the hydrogen gas. The kinetic energy of the gas' motion would be channelled into the magnetic field and cause the creation of randomly moving field 'irregularities'. Fermi now imagined that an energetic, charged particle collided with these irregularities, was reflected and thus underwent some kind of random walk.

We pursue Fermi in working out the physical consequences of this picture. If we assume that the irregularities can be considered as obstacles that are much heavier than the particles[1], then their motion and energy is unchanged by a collision with a particle. We compute the particle's energy change by transforming into the centre-of-mass frame, i.e. the rest frame of the irregularity. Since it is assumed that the particle is merely reflected, its momentum component perpendicular to the direction of motion of the magnetic field irregularity is unchanged in the collision while its parallel momentum component is flipped. Transforming back into the original reference frame yields the energy of the particle after the collision. Hence, we restrict our computation to the parallel momentum component.

Let $\boldsymbol{\Upsilon}$ be the velocity of the irregularity in units of c and $\Gamma = (1 - \Upsilon^2)^{-1/2}$ its Lorentz factor. Then the parallel four-momentum of a particle in the laboratory frame is $p^{\mu}_{\parallel} = (E/c, \boldsymbol{p}_{\parallel})$ where $\boldsymbol{p}_{\parallel} = (\boldsymbol{p} \cdot \boldsymbol{\Upsilon})\boldsymbol{\Upsilon}/\Upsilon^2$. We can express the particles' energy and parallel momentum in in the rest frame of the irregularity using a

[1]The assumption is well justified since the magnetic field irregularities should be in size comparable to multiples of the characteristic length scales of the MHD equations that govern the interstellar space plasma. This implies that in such an irregularity many particles are involved whose motion will not be drastically changed by a single particle interacting with them.

Lorentz transformation, i.e.

$$\begin{pmatrix} E'/c \\ \boldsymbol{p}'_\parallel \end{pmatrix} = \begin{pmatrix} \Gamma & -\Gamma\boldsymbol{\Upsilon}^\mathsf{T} \\ -\Gamma\boldsymbol{\Upsilon} & \Gamma\mathbf{1} \end{pmatrix} \begin{pmatrix} E/c \\ \boldsymbol{p}_\parallel \end{pmatrix} .$$

After the collision the particle's parallel four-momentum is $\bar{p}_\parallel^{\mu'} = (\bar{E}'/c, \bar{\boldsymbol{p}}'_\parallel) = (E'/c, -\boldsymbol{p}'_\parallel)$. Note the bar on top of the symbols representing the energy and the momentum of the particle after the collision. The energy of the particle did not change, because we assumed that the collision is elastic and the energy change of the irregularity can be neglected due to its relatively high mass. The particles' parallel four-momentum after the collision in the laboratory frame thus is[2]

$$\begin{pmatrix} \bar{E}/c \\ \bar{\boldsymbol{p}}_\parallel \end{pmatrix} = \begin{pmatrix} \Gamma & \Gamma\boldsymbol{\Upsilon}^\mathsf{T} \\ \Gamma\boldsymbol{\Upsilon} & \Gamma\mathbf{1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \Gamma & -\Gamma\boldsymbol{\Upsilon}^\mathsf{T} \\ -\Gamma\boldsymbol{\Upsilon} & \Gamma\mathbf{1} \end{pmatrix} \begin{pmatrix} E/c \\ \boldsymbol{p}_\parallel \end{pmatrix} .$$

The evaluation of the matrix product yields for the parallel momentum component

$$\bar{\boldsymbol{p}}_\parallel = \Gamma^2 \left(-\boldsymbol{p}_\parallel + 2\gamma m \boldsymbol{\Upsilon} c - \Upsilon^2 \boldsymbol{p}_\parallel \right)$$

and the particle's energy is

$$\frac{\bar{E}}{E} = \Gamma^2 \left(1 - 2\Upsilon\beta\cos\vartheta + \Upsilon^2 \right) ,$$

where we used that $\boldsymbol{\Upsilon}^\mathsf{T} \cdot \boldsymbol{p}_\parallel = \boldsymbol{p} \cdot \boldsymbol{\Upsilon} = \gamma m v \Upsilon \cos\vartheta = (E/c)\beta\Upsilon\cos\vartheta$ and ϑ is the angle between the direction of motion of the particle and the irregularities' direction of motion. Note that if $\cos\vartheta < 0$ the collision of the particle and the irregularity is head on, otherwise it is an overtaking collision. In agreement with this the ratio $\bar{E}/E$ is larger one for head-on collisions and smaller one for overtaking collisions; provided moderate values of Υ are assumed. If we now restrict ourselves to particles gyrating about a mean magnetic field along which the irregularities propagate (as Fermi seemingly did), we can express the ratio of energies as

$$\frac{\bar{E}}{E} = \frac{1 \pm 2\Upsilon\beta\cos\theta + \Upsilon^2}{1 - \Upsilon^2} ,$$

where θ is 'the angle of the inclination of the spiral'[3] and Υc still is the velocity of the 'perturbation'. The positive sign corresponds to the energy change in a head-on collision and the negative sign to the one of an overtaking collision. (Fermi, 1949, see eq. 13 and surrounding text).

[2] It is worth noting that

$$\begin{pmatrix} \Gamma & \Gamma\boldsymbol{\Upsilon}^\mathsf{T} \\ \Gamma\boldsymbol{\Upsilon} & \Gamma\mathbf{1} \end{pmatrix} \begin{pmatrix} \Gamma & -\Gamma\boldsymbol{\Upsilon}^\mathsf{T} \\ -\Gamma\boldsymbol{\Upsilon} & \Gamma\mathbf{1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \Gamma^2\left(1 - \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^\mathsf{T}\right) \end{pmatrix} \neq \mathbf{1} .$$

However, changing the sign of $\boldsymbol{\Upsilon}$ yields the correct inverse for the parallel component of $\boldsymbol{p}$, since $\Gamma^2\left(1 - \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^\mathsf{T}\right)\boldsymbol{p}_\parallel = \boldsymbol{p}_\parallel$.

[3] There is ambiguity in the definition of the inclination angle of the spiral. We note the above energy ratio is correct if the angle is restricted to values ranging between 0 and $\pi/2$, i.e.

After having derived the energy change in a single collision, Fermi demonstrates that particles whose spirals have an arbitrary inclination angle θ gain on average energy. To this end he derives the likelihood for head-on collisions: It is given by the rate of head-on encounters of particles and irregularities divided by the rate of all encounters. Hence, it is proportional to the relative velocity of the particles and the irregularities, i.e. $P_{\mathrm{HO}} = N(v\cos\theta + \Upsilon c)$, where N is the constant of proportionality. Analogously, the likelihood for overtaking collisions is $P_{\mathrm{OT}} = N(v\cos\theta - \Upsilon c)$. Since each collision is either head-on or overtaking, the sum of their respective probabilities has to equal one, i.e.

$$1 = P_{\mathrm{HO}} + P_{\mathrm{OT}} = N(v\cos\theta + \Upsilon c + v\cos\theta - \Upsilon c)\,.$$

Thus, the normalisation is $N = 1/(2v\cos\theta)$.

Fermi then computes the average energy gain up to terms of the order Υ^2. Concretely, he investigates the natural logarithm of the energy ratio, because

$$\ln\left(\frac{\bar{E}}{E}\right) = \ln\left(1 + \frac{\Delta E}{E}\right) = \frac{\Delta E}{E} + \mathcal{O}((\Delta E/E)^2)\,,$$

i.e. for small energy changes the natural logarithm approximates the relative energy change. The Taylor expansion of $\ln(\bar{E}/E)$ at $\Upsilon = 0$ is

$$\ln\left(\frac{\bar{E}}{E}\right) = \pm 2\Upsilon\beta\cos\theta + 2\left(1 - \beta^2\cos^2\theta\right)\Upsilon^2 + \mathcal{O}(\Upsilon^3)\,.$$

Its average is

$$\begin{aligned}\left\langle \ln\left(\frac{\bar{E}}{E}\right)\right\rangle_{\mathrm{Av}} \approx \left\langle\frac{\Delta E}{E}\right\rangle_{\mathrm{Av}} &= \left[2\Upsilon\beta\cos\theta + 2\left(1-\beta^2\cos^2\theta\right)\Upsilon^2\right]P_{\mathrm{HO}} \\ &\quad - \left[2\Upsilon\beta\cos\theta - 2\left(1-\beta^2\cos^2\theta\right)\Upsilon^2\right]P_{\mathrm{OT}} \\ &= 4\Upsilon^2 - 2\Upsilon^2\beta^2\cos^2\theta = 4\Upsilon^2\left(1 - \frac{\beta^2}{2}\cos^2\theta\right).\end{aligned} \tag{3.1}$$

Since $\theta \in [0, \pi/2)$, the relative energy change is largest for particles with small parallel velocity components. (Fermi, 1949, cf. eq. 14).

More importantly, if we assume that the collisions happen continuously and independently of each other, we expect the energies of the particles to grow exponentially. This is shown by Fermi (1949, Sec. 3) with the following back-of-the-envelope computation: Considering eq. (3.1), Fermi approximates

$\boldsymbol{B}_0$ is the mean magnetic field. If θ was defined to be in $[0, \pi]$, the absolute value of $\cos\theta$ should be used.

the average relative change of the particles' energies with $\epsilon = \langle \Delta E/E \rangle)_{\mathrm{Av}} \approx \Upsilon^2$ and starts the acceleration process with non-relativistically moving particles, i.e. the particles' energies are $E_0 = mc^2$. After n collisions their energies will be $E = (1+\epsilon)^n E_0$. This implies that

$$\ln(E/E_0) = n \ln(1+\epsilon) \approx n\epsilon$$

and hence

$$E = E_0 \exp(n\epsilon) = mc^2 \exp\left(n\Upsilon^2\right) .$$

If we denote with τ the time between collisions of particles and magnetic field irregularities, the energies of the particles at time t are

$$E(t) = mc^2 \exp(\Upsilon^2 t/\tau) .$$

The particles will not gain energy indefinitely. Fermi assumes that, e.g. a proton, may lose most of its energy in a nuclear collision. He denotes with T the mean time between these collisions. Under the assumption of a constant injection of particles into the acceleration process, we can compute their age distribution. Let Q be the injection rate. In steady state particle loss and injection balance and the total number of particles will be $\eta = QT$[4]. Moreover, the number of particles that have been injected during the infinitesimal time interval $\mathrm{d}t$ at $t_0 = 0$ and now have the age t is given by

$$N(t) = Q\,\mathrm{d}t \exp(-t/T) .$$

Since the particles are constantly injected the number of particles with age t is independent of the time t_0 at which a specific set of particles was injected. Dividing the number of particles with age t by the total number of particles η gives the probability distribution for the age of the particles, namely

$$P(t) = \frac{1}{T} \exp(-t/T)\,\mathrm{d}t \coloneqq p(t)\,\mathrm{d}t ,$$

where $P(t)$ is the probability to find a particle with age t and $p(t)$ is the corresponding probability density function.[5]

Knowing what the energy of a non-relativistic particle after time t is and knowing the age distribution of the particles in the steady state, it is possible to derive the probability distribution of the energies. To this end we express the age of a particle as a function of its energy, i.e.

$$t = \frac{\tau}{\Upsilon^2} \ln\left(\frac{E}{mc^2}\right) .$$

This implies that $\mathrm{d}t = \tau/\Upsilon^2\,\mathrm{d}E/E$. We note that the probability to find a particle with age t, i.e. $p(t)\,\mathrm{d}t = p(t(E))\,\mathrm{d}E$, is equal to the probability to find a particle

[4] This can be derived formally by solving $\mathrm{d}N/\mathrm{d}t = Q - N/T$ and evaluating $N(t \to \infty) =: \eta$.

[5] I thank Dr. Keno Riechers who discussed with me Fermi's derivation of the age distribution.

with energy $E(t)$. We thus define the probability density function of the particles' energies to be $\pi(E) \coloneqq p(t(E))$ and the probability that a particle has the energy in the range $E + \mathrm{d}E$ is

$$\begin{aligned}\pi(E)\,\mathrm{d}E &= \frac{\tau}{T\Upsilon^2}\exp\left(-\frac{\tau}{T\Upsilon^2}\ln\left(\frac{E}{mc^2}\right)\right)\frac{\mathrm{d}E}{E}\\ &= \frac{\tau}{T\Upsilon^2}(mc^2)^{\tau/(TB^2)}E^{-(1+\tau/(TY^2))}\,\mathrm{d}E\,.\end{aligned}$$

The form of the energy spectrum corresponding to such a probability distribution is an inverse power law as observed for cosmic rays arriving at earth. (Fermi, 1949, Sec. 4)

At the end of our summary of Fermi's original arguments, we remark that he derived the inverse power-law probability distribution of the particles' energies with a concrete physical situation in mind. The spectral index of the probability distribution thus depends on the parameters τ, T and Υ describing this situation. However, the observed spectral indices of many non-thermal astrophysical sources are about the same and, hence, they should not depend on the parameters of a specific physical scenario.

We thus abstract from the physical situation and identify the general features of the Fermi acceleration process: Firstly, particles are constantly injected into the acceleration process. Secondly, they are accelerated due to collisions with electromagnetic fluctuations sustained by a plasma through which the particles move. Thirdly, the energy change in such a collision is proportional to the particles' original energy and, fourth, the particles are not accelerated indefinitely, they leave the system or they lose all their energy in a collision with another particle. This escape process is independent of the particles' energies.

The third feature, i.e. the fact that the relative average energy gain is proportional to the energy of the particle, led to an exponential increase of a particle's energy, i.e. $E(t) = E_0\exp(\epsilon t/\tau)$. Taking the derivative with respect to t yields

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\epsilon}{\tau}E =: \frac{E}{\tau_{\mathrm{acc}}}\,,$$

where we defined the mean acceleration time. The fourth feature, namely the energy independent escape, was modelled using an exponential decay of the particles, i.e.

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\frac{N}{\tau_{\mathrm{esc}}}\,,$$

where τ_{esc}^{-1} is the probability to escape (or to lose all energy) per unit time. The number of particles with energy E at time t is $N(t,E)$. Since the particles are accelerated and decay, $N(t,E)$ evolves in time as

$$\frac{\mathrm{d}}{\mathrm{d}t}N(t,E) = \frac{\partial N}{\partial t} + \frac{\mathrm{d}E}{\mathrm{d}t}\frac{\partial N}{\partial E} = -\frac{N}{\tau_{\mathrm{esc}}}\,.$$

In steady state the above equation is an ordinary differential equation (ODE), namely

$$\frac{E}{\tau_{\mathrm{acc}}}\frac{\mathrm{d}N}{\mathrm{d}E} = -\frac{N}{\tau_{\mathrm{esc}}}\,.$$

Its solution is $N(E) = N(E_0)(E/E_0)^{-\tau_{\mathrm{acc}}/\tau_{\mathrm{esc}}}$. The derivative with respect E gives

$$\frac{\mathrm{d}N}{\mathrm{d}E} = -\frac{\tau_{\mathrm{acc}}}{\tau_{\mathrm{esc}}}N(E_0)E_0^{\tau_{\mathrm{acc}}/\tau_{\mathrm{esc}}}E^{-(1+\tau_{\mathrm{acc}}/\tau_{\mathrm{esc}})}\,, \tag{3.2}$$

which reproduces the probability distribution of the particles' energies. Hence, in general, the Fermi acceleration process can be parameterised by two parameters, namely the mean acceleration time and an the mean escape time. The ratio of these two parameters determines the spectral index of the particles' energy spectrum.

3.1.2 PARTICLE ACCELERATION AT PARALLEL SHOCKS

The fact that the observed spectral indices are about the same in many non-thermal sources requires the identification of acceleration sites at which the ratio of the acceleration time and the escape time is independent of, for example, the size of the source and others of its characteristics. It is plausible that most sources are surrounded by a plasma in which strong shocks are propagating. (cf. Longair, 1994, p. 355) Because of their ubiquity they are good candidate particle accelerators. In this section we will apply the Fermi acceleration process to strong shock waves yielding the desired inverse power law and evaluate its spectral index. In particular, the the ratio of acceleration time to the escape time will only depend on a single parameter characterising shock waves, namely the compression ratio.

In a first step we adopt a simple shock model. We presuppose that the particles that participate in the acceleration process do *not* contribute to the physics creating and sustaining the shock wave. Their large energies imply mean free paths that are large in comparison to the shock width, hence the charged particles merely experience the shock wave as a discontinuity in the fluid quantities of the background plasma. Thus, mathematically we model the shock using a discontinuous velocity profile $\boldsymbol{U}$ describing the flow of the background plasma and, if necessary, a discontinuous background magnetic field $\boldsymbol{B}$.

The relative orientation of $\boldsymbol{U}$ and $\mathbf{B}$ led to a taxonomy of astrophysical shocks dividing them into the categories *parallel shocks* and *oblique shocks*. For the sake of completeness, we mention that the latter is subdivided into *subluminal* and *superluminal* shocks. A parallel shock is characterised by the fact that its direction of propagation aligns with the orientation of the $\mathbf{B}$-field, i.e. $\boldsymbol{U} \parallel \boldsymbol{B}$. Oblique shocks are all shocks that are not parallel. The special case $\boldsymbol{U} \perp \boldsymbol{B}$ is called *perpendicular shock*. (Kirk, 1994, see Sec. 2.2) A parallel shock is special

in that the **B**-field is continuous, as we shall show later, across the shock and charged particles gyrating about it can cross the shock wave undisturbed. The latter is the reason why we investigate the acceleration of particles at parallel shocks.

If the shock wave is not interacting with the particles, the situation is similar to the scenario Fermi described when explicating his acceleration process. In Fermi's original account moving clouds produced magnetic irregularities that scattered the particles. In the case of a shock wave, the plasma behind the shock, i.e. the plasma in the *downstream*, can be assumed to be highly turbulent and following the shock wave with a fraction of its speed. (Bell, 1978, see Sec. 1) The charged particles interact with the MHD turbulence and, thus, are reflected by moving scatterers and may overtake the shock wave and (re-)enter the *upstream*. Bell (1978, Sec. 3) proposed that the charged particles streaming through the upstream excite MHD waves that once more act as scattering centres forcing all particles to repeatedly encounter the moving MHD turbulence of the downstream. Pictorially speaking, the particles are thus constantly reflected by moving scattering centres as they were in Fermi's initial description of the process. However, this time the electromagnetic field irregularities do not propagate into random directions, instead they move with the downstream plasma and hence follow the shock wave. This directed motion implies that the particles only experience head-on collisions. As we will show in the next paragraphs, this leads to an average energy change proportional to the velocity of the scatterers and not to their velocity squared.

We firstly investigate the discontinuities of the velocity and the magnetic field at a parallel shock. To this end we choose a reference frame in which the shock is at rest, called the *shock rest frame*. In this frame the upstream plasma is approaching the shock, located at $x = 0$, with the velocity of the shock wave, see Fig. 3.1.2. We assume that this velocity is known and that the shock wave is steady, i.e. all fluid quantities are time independent. It is then possible to relate the plasma velocity component parallel to the shock normal $\boldsymbol{n}$ to its downstream counterpart via an integration of the steady-state conservation of mass equation (2.75). We integrate it over a thin rectangular cuboid centred at the shock wave, e.g. $V = [-\varepsilon, \varepsilon] \times [-L/2, L/2] \times [-L/2, L/2]$, where L is arbitrary. The integration yields

$$\begin{aligned}\lim_{\varepsilon\to 0}\int_V \nabla_x \cdot (\rho \boldsymbol{U})\,\mathrm{d}^3x &= \lim_{\varepsilon\to 0}\int_{\partial V} \rho \boldsymbol{U}\cdot \mathrm{d}\boldsymbol{S}\\ &= L^2 \lim_{\varepsilon\to 0}\left[\rho(x+\varepsilon)U_x(x+\varepsilon) - \rho(x-\varepsilon)U_x(x-\varepsilon)\right]\\ &= L^2\left[\rho(x^+)U_x(x^+) - \rho(x^-)U_x(x^-)\right] = 0\,.\end{aligned}$$

We remark that x^- is in the upstream and x^+ in the downstream respectively and that the integrals over the lateral surfaces vanished because their surface area goes to zero in the limit $\varepsilon \to 0$. The above equation simply states that if mass is conserved the mass flux must be continuous across surfaces. It does not

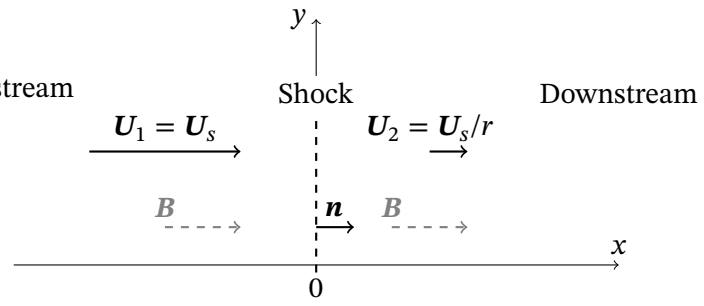

Figure 3.1: A parallel shock in its rest frame. $\boldsymbol{U}_s$ is the velocity of the shock wave, $\boldsymbol{B}$ is the mean magnetic field of the background plasma, $\boldsymbol{n}$ the shock normal and r is the compression ratio of the shock.

matter that the surface is a shock. If we use the subscripts 1 and 2 to denote upstream and downstream quantities, the above equation is equivalent to

$$U_2 = \frac{\rho_1}{\rho_2} U_1 =: \frac{U_1}{r}\,, \tag{3.3}$$

where we defined the *compression ratio r* of the shock. A shock that moves with a velocity U_s much larger than the speed of sound of the upstream c_s, i.e. its *Mach number* $M = U_s/c_s \gg 1$, is called a *strong shock*. The compression ratio of a strong shock is $r = 4$, see, for example, Blandford and Eichler (1987, Sec. 4.7).

Since $\boldsymbol{B}$ is divergence free, an analogue computation shows that B_x does not change across the shock, i.e.

$$\lim_{\epsilon\to 0} \int_V \boldsymbol{B}\cdot \mathrm{d}\boldsymbol{S} = L^2 \lim_{\epsilon\to 0} [B_x(x+\epsilon) - B_x(x-\epsilon)] = 0\,.$$

Again using the subscripts 1 and 2, the above equation reduces to

$$B_1 = B_2\,. \tag{3.4}$$

Equations (3.3) and (3.4) are part of the *MHD Rankine–Hugoniot conditions* that, in general, relate all upstream plasma quantities to their downstream counterparts assuming that the upstream quantities and the ratio of *specific heats* of the plasma is known. The complete set of equations and their solution can be found in Decker (1988, Sec. 2.2) and references therein. For the case of a parallel shock these conditions furthermore imply that the velocity components tangential to the shock plane, namely U_y and U_z, do not change across the shock. The same is true for the tangential magnetic field components. (Decker, 1988, eq. 12 and 13) Hence, in the downstream of a parallel shock the $\boldsymbol{B}$-field and the velocity $\boldsymbol{U}$ stay parallel, see Fig. (3.1.2).

Knowing how the velocity changes across the shock front, we can compute the energy change of a particle that starts in the upstream, enters the downstream

and on being scattered returns upstream. We compute its energy in the rest frame of the upstream plasma, i.e. the downstream plasma is flowing with velocity $\boldsymbol{\Upsilon} = (U_2 - U_1)/c\boldsymbol{e}_x = U_1/c(1-r)/r\boldsymbol{e}_x$.[6] The computation follows the steps presented in Sec. (3.1): We transform into the rest frame of the scattering centres, i.e. into the rest frame of the downstream plasma. The scattering event takes place and we subsequently transform back into the rest frame of the upstream. However, this time the particles are scattered into all directions and they are not simply reflected as before.

The particle's energy in rest frame of the downstream plasma is

$$\frac{E'}{c} = \Gamma\left(\frac{E}{c} - \boldsymbol{\Upsilon}\cdot\boldsymbol{p}\right) = \Gamma\frac{E}{c}\left(1 - \Upsilon\beta\cos\vartheta\right) .$$

The scattering is assumed to be elastic and once more the scattering centre's energy is unchanged in the event. Hence, the particle's energy before and after it is the same, i.e. $\bar{E}' = E'$. The momentum vector of the particle after the scattering event is denoted with $\bar{\boldsymbol{p}}'$ and, since the particle's energy did not change, $\bar{\boldsymbol{p}}'$ is related to $\boldsymbol{p}'$ by a rotation. We transform back into the rest frame of the upstream plasma. This yields

$$\frac{\bar{E}}{c} = \Gamma\left(\frac{\bar{E}'}{c} + \boldsymbol{\Upsilon}\cdot\bar{\boldsymbol{p}}'\right) = \Gamma\frac{E'}{c}\left(1 + \Upsilon\beta'\cos\bar{\vartheta}'\right) ,$$

where we used that the magnitude of the particles velocity is unchanged in the scattering event, namely $\bar{\beta}' = \beta'$. Plugging our expression for E' into the above formula gives an expression for the energy of the particle in the upstream rest frame after being scattered, i.e.

$$\begin{aligned}\bar{E} &= \Gamma^2 E\left(1 + \Upsilon\beta'\cos\bar{\vartheta}'\right)\left(1 - \Upsilon\beta\cos\vartheta\right)\\ &= \Gamma^2 E\left(1 - \Upsilon\beta'\cos\bar{\theta}'\right)\left(1 + \Upsilon\beta\cos\theta\right)\\ &= \Gamma^2 E\left[1 - \frac{U_1 - U_2}{c}\beta'\cos\bar{\theta}'\right]\left[1 + \frac{U_1 - U_2}{c}\beta\cos\theta\right] .\end{aligned} \tag{3.5}$$

Cf. Gaisser et al. (2016, Sec. 12.2.2). Notice that the angle ϑ is the angle between the particle's momentum vector $\boldsymbol{p}$ and the velocity $\boldsymbol{\Upsilon}$. We prefer to use the angle between the particle's momentum and $\boldsymbol{e}_x$, i.e. the x-axis, and denote it with the symbol θ. Since $\boldsymbol{\Upsilon}$ is the velocity of the downstream plasma in the upstream rest frame, it points in the negative x-direction and hence the relation between the two angles is $\vartheta = \pi - \theta$, whence the sign change in the above equation. The angle θ is referred to as the *pitch angle*, because the **B**-field is aligned with the x-axis and, thus, θ determines the pitch of the spiral traced out by the particles' trajectories.

In a next step, we compute the *average* energy gain of a particle performing one cycle, i.e. of a particle that leaves the upstream, is scattered in the downstream and returns to the upstream. If the velocity of the particle β is fixed,

[6] We neglected terms of order $\mathcal{O}(U_1U_2/c^2)$. The application of the velocity addition formula yields $\boldsymbol{\Upsilon} = (U_2 - U_1)/(1 - U_1U_2/c^2)\boldsymbol{e}_x$.

then the energy $\bar{E}$ can be considered a random variable of the two angles θ and $\bar{\theta}'$ whose joint probability density function, say $p(\theta, \bar{\theta}')$, can be written as the product of the probability density function of θ and the probability density function of $\bar{\theta}'$ that we denote with $\chi(\theta)$ and $\psi(\bar{\theta}')$ respectively. The reason is that the joint probability density function can be factorized is that the particle is scattered enough times to 'forget' about its original direction of motion before returning to the upstream. Hence, the angle θ and $\bar{\theta}'$ are *statistically independent*. Thinking of a phase-space distribution of particles and not of a single particle, we express this idea by saying that the particle distribution function is 'isotropised', i.e. it becomes isotropic on a time scale that is short in comparison to the time scale of the acceleration process.

First, we derive an expression for $\chi(\theta)$. The probability density function $\chi(\theta)$ is defined by the number of particles in the upstream that have a specific pitch angle $\theta \in [0, \pi/2)$ divided by the total number of particles that cross the shock. Note only particles with θ in $[0, \pi/2)$ do cross the shock. We emphasise that θ is defined in the rest frame of the upstream plasma and that the phase-space distribution of the particles in this frame is isotropic, because of Alfvén waves that scatter them and that are created by their streaming through the upstream, see Bell (1978, Sec. 1). Thus the total number of particles crossing the shock per unit time, surface area and momentum interval is

$$\begin{aligned}\frac{\mathrm{d}j_S^x}{\mathrm{d}p}(p) &= \frac{\mathrm{d}N}{\mathrm{d}t\,\mathrm{d}A\,\mathrm{d}p} \\ &= \int_0^{2\pi}\int_0^{\pi/2+\arcsin(U_1/v)} f(p)p^2\left(c\frac{p^x}{p^0}+U_1\right)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \\ &= 2\pi f(p)p^2\int_{-U_1/v}^{1}(\mu v+U_1)\,\mathrm{d}\mu = \pi f(p)p^2 v\left(1+\frac{U_1}{v}\right)^2,\end{aligned} \tag{3.6}$$

where we employed the definition of the *number-flux-4-vector* as can, for example, be found in Thorne and Blandford (2017, cf. eq. 3.35c). Moreover, $p^0 = \gamma mc$ is the zeroth component of the particles' four-momenta and $f(p)$ is their *upstream* phase-space distribution function at the shock. Notice that the shock is moving with U_1 in the negative x-direction and thus its velocity had to be included in the computation of the flux through the shock surface. This also explains the appearance of $\arcsin(U_1/v)$ in the upper limit of the θ integration; a particle whose velocity in the negative x-direction is smaller than the shock's velocity is overtaken by the shock wave. Moreover, in the third line we changed the variable of integration to $\mu \coloneqq \cos\theta$. The number of particles that have a *specific* pitch angle and momentum when crossing the shock is

$$\begin{aligned}\frac{\mathrm{d}j_S^x}{\mathrm{d}p}(p,\theta) &= \int_0^{2\pi} f(p)p^2\left(c\frac{p^x}{p^0}+U_1\right)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \\ &= 2\pi f(p)p^2 v\left(\cos\theta+\frac{U_1}{v}\right)\sin\theta\,\mathrm{d}\theta\,.\end{aligned}$$

Hence, the probability for a particle with pitch angle θ and momentum p (or velocity v) to cross the shock is

$$\begin{aligned} P(\theta) &= \frac{2\pi f(p) p^2 v \,(\cos\theta + U_1/v) \sin\theta \,\mathrm{d}\theta}{\pi f(p) p^2 v \,(1 + U_1/v)^2} \\ &= \frac{2\,(\cos\theta + U_1/v)\sin\theta}{(1 + U_1/v)^2}\,\mathrm{d}\theta =: \chi(\theta)\,\mathrm{d}\theta\,. \end{aligned} \tag{3.7}$$

Notice that the derivation of $\chi(\theta)$ implicitly restricts its domain to $[0, \pi/2 + \arcsin(U_1/v))$.

Secondly, we derive the probability density function $\psi(\bar{\theta}')$. We remind ourselves that the prime indicates that $\bar{\theta}'$ is defined in the rest frame of the *downstream* plasma. The probability that a particle with a specific pitch angle $\bar{\theta}'$ crosses the shock is thus determined by the phase-space distribution function of the particles as seen in that same reference frame. The phase-space distribution function in the downstream rest frame is isotropic, because the particles are scattered by the MHD turbulence of the shocked plasma. The bar that adorns the angle $\bar{\theta}'$ tells us that we are interested in particles that were scattered; only particles that changed their pitch angle to be in the range $(\pi/2 + \arcsin(U_2/v'), \pi]$ are able to return upstream. Note that the shock in the downstream rest frame propagates along the negative x-direction with velocity U_2, whence the $\arcsin(U_2/v')$. This implies that the total number of particles crossing the shock wave from downstream to upstream can as well be computed with an adapted version of the formula given in eq. (3.6), i.e.

$$\begin{aligned} \frac{\mathrm{d}j_S'^x}{\mathrm{d}p'}(p') &= 2\pi f'(p') p'^2 \int_{-1}^{-U_2/v'} (v'\bar{\mu}' + U_2)\,\mathrm{d}\bar{\mu}' \\ &= -\pi f(p') p'^2 v' \left(1 - \frac{U_2}{v'}\right)^2 . \end{aligned}$$

The minus sign reflects that the particles are propagating in the negative x-direction. In this context, probabilities are ratios of particle numbers and, hence, we use the absolute value of the above number flux current density to compute the probability density function $\psi(\bar{\theta}')$. Proceeding as in eq. (3.7), we arrive at

$$\psi(\bar{\theta}') = \frac{2|(\cos\bar{\theta}' + U_2/v')\sin\bar{\theta}'|}{(1 - U_2/v')^2} = -\frac{2(\cos\bar{\theta}' + U_2/v')\sin\bar{\theta}'}{(1 - U_2/v')^2}\,. \tag{3.8}$$

Note that it was possible to replace the absolute value in the numerator with a minus sign, because the domain of ψ is $(\pi/2 + \arcsin(U_2/v'), \pi]$. (Kirk, 1994, Sec. 5.1 and, in particular, eq. 95 for the probability density functions)

Eventually, we compute the average relative energy gain $\epsilon = \langle \ln(\bar{E}/E) \rangle \approx$

$\langle \Delta E/E \rangle$, i.e. [7]

$$\begin{aligned}\left\langle \ln\left(\frac{\bar{E}}{E}\right)\right\rangle &= \int_{-U_1/v}^{1}\int_{-1}^{-U_2/v'} \ln\left(\frac{\bar{E}}{E}\right)\chi(\mu)\psi(\bar{\mu}')\,\mathrm{d}\mu\,\mathrm{d}\bar{\mu}' \\ &= \ln\Gamma^2 + \int_{-1}^{-U_2/v'} \ln\left(1-\frac{U_1-U_2}{c}\beta'\bar{\mu}'\right)\psi(\bar{\mu}')\,\mathrm{d}\bar{\mu}' \\ &\quad + \int_{-U_1/v}^{1} \ln\left(1+\frac{U_1-U_2}{c}\beta\mu\right)\chi(\mu)\,\mathrm{d}\mu \\ &= \frac{4}{3}\frac{U_1-U_2}{c}\beta + \mathcal{O}\left((U_1/c)^2\right) \\ &= \frac{4}{3}\frac{U_1-U_2}{c} + \mathcal{O}\left((U_1/c)^2\right) + \mathcal{O}(\gamma^{-2})\,.\end{aligned} \tag{3.9}$$

Notice that the velocity addition formula can be applied to show that $\beta' = \beta + \mathcal{O}(U_1/c)$. Moreover, $\beta = 1+\mathcal{O}(\gamma^{-2})$ and, hence, for highly relativistic particles it is almost one. We also remark that $\mathcal{O}((U_1-U_2)/c) = \mathcal{O}(U_1/c) = \mathcal{O}(U_2/c)$, since $U_1 = rU_2$.

At this point we highlight that, as announced in the introduction to this section, the average relative energy gain is proportional to the velocity of the scatterers and not their velocity squared, cf. eq. (3.1). This led to the distinction between *Fermi I* and *Fermi II acceleration* processes that refer to the former and latter respectively.

In Sec. 3.1, we computed the energy of the particles after n encounters with magnetic irregularities assuming that their energy increased each time by the relative *average* energy gain ϵ, i.e. $E(n) = (1+\epsilon)^n E_0$. Bell (1978) uses the *central limit theorem* to show that $\ln(E(n)/E_0)$ is normal distributed with mean $n\epsilon$ and variance $n\sigma^2$, where σ is the standard deviation of $\ln(\bar{E}/E)$ and we remind ourselves that ϵ is the mean of $\ln(\bar{E}/E)$. Taking $\ln(E(n)/E_0)$ as a proxy for the change of the particles' energies, this means that almost all particles have the energy $E(n)$ after n cycles, because the normal distribution will be strongly peaked if n is large; note that the ratio of the variance to the mean squared is $n\sigma^2/n^2\epsilon^2 = 1/n(\sigma/\epsilon)^2$. This together with the fact that the mean of $\ln(E(n)/E_0)$ is $n\epsilon$ justifies the usage of the relative average energy gain to update the particle's energy, because this leads to $\ln(E(n)/E_0) = n\ln(\epsilon+1) = n\epsilon + \mathcal{O}(\epsilon^2)$, which is the expected value of the sharply peaked normal distribution.

We reconstruct Bell's argument in more detail. We begin with explicitly stating the central limit theorem: Suppose that X_n is a sequence of *independent and identically distributed* random variables with mean $\mu = \langle X_k \rangle$ and finite

[7] We remark that expressing the probability density functions in terms of μ allows us to drop $\sin\theta$ in eq. (3.7), because

$$\chi(\theta)\,\mathrm{d}\theta = \chi(\mu)\frac{\mathrm{d}\theta}{\mathrm{d}\mu}\,\mathrm{d}\mu = -\frac{\chi(\mu)}{\sin\theta}\,\mathrm{d}\mu\,.$$

positive variance $\sigma^2 = \langle X_k^2 \rangle - \langle X_k \rangle^2$. If $S_n = X_1 + \cdots + X_n$, then as n approaches infinity, the random variables $(S_n - n\mu)/(\sqrt{n}\sigma)$ *converge in distribution* to the normal distribution $\mathcal{N}(0,1)$. (Billingsley, 1995, Theorem 27.1)

Leaving aside the technical details of the convergence concept, the theorem essentially states that the probability density function of S_n is becoming a Gaussian distribution with mean $n\mu$ and variance $n\sigma^2$ as $n \to \infty$.

We now apply the theorem, we set

$$\begin{aligned} S_n &= \ln\left(\frac{E(n)}{E_0}\right) \\ &= \ln\left(\Gamma^{2n} \prod_{k=0}^{n-1} \left[1 - \frac{U_1 - U_2}{c}\beta' \cos\bar{\theta}'_k\right]\left[1 + \frac{U_1 - U_2}{c}\beta \cos\theta_k\right]\right) \\ &= n \ln \Gamma^2 + \sum_{k=0}^{n-1} \ln\left(\left[1 - \frac{U_1 - U_2}{c}\beta' \cos\bar{\theta}'_k\right]\left[1 + \frac{U_1 - U_2}{c}\beta \cos\theta_k\right]\right), \end{aligned}$$

where we repeatedly used the formula for the energy ratio $\bar{E}/E$ derived in eq. (3.5) and introduced the index k to number the cycles, i.e. going from upstream to downstream and back, performed by the particle. Hence, the elements of the sequence of random variables are

$$X_k = \ln\left(\left[1 - \frac{U_1 - U_2}{c}\beta' \cos\bar{\theta}'_k\right]\left[1 + \frac{U_1 - U_2}{c}\beta \cos\theta_k\right]\right),$$

i.e. the relative changes in energy. Because of the scattering of the particles in the upstream and the downstream, each cycle is independent of the previous one and the probability density functions of the relative energy changes are always given by $p(\theta_k, \bar{\theta}'_k) = \chi(\theta_k)\psi(\bar{\theta}'_k)$, see eqs. (3.7) and (3.8). This implies that the random variables in the sequence $(X_k)_{k=1,\ldots,n}$ are independent and identically distributed and that the mean of each element is $\langle X_k \rangle = \epsilon$. Additionally, we make the assumption that the variance of X_k is finite and positive.

This means that the central limit theorem applies and that the random variable $\ln(E(n)/E_0)$ is normal distributed, i.e. $\ln(E(n)/E_0) \sim \mathcal{N}(n\epsilon, n\sigma^2)$, if $n \to \infty$. This in turn justifies to work with the average relative energy change ϵ, because the normal distribution will be sharply peaked.

After having investigated the average relative energy change of a particle that crosses and recrosses the shock wave, we turn our attention to the next ingredient in the Fermi acceleration process, namely the escape of the particles. Unlike the original escape, the particles leave the acceleration process due to being advected downstream and not because they lose their energy in a collision with an ion of the background plasma (Bell, 1978, Sec. 2).

We remark that our computation of the escape probability follows Kirk (1994, Sec. 5.1). Since the particles are scattered in the downstream, it gets increasingly difficult for a particle to return to the shock wave the further downstream it got.[8]

[8] A mathematical formulation of this statement can be found in Drury (1983, in particular eq. 3.42 and the surrounding text).

At some distance L from the shock the return probability is quasi zero. We thus think up a box of length L bounded by the shock wave and an imaginary surface that moves together with the shock. We note that the box may have a different size for each momentum range $p + \mathrm{d}p$, because the distance L will depend on the energy of the particles if the scattering process is energy dependent. The ratio of the number of particles that leave the box to the number of particles that enter the downstream is the escape probability.

We investigate the escape of the particles in the downstream rest frame. In this reference frame, the distribution function of the energetic and charged particles is isotropic. Notice that up to now, we only worked with the distribution function at the shock. Since we are interested in the particles that leave the box, we need to know the values of the distribution function at the imaginary surface. To this end we assume that the distribution function is homogeneous in the downstream, i.e. for an arbitrary but fix momentum p, $f(x, p) = \text{const.}$ for $x > 0$. We show later that this assumption is correct, see eq. (3.74). The total number of particles that cross the imaginary surface per unit time, surface area and momentum interval is

$$\begin{aligned}\frac{\mathrm{d}j_I'^x}{\mathrm{d}p'}(p') &= \int_0^{2\pi}\int_{-1}^{1}\left(c\frac{p'^x}{p^0}+U_2\right)f'(p')p'^2\,\mathrm{d}\mu'\,\mathrm{d}\phi' \\ &= 2\pi f'(p')p'^2\int_{-1}^{1}(\mu' v'+U_2)\,\mathrm{d}\mu' = 4\pi f'(p')p'^2U_2\,.\end{aligned}$$

Notice that particles are leaving and entering the box, though there is a net flux that is determined by the velocity of the imaginary surface. The number of particles that *enters* the downstream is

$$\begin{aligned}\frac{\mathrm{d}j_s'^x}{\mathrm{d}p'}(p') &= \int_0^{2\pi}\int_{-U_2/v}^{1}\left(c\frac{p'^x}{p^0}+U_2\right)f'(p')p'^2\,\mathrm{d}\mu'\,\mathrm{d}\phi' \\ &= \pi f(p')p'^2v'\left(1+\frac{U_2}{v'}\right)^2\,.\end{aligned}$$

We again account for the fact that the shock is moving, namely with velocity U_2 in downstream rest frame, and that only particles whose pitch angle is in the range $[0, \pi/2 + \arcsin(U_2/v')]$ enter the downstream, cf. eq. (3.6).

The escape probability is

$$P_{\mathrm{esc}} = \frac{\mathrm{d}j_I'^x/\,\mathrm{d}p'}{\mathrm{d}j_s'^x/\,\mathrm{d}p'} = \frac{4U_2}{v'\left(1+\frac{U_2}{v'}\right)^2} = 4\frac{U_2}{v'} + \mathcal{O}((U_2/v)^2)\,. \tag{3.10}$$

At the end of Sec. (3.1), we pointed out that Fermi assumed an energy independent escape probability. We emphasise that the escape probability P_{esc} is energy dependent. However, for relativistic particles $v \approx c$ it is energy independent. We mention that we computed the relative average energy gain in the upstream

rest frame, whereas we computed the escape probability in the downstream rest frame. Because the energy is frame-dependent, we compute the energy spectrum of the particles in the upstream rest frame and note that P_{esc} is the same in all reference frames, because it is a ratio of particle numbers.

The average relative energy gain ϵ in conjunction with the escape probability enable us to repeat Fermi's arguments for the case of a parallel shock and, thus, to compute the energy spectrum of the particles. The probability for a particle to at least perform n cycles and to thus have at least the energy $E(n)$ is

$$\begin{aligned} P(n) &= \sum_{k=n}^{\infty} (1-P_{\mathrm{esc}})^k P_{\mathrm{esc}} \\ &= P_{\mathrm{esc}}(1-P_{\mathrm{esc}})^n \sum_{k=0}^{\infty} (1-P_{\mathrm{esc}})^k = (1-P_{\mathrm{esc}})^n \,, \end{aligned} \tag{3.11}$$

where we used that the geometric series $\sum q^k$ converges to $1/(1-q)$ if $|q| < 1$. If we assume a constant injection of particles at the shock wave, a steady state will be reached and we can compute the age distribution of the particles, cf. Sec. 3.1. To this end we introduce τ_c, i.e. the time it takes to perform a cycle. After $t = n\tau_c$ the number of particles injected during the infinitesimal time interval $\mathrm{d}t$ at $t_0 = 0$ reduced to

$$N(t) = Q\,\mathrm{d}t(1-P_{\mathrm{esc}})^n = Q\,\mathrm{d}t(1-P_{\mathrm{esc}})^{t/\tau_c} \,,$$

because all particles that have performed less than n cycles escaped. Dividing this number by the total number of particles at the shock in steady state yields the probability that a particle has a specific age. The steady-state number of particles is $\eta = Q\tau_{\mathrm{esc}}$ where the reciprocal of τ_{esc} is the probability to escape per unit of time, i.e. $\tau_{\mathrm{esc}}^{-1} = P_{\mathrm{esc}}/\tau_c$. Hence, the probability that a particle has age t is

$$P(t) = \frac{P_{\mathrm{esc}}}{\tau_c}(1-P_{\mathrm{esc}})^{t/\tau_c}\,\mathrm{d}t =: p(t)\,\mathrm{d}t \,.$$

A particle with age $t = n\tau_c$ has on average the energy $\ln(E/E_0) = n\epsilon = \epsilon t/\tau_c =: t/\tau_{\mathrm{acc}}$. We can thus compute the probability that a particle has a certain energy by setting $\pi(E)\,\mathrm{d}E \coloneqq p(t(E))\,\mathrm{d}E$. This yields

$$\begin{aligned} \pi(E)\,\mathrm{d}E &= \frac{\tau_{\mathrm{acc}}}{\tau_{\mathrm{esc}}}(1-P_{\mathrm{esc}})^{\ln(E/E_0)\tau_{\mathrm{acc}}/\tau_c}\frac{\mathrm{d}E}{E} \\ &= \frac{\tau_{\mathrm{acc}}}{\tau_{\mathrm{esc}}}\frac{1}{E_0}\left(\frac{E}{E_0}\right)^{-(1+\tau_{\mathrm{acc}}/\tau_{\mathrm{esc}})+\mathcal{O}(P_{\mathrm{esc}}^2)}\mathrm{d}E \,, \end{aligned} \tag{3.12}$$

where we used that $\ln(1-P_{\mathrm{esc}}) = -P_{\mathrm{esc}} + \mathcal{O}(P_{\mathrm{esc}}^2)$. Notice that $P_{\mathrm{esc}} = 4U_2/c$, hence $\mathcal{O}(P_{\mathrm{esc}}^2) = \mathcal{O}((U_2/c)^2)$; the spectral index neglects corrections of the order of the shock velocity divided by the speed of light squared. Moreover, $\pi(E) = C|\mathrm{d}N/\,\mathrm{d}E|$ where $C > 0$ is a constant, because $\pi(E)\,\mathrm{d}E$ is by definition the

number of particles $\mathrm{d}N$ with energies in the range $[E, E + \mathrm{d}E]$ divided by the total number of particles, also see eq. (3.2). Thus, $\pi(E)$ is proportional to the particle energy spectrum.

We remark that the energy spectrum of the particles can be directly computed, i.e. without the introduction of the cycle time τ_c, from the probability $P(n)$ given in eq. (3.11) and the relative energy change ϵ derived in eq. (3.9). See, for example, eq.(8) in Bell (1978), namely

$$\begin{aligned}\ln P(n(E)) = n(E)\ln(1-P_{\mathrm{esc}}) &= -\frac{4}{\epsilon}\frac{U_2}{c}\ln\left(\frac{E}{E_0}\right) + \mathcal{O}((U_2/c)^2) \\ &= -\frac{3U_2}{U_1-U_2}\ln\left(\frac{E}{E_0}\right) + \mathcal{O}((U_2/c)^2)\,,\end{aligned}$$

where we once more used that the energy of the particle and the number of cycles it performed are interchangeable, because $n = \ln(E/E_0)/\epsilon$. Since $P(n(E))$ is proportional to the number of particles with energies greater or equal to E, i.e.

$$P(E) \propto \int_E^\infty \frac{\mathrm{d}N}{\mathrm{d}E'}\,\mathrm{d}E'\,,$$

its derivative with respect to E is proportional to the energy spectrum $\mathrm{d}N/\,\mathrm{d}E$ of the particles. Its derivative is

$$\begin{aligned}\frac{\mathrm{d}P(E)}{\mathrm{d}E} &= \frac{\mathrm{d}}{\mathrm{d}E}\left(\frac{E}{E_0}\right)^{-3U_2/(U_1-U_2)+\mathcal{O}((U_2/c)^2)} \\ &\approx -\frac{1}{E_0}\frac{3U_2}{U_1-U_2}\left(\frac{E}{E_0}\right)^{-(1+3U_2/(U_1-U_2))}\,,\end{aligned} \tag{3.13}$$

where we neglected second order corrections in U_2/c. Keeping in mind that

$$\frac{\tau_{\mathrm{acc}}}{\tau_{\mathrm{esc}}} = \frac{\tau_c}{\epsilon}\frac{P_{\mathrm{esc}}}{\tau_c} = \frac{3U_2}{U_1-U_2}\,, \tag{3.14}$$

shows that the derived energy spectrum is the same as the one we derived in eq. (3.12) taking the detour over the cycle time τ_c.

The merit of the introduction of the cycle time τ_c is to bring out clearly the analogy to Fermi's original computations and to address the issue that to account for the observations Fermi had to require that the ratio of the acceleration time and escape time is the same in all sources; and 'there is no good reason why this should be so' (Blandford & Eichler, 1987, Sec. 3.2). In Fermi's computations two time scales that represent two separate physical processes play a role, namely the the time between accelerating collisions with magnetic irregularities, denoted with τ, and the time T between the collisions in which a particle loses all its energy. In the case of the acceleration of particles at a parallel shock, there is only the scattering of the particles and the related timescale τ_c. The physical details of the scattering process, which we deliberately left in the dark, will

eventually determine the cycle time τ_c and thus the temporal evolution of the spectrum. However, the steady-state spectrum is independent of these details, because τ_c cancels in eq. (3.14). This implies that the spectral index is solely determined by the kinematics of the shock wave, i.e. by the compression ratio r of the shock, because

$$\frac{\tau_{\text{acc}}}{\tau_{\text{esc}}} = \frac{3U_2}{U_1 - U_2} = \frac{3}{r-1}\,.$$

Notice that if $r = 4$, then this is in agreement with the observations which require the spectral index to have a value around -2, i.e. $\mathrm{d}N/\mathrm{d}E \propto E^{-2}$.

3.2 THE SEMI-RELATIVISTIC VFP EQUATION

In the last section, we have seen that it is possible to derive the probability that a particle has an energy in the range $[E, E + \mathrm{d}E]$ by considering possible particle trajectories, evaluating their likelihoods and, subsequently, averaging over them. This procedure is at the heart of statistical methods as presented in Chapter 2 and, eventually, leads to the Boltzmann equation.[9] However, there is a difference: In statistical methods the probability density functions[10] are not derived directly via particle trajectories but they are solutions of kinetic equations like the Boltzmann equation. The information contained in the particle trajectories is represented in the kinetic equations in the form of the coefficients of the PDE, e.g. the $\boldsymbol{B}$- and $\boldsymbol{E}$-fields of the Lorentz force and as boundary conditions. Hence, we are also able to derive the energy spectrum of charged particles accelerated at a parallel shock by solving the Boltzmann equation, if we prescribe the electromagnetic fields and suitable boundary conditions. This is the aim of this section. Notice that a key advantage of this approach is that we get information about the spatial and temporal evolution of the particle distribution, since the single particle distribution function f depends on $t, \boldsymbol{x}, \boldsymbol{p}$. Up to now we computed the energy spectrum at the shock wave only.

The charged particles propagate relativistically through a magnetised background plasma in which they may be accelerated. The corresponding transport equation is the relativistic VFP equation

$$\frac{p^\mu}{m}\frac{\partial \mathcal{F}}{\partial x^\mu} + \frac{q}{m}F_\mu{}^\nu p_\nu \frac{\partial \mathcal{F}}{\partial p_\mu} = \left(\frac{\delta \mathcal{F}}{\delta \tau}\right)_c\,, \tag{3.15}$$

see the last paragraphs of Section (2.1.5) for further explanations. The energetic and charged particles are non-virtual and massive particles, thus the phase-space

[9] Maybe it is possible to explicitly relate the probability density function $p(\theta, \bar{\theta}') = \chi(\theta)\psi(\bar{\theta})'$ and $D(t, \Xi)$.

[10] Notice that the energy spectrum of the particles $\mathrm{d}N/\mathrm{d}E$ is proportional to the probability density function $\pi(E)$ and that the one particle distribution function is also a probability density function (modulo a constant of normalisation)

density $\mathcal{F}(x^\mu, p_\nu)$ is related to the single particle distribution function through

$$\mathcal{F}(x^\mu, p_\nu) = 2mH(E)f(t, \boldsymbol{x}, \boldsymbol{p})\delta(p^\nu p_\nu - m^2c^2)\,, \tag{3.16}$$

where the Heaviside step function $H(E)$ guarantees that the energy of the particles is positive and the delta distribution enforces the relativistic energy-momentum relation. The electromagnetic field tensor models the $\boldsymbol{E}$- and $\boldsymbol{B}$-fields of the background plasma, namely

$$qF_\mu{}^\nu p_\nu = q\gamma\left(\boldsymbol{E}^{\mathrm{BG}} \cdot \boldsymbol{v}, -[\boldsymbol{E}^{\mathrm{BG}} + \boldsymbol{v} \times \boldsymbol{B}^{\mathrm{BG}}]\right)\,.$$

At this point a key assumption of the proposed particle acceleration model becomes visible. The electromagnetic fields of the background plasma are unaffected by the energetic particles. We call this assumption the *test particle limit*.

If we would like to solve eq. (3.15) we need an explicit expression for the collision term. In the last section we mentioned that the energetic particles are scattered by Alfvén waves that they generate when they stream through the upstream region of the shock wave or by electromagnetic turbulence in the downstream. We take this description as the starting point of our derivation of a collision term. In a first step we investigate what Alfvén waves are.

3.2.1 ALFVÉN WAVES

In general, waves are characterised by their *phase velocity* and a *dispersion relation*. We model the background plasma with the ideal MHD equations (2.93) – (2.96). To derive the dispersion relations of possible MHD waves we assume that the background plasma is in an *equilibrium state*, i.e. it can be characterized by the uniform quantities $\rho_0, P_0, \boldsymbol{U}_0, \boldsymbol{B}_0$ that do not change over time. For simplicity we set its velocity to zero, i.e. $\boldsymbol{U}_0 = 0$. As before the pressure P_0 is isotropic and the adiabatic equation of state holds, we say the plasma is *isentropic*. (Thorne & Blandford, 2017, Sec. 19.7.2) We disturb the equilibrium by increasing the plasma velocity to $\delta\boldsymbol{U}$. In our analysis of the consequences of this disturbance we ignore terms of order $\mathcal{O}(|\delta\boldsymbol{U}|^2)$ and we presuppose that the derivatives of all implied changes, namely $\partial/\partial t$ and $\partial/\partial x^i$ of $\delta\rho, \delta P$ and $\delta\boldsymbol{B}$ are of order $\mathcal{O}(|\delta\boldsymbol{U}|)$. Notice that this is plausible if one anticipates that a 'small' perturbation of an equilibrium will be describable as a superposition of plane waves whose amplitudes will depend on the magnitude of the stimulus producing them. We note that the following computations can be found in Dendy (1990, cf. Sec. 4.3). Under the above assumptions, the conservation of mass equation becomes

$$\frac{\partial}{\partial t}(\rho_0 + \delta\rho) + \nabla_x \cdot ((\rho_0 + \delta\rho)\delta\boldsymbol{U}) = 0\,.$$

Neglecting second-order terms and keeping in mind that ρ_0 is independent of t and $\boldsymbol{x}$, yields

$$\frac{\partial}{\partial t}\delta\rho = -\rho_0 \nabla_x \cdot \delta\boldsymbol{U}. \tag{3.17}$$

Proceeding analogously with the adiabatic equation of state gives

$$\begin{aligned}\frac{\mathrm{D}}{\mathrm{D}t}((P_0+\delta P)(\rho_0+\delta\rho)^{-\gamma}) &= \frac{\partial}{\partial t}\delta P(\rho_0+\delta\rho)^{-\gamma} - \gamma(P_0+\delta P)(\rho_0+\delta\rho)^{-\gamma-1}\frac{\partial}{\partial t}\delta\rho \\ &= \frac{\partial}{\partial t}\delta P\rho_0^{-\gamma} - \gamma P_0\rho_0^{-\gamma-1}\frac{\partial}{\partial t}\delta\rho + \mathcal{O}(|\delta\boldsymbol{U}|^2) \\ &= 0\,.\end{aligned}$$

Note that we did a Taylor expansion of $(\rho_0+\delta\rho)^{-\gamma}$. The previous equation is equivalent to

$$\frac{\partial}{\partial t}\delta P \approx \frac{\gamma P_0}{\rho_0}\frac{\partial}{\partial t}\delta\rho =: c_s^2\frac{\partial}{\partial t}\delta\rho = -\rho_0 c_s^2 \nabla_x\cdot\delta\boldsymbol{U}, \tag{3.18}$$

where we defined the *speed of sound* c_s and used the mass conservation equation.

Before we investigate how the disturbance changes the momentum of the background plasma, we reformulate the $\boldsymbol{J}\times\boldsymbol{B}$ term in the momentum conservation equation (2.94), namely

$$\begin{aligned}\rho\frac{\mathrm{D}U^i}{\mathrm{D}t} - \frac{\partial P}{\partial x^i} = (\boldsymbol{J}\times\boldsymbol{B})^i &= \frac{1}{\mu_0}((\nabla_x\times\boldsymbol{B})\times\boldsymbol{B})^i \\ &= \frac{1}{\mu_0}\epsilon_{ikl}\epsilon_{kmn}\frac{\partial B^n}{\partial x^m}B^l \\ &= \frac{1}{\mu_0}(\delta_{in}\delta_{lm}-\delta_{im}\delta_{ln})\frac{\partial B^n}{\partial x^m}B^l \\ &= -\frac{1}{\mu_0}B_l\frac{\partial B^l}{\partial x^i} + \frac{1}{\mu_0}\frac{\partial B^i}{\partial x^l}B^l \\ &= -\frac{\partial}{\partial x^i}\left(\frac{B^2}{2\mu_0}\right) + \frac{1}{\mu_0}(\boldsymbol{B}\cdot\nabla_x)B^i\,.\end{aligned}$$

Notice that we used Ampère's law as presented in eq. (2.91) and the formula $\epsilon_{ijk}\epsilon_{ilm} = \delta_{il}\delta_{km} - \delta_{im}\delta_{kl}$. The first term on the right-hand side is the *magnetic pressure* and the second term is the *magnetic tension*. The disturbed quantities evolve as

$$\frac{\partial}{\partial t}\delta U^i = -\frac{1}{\rho_0}\frac{\partial}{\partial x^i}\delta P - \frac{c_A^2}{B_0}\left(\boldsymbol{n}\cdot\frac{\partial}{\partial x^i}\delta\boldsymbol{B} - (\boldsymbol{n}\cdot\nabla_x)\delta B^i\right), \tag{3.19}$$

where $\boldsymbol{n} \coloneqq \mathbf{B}_0/B_0$ is the unit vector that points into the direction of the guide field $\mathbf{B}_0$ and $c_A \coloneqq B_0/\sqrt{\rho_0\mu_0}$ is the *Alfvén velocity*.

We derived an evolution equation for the pressure and the velocity perturbation. We lack an equation for $\delta\boldsymbol{B}$ to complete the set of equations. The

evolution of the magnetic field in ideal MHD is determined by the induction equation (2.90), i.e.

$$\begin{aligned}\frac{\partial}{\partial t}\delta B^i &= \epsilon_{ijk}\frac{\partial}{\partial x^j}\left(\delta \boldsymbol{U}\times \boldsymbol{B}_0\right)^k \\ &= \epsilon_{ijk}\epsilon_{klm}B_0^m\frac{\partial}{\partial x^j}\delta U^l \\ &= (\boldsymbol{B}_0\cdot\nabla_x)\delta U^i - B_0^i(\nabla_x\cdot\delta\boldsymbol{U})\,.\end{aligned} \tag{3.20}$$

We combine the three equations (3.18)–(3.20) to solve for the velocity perturbation. We take the derivative of eq. (3.19) with respect to time and replace the time derivatives of the pressure and magnetic field disturbances using the equations (3.18) and (3.20) respectively. This yields

$$\begin{aligned}\frac{\partial^2}{\partial t^2}\delta U^i = (c_s^2+c_A^2)\frac{\partial}{\partial x^i}\nabla_x\cdot\delta\boldsymbol{U} \\ - c_A^2\left((\boldsymbol{n}\cdot\nabla_x)\boldsymbol{n}\cdot\frac{\partial}{\partial x^i}\delta\boldsymbol{U} - (\boldsymbol{n}\cdot\nabla_x)^2\delta U^i + n^i(\boldsymbol{n}\cdot\nabla_x)(\nabla_x\cdot\delta\boldsymbol{U})\right).\end{aligned}$$

We choose a plane wave ansatz for the velocity perturbation, namely $\delta\boldsymbol{U} = \boldsymbol{A}\exp(i\boldsymbol{k}\cdot\boldsymbol{x}-i\omega t)$ where $\boldsymbol{k}$ is the *wave vector* and ω the *(angular) wave frequency*. We denote with $k_\parallel \coloneqq \boldsymbol{k}\cdot\boldsymbol{n}$ the component of the wave vector that is parallel to guide field $\boldsymbol{B}_0$. The ansatz implies that

$$\left[c_A^2k_\parallel^2\mathbf{1} + (c_s^2+c_A^2)\boldsymbol{k}\otimes\boldsymbol{k} - c_A^2k_\parallel(\boldsymbol{k}\otimes\boldsymbol{n}+\boldsymbol{n}\otimes\boldsymbol{k}) - \omega^2\mathbf{1}\right]\boldsymbol{A} = 0\,, \tag{3.21}$$

where we introduced the *dyadic product* of two vectors, i.e. $(\boldsymbol{v}\otimes\boldsymbol{w})^{ij} \coloneqq v^iw^j$. Note that this is an eigenvalue problem. The eigenvalues ω^2 are the roots of the *characteristic polynomial* and since the matrix is built out of the wave vector $\boldsymbol{k}$, the wave frequency ω has to depend on $\boldsymbol{k}$ as well, i.e. the roots of the characteristic polynomial are the dispersion relations of the MHD waves.

Instead of formally solving the eigenvalue problem, we note that all dyadic products in eq. (3.21) vanish if we multiply them with a vector orthogonal to both the wave vector $\boldsymbol{k}$ and the direction of the guide field $\boldsymbol{n}$.[11] Hence, $\boldsymbol{A} = a\boldsymbol{n}\times\boldsymbol{k}$ with $a\in\mathbb{R}$ is an eigenvector and plugging it into equation (3.21) yields the dispersion relation

$$\omega^2 = c_A^2k_\parallel^2 = c_A^2(\boldsymbol{k}\cdot\boldsymbol{n})^2 = c_A^2k^2\cos^2\theta\,, \tag{3.22}$$

where θ is the angle between the direction into which the wave propagates and $\boldsymbol{B}_0$. This wave is called is the *intermediate magnetosonic mode* or *Alfvén mode*. Plugging $\delta\boldsymbol{U} = a(\boldsymbol{n}\times\boldsymbol{k})\exp(i\boldsymbol{k}\cdot\boldsymbol{x}-i\omega t)$ into eq. (3.20) gives

$$\delta\boldsymbol{B} = a\left(\frac{\boldsymbol{k}}{\omega}\cdot\boldsymbol{B}_0\right)(\boldsymbol{n}\times\boldsymbol{k})\exp(i\boldsymbol{k}\cdot\boldsymbol{x}-i\omega t)\,. \tag{3.23}$$

[11] $\boldsymbol{v}\otimes\boldsymbol{w}\boldsymbol{x} = \boldsymbol{v}(\boldsymbol{w}\cdot\boldsymbol{x})$, hence a dyadic product multiplied with a vector that is orthogonal to its second factor equals zero.

The $\boldsymbol{B}$-field perturbation and the velocity perturbation oscillate into the same direction, namely perpendicularly to the magnetic field $\boldsymbol{B}_0$ and the wave vector $\boldsymbol{k}$. The Alfvén wave is thus a *transverse wave*. Moreover, eq. (3.17) tells us that the wave does not compress the plasma, i.e. $\delta\rho = 0$. For the sake of completeness, we mention that the other two eigenmodes of the system correspond to the *fast and slow magnetosonic mode*. (Thorne & Blandford, 2017, cf. Sec. 19.7.2)

For later reference, we derive an expression for the energy density of an Alfvén wave. Since an Alfvén wave does not compress the plasma, we assume that its energy density is

$$\mathcal{U} = \frac{1}{2}\rho_0 \delta U^2 + \frac{\delta B^2}{2\mu_0}\,,$$

where the first term is its kinetic energy density and the second term is the energy density of an electromagnetic wave in vacuum, which we assume to equal the electromagnetic energy density of the Alfvén wave. Because we set $\boldsymbol{A} = a\boldsymbol{n} \times \boldsymbol{k}$ to be the amplitude of the velocity perturbation, eq. (3.23) implies that

$$\delta U^2 = \frac{\omega^2 \delta B^2}{(\boldsymbol{k}\cdot\boldsymbol{B}_0)^2} = c_A^2 \frac{\delta B^2}{B_0^2}\,,$$

where we used the dispersion relation given in eq. (3.22). Together with the definition of the Alfvén speed, namely $c_A = B_0/\sqrt{\rho_0\mu_0}$, the energy density of an Alfvén wave is

$$\mathcal{U} = \frac{1}{2}c_A^2 \frac{\delta B^2}{B_0^2}\rho_0 + \frac{\delta B^2}{2\mu_0} = \frac{\delta B^2}{\mu_0}\,. \tag{3.24}$$

For a more careful and detailed discussion of the energy densities of plasma waves we refer the reader to Kulsrud (2005, Sec. 5.5) and Bellan (2006, Sec. 7.4 – 7.6).

3.2.2 COLLISION OPERATOR

Having a mathematical description of an Alfvén wave, we compute how a charged and relativistic particle interacts with it. We follow Blandford and Eichler (1987, eq. 3.14 – 3.17) and apply *classical perturbation theory*. We start in a reference frame in which the guide field $\boldsymbol{B}_0$ is aligned with the x-axis and work with a circularly polarised Alfvén wave, i.e.

$$\boldsymbol{B}(t,\boldsymbol{x}) = B_0\boldsymbol{e}_x + A\delta\hat{\boldsymbol{B}} = \begin{pmatrix} B_0 \\ A\cos(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t + \Phi) \\ A\sin(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t + \Phi) \end{pmatrix} = \begin{pmatrix} B_0 \\ A\cos([\boldsymbol{k}\cdot\boldsymbol{v} - \omega]t + \Phi) \\ A\sin([\boldsymbol{k}\cdot\boldsymbol{v} - \omega]t + \Phi) \end{pmatrix},$$

where $A \ll |\boldsymbol{B}_0|$ is the amplitude of the wave, $\delta\hat{\boldsymbol{B}}$ is a unit vector and Φ is the phase of the wave. Moreover, $\boldsymbol{x} = \boldsymbol{v}t$ is the position of the particle . The velocity of the wave is $\boldsymbol{c}_A = c_A\boldsymbol{k}/k$. Assuming that $c_A \ll c$, we do a Galilean transformation into a frame in which the wave is static. Notice that the magnetic

field, time and space are unchanged by Galilean transformations, i.e. $\boldsymbol{B}' = \boldsymbol{B}$, $t' = t$ and $k' = k$. In this frame the particle has velocity $\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{c}_A$ and

$$\delta B_y = A\cos([\boldsymbol{k}\cdot\boldsymbol{v}' + \boldsymbol{k}\cdot\boldsymbol{c}_A - \omega]t + \Phi) = A\cos(\boldsymbol{k}\cdot\boldsymbol{v}'t + \Phi)\,,$$

analogously for δB_z. Thus, the particle is experiencing a static $\boldsymbol{B}$-field helix. Since the **B**-field does not vary in time, there is no $\boldsymbol{E}$-field because of the Maxwell-Faraday equation (2.38). Hence, in the wave rest frame only the magnetic force acts on the particle and its energy is conserved, because

$$\frac{\mathrm{d}p'^2}{\mathrm{d}\tau} = 2\boldsymbol{p}'\cdot\frac{\mathrm{d}\boldsymbol{p}'}{\mathrm{d}\tau} = 2\gamma q\boldsymbol{p}'\cdot(\boldsymbol{v}'\times\boldsymbol{B}) = 0\,,$$

where τ is the proper time of the particle.

Since $\delta\boldsymbol{B}$ is assumed to be a small disturbance of $\boldsymbol{B}_0$, we expect the particles trajectory to be mainly determined by the guide field. Hence, we choose the ansatz $\boldsymbol{p} = \boldsymbol{p}_0 + \delta\boldsymbol{p}$ with $\delta\boldsymbol{p} \coloneqq A\boldsymbol{p}_1$, i.e. the perturbation of the particle's trajectory is of the same order as the perturbation of the guide field.[12] Note that we dropped the prime; for the rest of the computation, momentum is defined in the rest frame of the wave. Plugging this ansatz in the equation of motion yields

$$\frac{\mathrm{d}\boldsymbol{p}_0}{\mathrm{d}\tau} - \gamma q\,(\boldsymbol{v}_0\times B_0) + A\left[\frac{\mathrm{d}\boldsymbol{p}_1}{\mathrm{d}\tau} - \gamma q\left(\boldsymbol{v}_0\times\delta\hat{\boldsymbol{B}}_0\right) - \gamma q\boldsymbol{v}_1\times\boldsymbol{B}_0\right] + \mathcal{O}(A^2) = 0\,.$$

Notice the subscript of $\delta\hat{\boldsymbol{B}}_0$, we also had to expand its sine and cosine components, because $\boldsymbol{k}\cdot\boldsymbol{v} = \boldsymbol{k}\cdot(\boldsymbol{v}_0 + A\boldsymbol{v}_1)$.

The solution of the zeroth order equation is the helical motion of a charged particle moving along a constant **B**-field, namely

$$\boldsymbol{v}_0 = \begin{pmatrix} v_{0,x} \\ v_{0,\perp}\cos(\omega_g t + \phi) \\ v_{0,\perp}\sin(\omega_g t + \phi) \end{pmatrix},$$

where $v_\perp^2 = v_y^2 + v_z^2$ and $\omega_g = qB_0/\gamma m$ is the relativistic gyro frequency, see Dendy (1990, eq. 2.8a, b). The equation of motion determining the evolution of the perturbation is

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{v}_1}{\mathrm{d}t} &= \frac{q}{\gamma m}\left[\boldsymbol{v}_0\times\delta\hat{\boldsymbol{B}}_0 + \boldsymbol{v}_1\times\boldsymbol{B}_0\right] \\ &= \frac{q}{\gamma m}\begin{pmatrix} v_{0,\perp}\left[\cos(\omega_g t+\phi)\sin(\boldsymbol{k}\cdot\boldsymbol{v}_0 t + \Phi) - \sin(\omega_g t+\phi)\cos(\boldsymbol{k}\cdot\boldsymbol{v}_0 t+\Phi)\right] \\ -v_{0,x}\sin(\boldsymbol{k}\cdot\boldsymbol{v}_0 t+\Phi) + B_0 v_{1,z} \\ v_{0,x}\cos(\boldsymbol{k}\cdot\boldsymbol{v}_0 t+\Phi) - B_0 v_{1,y}\end{pmatrix}\end{aligned}$$

Notice that $\mathrm{d}/\mathrm{d}\tau = \gamma\mathrm{d}/\mathrm{d}t$. We furthermore used that the energy of the particle does not change in the rest frame of the wave, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma m\boldsymbol{v}_i = \gamma m\frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} \quad \text{for } i\in\{0,1\}\,.$$

[12] Since the perturbation parameter is A and the units of $\boldsymbol{A}$ are tesla, the units of the perturbed quantities need to be divided by T. For example, $\boldsymbol{p}_1$ has the units $\mathrm{kg\,T^{-1}\,m\,s^{-1}}$.

We proceed with solving the equation for $v_{1,x}$, i.e.

$$\frac{\mathrm{d}v_{1,x}}{\mathrm{d}t} = v\frac{\mathrm{d}}{\mathrm{d}t}\mu_1 = \frac{v_{0,\perp}q}{\gamma m}\sin(\boldsymbol{k}\cdot\boldsymbol{v}_0 t - \omega_g t + \chi)\,, \tag{3.25}$$

where we took advantage of the fact that particles' energy does not change, i.e. $v_x = v_{x,0} + Av_{1,x} = v(\mu_0 + A\mu_1)$ which implies that $v_{1,x} = v\mu_1$. Moreover, we used the trigonometric identity $\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$ and we defined the relative phase $\chi \coloneqq \Phi - \phi$ between the Alfvén wave and the gyro motion.[13] (cf. Blandford & Eichler, 1987, eq. 3.14)

The solution of eq. (3.25) is

$$\begin{aligned}\mu_1(t) &= \frac{v(1-\mu_0^2)^{1/2}q}{p}\int_0^t \sin\left[(\boldsymbol{k}\cdot\boldsymbol{v}_0 - \omega_g)t' + \chi\right]\mathrm{d}t' \\ &= \frac{v(1-\mu_0^2)^{1/2}q}{p}\frac{1}{\boldsymbol{k}\cdot\boldsymbol{v}_0 - \omega_g}\left(\cos\left[(\boldsymbol{k}\cdot\boldsymbol{v}_0 - \omega_g)t + \chi\right] - \cos\chi\right)\,,\end{aligned} \tag{3.26}$$

where $v_{0,\perp} = v(1-\mu_0^2)^{1/2}$.

We solve the equations for $v_{1,y}$ and $v_{1,z}$ in one go by setting $v_{1,\perp} \coloneqq v_{1,y} + \mathrm{i}v_{1,z}$. With this definition at hand, the corresponding equation of motion is

$$\begin{aligned}\frac{\mathrm{d}v_{1,\perp}}{\mathrm{d}t} &= \mathrm{i}\frac{qv_{0,x}}{\gamma m}\left[\cos(\boldsymbol{k}\cdot\boldsymbol{v}_0 t + \Phi) + \mathrm{i}\sin(\boldsymbol{k}\cdot\boldsymbol{v}_0 t + \Phi)\right] - \mathrm{i}\omega_g(v_{1,y} + iv_{1,z}) \\ &= \mathrm{i}\frac{qv_{0,x}}{\gamma m}\exp[\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{v}_0 t + \Phi)] - \mathrm{i}\omega_g v_{1,\perp}\,.\end{aligned}$$

Assuming that the perturbation of the particles' velocity was zero at the beginning of the wave-particle interaction, i.e. $v_{1,\perp}(t=0) = 0$, the solution to this ODE is

$$v_{1,\perp}(t) = \frac{qv_{0,x}}{\gamma m}\frac{\exp(\mathrm{i}\Phi)}{\boldsymbol{k}\cdot\boldsymbol{v}_0 + \omega_g}\left[\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{v}_0 t) - \exp(-\mathrm{i}\omega_g t)\right]\,. \tag{3.27}$$

Since the energy of the particle does not change during the interaction, it is the direction of motion that changes. If we use spherical coordinates for the momentum of the particle, i.e. $\boldsymbol{p} = p(\cos\theta, \sin\theta\cos\varphi, \sin\theta\sin\varphi)^{\mathsf{T}}$, then this direction is given by the the angles θ and φ. We thus also work out the perturbation φ_1 of the particles' gyro phase φ. The gyro phase is

$$\varphi = \arctan\left(\frac{\Im(v_\perp)}{\Re(v_\perp)}\right) = \arctan\left(\frac{\Im(v_{0,\perp} + Av_{1,\perp})}{\Re(v_{0,\perp} + Av_{1,\perp})}\right)\,,$$

where $\Im(z)$ and $\Re(z)$ denote the imaginary or the real part of the number $z \in \mathbb{C}$. $v_{0,\perp}$ is defined analogously to $v_{1,\perp}$, namely $v_{0,\perp} \coloneqq v_{0,y} + \mathrm{i}v_{0,z}$. A Taylor

[13] Note that in Blandford and Eichler (1987) χ is shifted by $\pi/2$, whence the cosine instead of the sine. Notice that the relative phase is arbitrary.

expansion in A gives

$$\begin{aligned}\varphi &\approx \arctan\left(\frac{\mathfrak{I}(v_{0,\perp})}{\mathfrak{R}(v_{0,\perp})}\right) + \frac{\mathfrak{R}(v_{0,\perp})\mathfrak{I}(v_{1,\perp}) - \mathfrak{I}(v_{0,\perp})\mathfrak{R}(v_{1,\perp})}{\mathfrak{I}^2(v_{0,\perp}) + \mathfrak{R}^2(v_{0,\perp})} A \\ &= \varphi_0 + \frac{1}{v_{0,\perp}^2}\left(\mathfrak{R}(v_{0,\perp})\mathfrak{I}(v_{1,\perp}) - \mathfrak{I}(v_{0,\perp})\mathfrak{R}(v_{1,\perp})\right) A \\ &=: \varphi_0 + \varphi_1 A\,,\end{aligned}$$

where we neglected terms of order $\mathcal{O}(A^2)$ and used that $\mathfrak{I}^2(v_{0,\perp}) + \mathfrak{R}^2(v_{0,\perp}) = v_{0,\perp}^2(\cos^2(\omega_g t + \phi) + \sin^2(\omega_g t + \phi)) = v_{0,\perp}^2$.

The explicit expression for φ_1 is very similar to the expression for μ_1, namely

$$\varphi_1(t) = \frac{v_{0,x}}{v_{0,\perp}} \frac{q}{\gamma m} \frac{\sin\left[(\boldsymbol{k}\cdot\boldsymbol{v}_0 - \omega_g)t + \chi\right] + \sin(2\omega_g t - \chi)}{\boldsymbol{k}\cdot\boldsymbol{v}_0 + \omega_g}\,. \tag{3.28}$$

Having derived an expression for an Alfvén wave and having worked out how such a wave interacts with a charged particle, we are in a position to devise a mathematical formulation for the collision term. Because the changes of the particle's trajectories could be treated as small deviations from their helical motions, it makes sense to apply the Fokker–Planck theory in going from one particle to a 'statistical' set of particles. At the end of Sec. 2.1.5 we showed that if the changes in momentum $\delta\boldsymbol{p}$ (or velocity) are small, the collision term can be computed with the help of the transition probability $\psi(\boldsymbol{p}, \delta\boldsymbol{p})$, see eqs. (2.60) – (2.62). The perturbation of the momentum, i.e. of the pitch angle, eq. (3.26), and the gyro phase, eq. (3.28), depends on the relative phase χ between the Alfvén wave and the gyro motion of the particle. For the case of many wave-particle interactions, we assume that χ is a uniformly distributed random variable and, hence, the transition probability (density) is

$$\psi(\boldsymbol{p}, \delta\boldsymbol{p}) \propto \frac{1}{2\pi}\,.$$

This implies

$$\psi(\boldsymbol{p} - \delta\boldsymbol{p}, \delta\boldsymbol{p}) = \psi(\boldsymbol{p}, -\delta\boldsymbol{p})\,, \tag{3.29}$$

because for a perturbation $\delta\boldsymbol{p}$ corresponding to an arbitrary but fixed χ, there is an equally likely encounter of a particle and a wave with a relative phase χ' that corresponds to a change in momentum equal to $-\delta\boldsymbol{p}$, see eqs. (3.26) and (3.28). If this is the case we speak of *detailed balance*. Notice that we assumed a fixed duration t of the interaction.

Detailed balance allows us to simplify the right-hand side of the Fokker–Planck equation (2.62). We integrate equation (3.29) and Taylor expand the transition probability density in its first argument at $\boldsymbol{p}$, i.e.

$$\begin{aligned}&\int_{\mathbb{R}^3} \psi(\boldsymbol{p}, -\delta\boldsymbol{p})\,\mathrm{d}^3\delta p \\ &\qquad = \int_{\mathbb{R}^3} \left(\psi(\boldsymbol{p}, \delta\boldsymbol{p}) - \frac{\partial\psi}{\partial p^i}\delta p^i + \frac{1}{2}\frac{\partial^2\psi}{\partial p^i\,\partial p^j}\delta p^i \delta p^j\,\mathrm{d}^3\delta p + \mathcal{O}(|\delta\boldsymbol{p}|^3)\right)\mathrm{d}^3\delta p\,.\end{aligned}$$

The integration over $\delta \boldsymbol{p}$ and the derivatives $\partial/\partial p^i$ with respect to the components can be interchanged and keeping in mind that ψ is normalised to one[14] shows that the above equation is equivalent to

$$\frac{\partial}{\partial p^i}\left[\left\langle\frac{\delta p^i}{\delta t}\right\rangle - \frac{1}{2}\frac{\partial}{\partial p^j}\left\langle\frac{\delta p^i \delta p^j}{\delta t}\right\rangle\right] = \frac{\partial}{\partial p^i}\left[F^i - \frac{\partial D^{ij}}{\partial p^j}\right] = 0\,,$$

where we divided by δt to recover the drift vector $\boldsymbol{F}$ and the diffusion tensor D^{ij} that we defined in the eqs. (2.60) and (2.61) respectively. We integrate this equation over an arbitrary volume V, namely

$$0 = \int_V \frac{\partial}{\partial p^i}\left[F^i - \frac{\partial D^{ij}}{\partial p^j}\right] \mathrm{d}^3 p = \int_{\partial V}\left[F^i - \frac{\partial D^{ij}}{\partial p^j}\right] n_i \,\mathrm{d}S\,.$$

This implies that the term in square brackets must equal zero, because the integral is equal to zero for arbitrary surfaces ∂V. Hence, if detailed balance holds the drift vector and the diffusion tensor are related, i.e.

$$F^i = \frac{\partial}{\partial p^j} D^{ij}\,. \tag{3.30}$$

This relation turns the Fokker–Planck equation (2.62) into

$$\begin{aligned}\left(\frac{\delta f}{\delta t}\right)_c &= -\frac{\partial f F^i}{\partial p^i} + \frac{\partial^2 f D^{ij}}{\partial p^i \partial p^j} \\ &= \frac{\partial}{\partial p^i}\left[-f\frac{\partial D^{ij}}{\partial p^j} + D^{ij}\frac{\partial f}{\partial p^j} + f\frac{\partial D^{ij}}{\partial p^j}\right] = \frac{\partial}{\partial p^i}\left[D^{ij}\frac{\partial f}{\partial p^j}\right],\end{aligned} \tag{3.31}$$

i.e. a diffusion equation.

Because the energy of the particle does not change in the wave-particle interaction, we use spherical coordinates in momentum space. However, the components of the diffusion tensor D^{ij} are defined in a Cartesian coordinate system, see its definition in eq. (2.61). We now derive expressions for the components of D in spherical coordinates. The nabla operator in spherical coordinates is

$$\nabla_p f = \boldsymbol{e}_p \frac{\partial f}{\partial p} + \boldsymbol{e}_\theta \frac{1}{p}\frac{\partial f}{\partial \theta} + \boldsymbol{e}_\varphi \frac{1}{p\sin\theta}\frac{\partial f}{\partial \varphi}\,.$$

The contraction of $\nabla_p f$ and the diffusion tensor yields

$$z^i := D^{ij} e_{p,j}\frac{\partial f}{\partial p} + D^{ij} e_{\theta,j}\frac{1}{p}\frac{\partial f}{\partial \theta} + D^{ij} e_{\varphi,j}\frac{1}{p\sin\theta}\frac{\partial f}{\partial \varphi}\,.$$

Notice that the right-hand side of the Fokker–Planck equation (3.31) equals $\nabla_p \cdot \boldsymbol{z}$. The divergence in spherical coordinates is

$$\nabla_p \cdot \boldsymbol{z} = \frac{1}{p^2}\frac{\partial}{\partial p}\left(p^2 z_p\right) + \frac{1}{p\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta z_\theta\right) + \frac{1}{p\sin\theta}\frac{\partial z_\varphi}{\partial \varphi}\,,$$

[14] It might not be obvious that $\int_{\mathbb{R}^3} \psi(\boldsymbol{p}, -\delta\boldsymbol{p})\,\mathrm{d}^3\delta p = 1$ as well. The integration over $\mathbb{R}^3$ can be formalised as an integration over a sphere with infinite radius and a sphere is invariant under parity transformations.

where z_p, z_θ and z_φ are the coordinates of $\boldsymbol{z}$ in the $\boldsymbol{e}_p, \boldsymbol{e}_\theta, \boldsymbol{e}_\varphi$ coordinate system, namely $z_p = \boldsymbol{e}_p \cdot \boldsymbol{z}, z_\theta = \boldsymbol{e}_\theta \cdot \boldsymbol{z}$ and $z_\varphi = \boldsymbol{e}_\varphi \cdot \boldsymbol{z}$. We note that the expression for the nabla operator and the divergence in spherical coordinates can, for example, be found in Beresnyak (2023, Spherical coordinates). In total,

$$\begin{aligned}\frac{\partial}{\partial p^i}\left(D^{ij}\frac{\partial f}{\partial p^j}\right) = \frac{1}{p^2}\frac{\partial}{\partial p}\left[p^2\left(D_{pp}\frac{\partial f}{\partial p} + D_{p\theta}\frac{\partial f}{\partial \theta} + D_{p\varphi}\frac{\partial f}{\partial \varphi}\right)\right] \\ + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left[\sin\theta\left(D_{\theta p}\frac{\partial f}{\partial p} + D_{\theta\theta}\frac{\partial f}{\partial \theta} + D_{\theta\varphi}\frac{\partial f}{\partial \varphi}\right)\right] \\ + \frac{\partial}{\partial \varphi}\left[D_{\varphi p}\frac{\partial f}{\partial p} + D_{\varphi\theta}\frac{\partial f}{\partial \theta} + D_{\varphi\varphi}\frac{\partial f}{\partial \varphi}\right],\end{aligned} \tag{3.32}$$

where we defined the diffusion coefficients $D_{ab} \coloneqq \partial a/\partial p^i \; D^{ij} \; \partial b/\partial p^j$ with $a, b \in \{p, \theta, \varphi\}$ (Risken, 1996, eq. 4.132). We remark that

$$\nabla_p p = \boldsymbol{e}_p\,, \quad \nabla_p \theta = \frac{1}{p}\boldsymbol{e}_\theta \quad \text{and} \quad \nabla_p \varphi = \frac{1}{p\sin\theta}\boldsymbol{e}_\varphi\,, \tag{3.33}$$

where p, θ and φ are considered to be functions of p^x, p^y and p^z. The definition of the diffusion coefficients D_{ab} implies that $D_{ab} = D_{ba}$, because the diffusion tensor is symmetric, namely $D^{ij} = \langle \delta p^i \delta p^j / 2\delta t\rangle = D^{ji}$.

We would like to demonstrate that the definition of the diffusion coefficients D_{ab} agrees as well with the *Einstein–Smoluchowski expression* developed in the context of Brownian motion, namely $D = \langle x^2\rangle/2t$, see, for example, Chandrasekhar (1943, eq. 174). The perturbation $p + \delta p, \theta + \delta\theta, \varphi + \delta\varphi$ of the momentum vector $\boldsymbol{p}$ leads to

$$\begin{aligned}\boldsymbol{p} + \delta\boldsymbol{p} &= (p + \delta p)\begin{pmatrix}\cos\theta - \sin\theta\delta\theta \\ (\sin\theta + \cos\theta\delta\theta)(\cos\varphi - \sin\varphi\delta\varphi) \\ (\sin\theta + \cos\theta\delta\theta)(\sin\varphi + \cos\varphi\delta\varphi)\end{pmatrix} \\ &= \boldsymbol{p} + \boldsymbol{e}_p\delta p + \boldsymbol{e}_\theta p\delta\theta + \boldsymbol{e}_\varphi p\sin\theta\delta\varphi\,.\end{aligned}$$

Note that we neglected all second order terms of the perturbation variables $\delta p, \delta\theta$ and $\delta\varphi$. The computation of the diffusion coefficient $D_{\theta\theta}$ may serve as an example:

$$\begin{aligned}D_{\theta\theta} &= \frac{e_{\theta,i}}{p}\left\langle\frac{\delta p^i \delta p^j}{2\delta t}\right\rangle\frac{e_{\theta,j}}{p} \\ &= \left\langle\frac{1}{2\delta t}\frac{e_{\theta,i}}{p}\left[e_p^i\,\delta p + e_\theta^i\, p\delta\theta + e_\varphi^i\, p\sin\theta\delta\varphi\right]\left[e_p^j\delta p + e_\theta^j p\delta\theta + e_\varphi^j p\sin\theta\delta\varphi\right]\frac{e_{\theta,j}}{p}\right\rangle \\ &= \left\langle\frac{\delta\theta^2}{2\delta t}\right\rangle,\end{aligned}$$

where we exploited the linearity of the expectation value and the orthogonality of $\boldsymbol{e}_p, \boldsymbol{e}_\theta$ and $\boldsymbol{e}_\varphi$. The other diffusion coefficients D_{ab} are computed analogously and, thus, they have the Einstein–Smoluchowski form we expect them to have.

We turn our attention to the evaluation of the diffusion coefficients that model the Alfvén wave-particle interaction, i.e. the coefficients determining the collision operator of the particle transport problem that we are investigating. First, diffusion coefficients related to changes in the magnitude of momentum are zero, because the energy of the particle does not change. This is a consequence of their definition: $D_{ap} = D_{pa} = e_{p,i}\langle \delta p^i \delta p^j / 2\delta t \rangle e_{a,j}$ and $\delta \boldsymbol{p}$ does not contain $\boldsymbol{e}_p$ if $\delta p = 0$. Hence, $D_{pa} = 0$ for $a \in \{p, \theta, \varphi\}$. This leaves us with the three diffusion coefficients $D_{\theta\theta}, D_{\theta\varphi}, D_{\varphi\varphi}$.

The computation of the diffusion coefficients require us to know more about the transition probability density $\psi(\boldsymbol{p}, \delta\boldsymbol{p})$. We mentioned that it will depend on the relative phase χ. However, notice that we kept quite about the fact that the change in the pitch angle (and the gyro phase) also depends on the angle between the direction of propagation of the wave $\boldsymbol{k}/k$ and the particles velocity $\boldsymbol{v}_0$. We circumvent this difficulty by restricting ourselves to a specific direction of $\boldsymbol{k}$, namely we require that $\boldsymbol{k} \parallel \boldsymbol{B}_0$. Furthermore, the time t the interaction takes influences the perturbation. Since we do not have any a priori knowledge about the interactions' durations, we again presuppose that all possibilities are equally likely. Eventually, we have to take into account that the particle interacts not with one, but with many waves of varying amplitude and wave length.

We demonstrate for the pitch-angle diffusion coefficient $D_{\theta\theta}$ how the averaging over all possible changes of momenta $\delta\boldsymbol{p}$ works. We start with eq. (3.26), i.e. the expression for the perturbation of μ. Notice that because $\mu = \mu_0 + A\mu_1 = \cos\theta = \cos(\theta_0 + \delta\theta) = \cos\theta_0 - \sin\theta_0\delta\theta$, the perturbation of θ is $\delta\theta = -A\mu_1/\sin\theta_0$. The diffusion coefficient thus is

$$
\begin{aligned}
\left\langle \frac{\delta\theta^2}{2\delta t} \right\rangle &= \frac{\mu_{\mathrm{mp}}}{2(1-\mu_0^2)} \lim_{t\to\infty} \frac{1}{t} \frac{1}{2\pi} \int_0^\infty \int_0^t \int_0^t \int_0^{2\pi} \frac{\mathrm{d}\mu_1}{\mathrm{d}t'} \frac{\mathrm{d}\mu_1}{\mathrm{d}t''} \frac{A^2}{\mu_{\mathrm{mp}}} \,\mathrm{d}\chi \,\mathrm{d}t' \,\mathrm{d}t'' \,\mathrm{d}k \\
&= \frac{\mu_{\mathrm{mp}}}{4} \frac{v^2 q^2}{p^2} \lim_{t\to\infty} \frac{1}{t} \int_0^\infty \int_0^{2\pi} \left(\frac{\cos\left[(kv\mu_0 - \omega_g)t + \chi\right] - \cos\chi}{(kv\mu_0 - \omega_g)} \right)^2 \mathcal{E}(k) \frac{\mathrm{d}\chi \,\mathrm{d}k}{2\pi} \\
&= \frac{\mu_{\mathrm{mp}}}{4} \frac{\omega_g^2}{v\mu_0 B_0^2} \lim_{t\to\infty} tv\mu_0 \int_0^\infty \frac{1 - \cos\left[(k - \omega_g/v\mu_0)(tv\mu_0)\right]}{(k - \omega_g/v\mu_0)^2 (tv\mu_0)^2} \mathcal{E}(k) \,\mathrm{d}k \\
&= \frac{1}{2}\frac{\pi}{4} \frac{\omega_g^2}{(B_0^2/2\mu_{\mathrm{mp}})v\mu_0} \int_0^\infty \delta(k - \omega_g/v\mu_0) \mathcal{E}(k) \,\mathrm{d}k \\
&= \frac{1}{2}\frac{\pi}{4} \frac{k\mathcal{E}(k)|_{k=\omega_g/v\mu_0}}{(B_0^2/2\mu_{\mathrm{mp}})} \omega_g =: \frac{\nu}{2}
\end{aligned}
\tag{3.34}
$$

Note the unusual notation μ_{mp} for the magnetic permeability; in the current context μ_0 denotes the cosine of the undisturbed pitch angle. In the first line we average over the duration of the interaction t and the relative phase χ. However, the possible values of χ are in $[0, 2\pi]$ whereas t can have values ranging from 0 to infinity. Keeping in mind that all relative phases and durations are equally likely, we arrive at the a factor $1/2\pi$ and the limit $\lim_{t\to\infty} 1/t$.

The integral over all wave vectors reveals its meaning when going from the first to the second equation. In doing so we integrated over t' and t'' and introduced the *energy density spectrum* $\mathcal{E}(k)$, i.e. $\mathcal{E}(k)\,\mathrm{d}k$ is the wave energy density of waves with wave vector in the interval $[k, k+\mathrm{d}k]$. Notice that A^2/μ_{mp} is the energy density of an Alfvén wave, see eq. (3.24), and integrating over an energy density spectrum allows us to represent that a particle interacts with waves of varying amplitude and wave length.[15] We included a factor of 1/2 to account for the fact that the energy density spectrum is defined such that it does not distinguish between left- and right-circularly polarised waves; a particle with a given sign of $\cos\theta_0$ and a charge q only interacts with one of them. Also note that we set $\boldsymbol{k} \parallel \boldsymbol{B}_0$, cf. (3.26).

The delta distribution appears because it can be shown with the help of a contour integral that

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\eta(x)\,\mathrm{d}x \coloneqq \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos(x)}{x^2}\,\mathrm{d}x = 1\,,$$

see Stein et al. (2003, Chap. 2, Example 2). Hence, η can be used as a *nascent delta function*, i.e.

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{-\infty}^{\infty}\eta(x/\varepsilon)f(x)\,\mathrm{d}x = f(0) = \int_{-\infty}^{\infty}\delta(x)f(x)\,\mathrm{d}x\,, \tag{3.35}$$

see Stein et al. (1972, Theorem 1.18.) for the mathematical meaning of the limit. In eq. (3.34) we set $\varepsilon = 1/t\upsilon\mu_0$ and the limit $t\to\infty$ is turned into $\varepsilon\to 0$. The appearance of the delta distribution tells us that only *resonant waves* contribute to the diffusion in pitch angle, namely Alfvén waves whose helical geometry is followed by the gyro motion of the particle or, equivalently, Alfvén waves whose wave length is such that the particle crosses it during one gyro period; notice that $k = \omega_g/\upsilon\mu_0$ is equivalent to $\lambda = T_g\upsilon_{0,x}$, where T_g is the gyro period.

In the last line of eq. (3.34) we defined the collision frequency ν, cf. result and the arguments that led to it with Blandford and Eichler (1987, p. 20).

We proceed with $D_{\varphi\varphi}$. An analogue computation yields

$$\begin{aligned}\left\langle\frac{\delta\varphi^2}{2\delta t}\right\rangle &= \frac{\mu_{\mathrm{mp}}}{4}\lim_{t\to\infty}\frac{1}{t}\int_0^{\infty}\int_0^{2\pi}[\varphi_1(t)-\varphi(0)]^2\mathcal{E}(k)\frac{\mathrm{d}\chi}{2\pi}\,\mathrm{d}k\\ &= \frac{\mu_{\mathrm{mp}}}{4}\frac{\upsilon_{0,x}^2\omega_g^2}{\upsilon_{0,\perp}^2B_0^2}\lim_{t\to\infty}\frac{1}{t}\int_0^{\infty}\frac{1-\cos[(k\upsilon\mu_0+\omega_g)t]}{(k\upsilon\mu_0+\omega_g)^2}\mathcal{E}(k)\,\mathrm{d}k\\ &= \frac{1}{2}\frac{\pi}{4\tan^2\theta_0}\frac{k\mathcal{E}(k)|_{k=-\omega_g/\upsilon\mu_0}}{(B_0^2/2\mu_{\mathrm{mp}})}\omega_g \eqqcolon \frac{\nu_\varphi}{2}\,.\end{aligned} \tag{3.36}$$

[15] In the first equation A^2/μ_{mp} is to be understood symbolically, since A denotes a specific amplitude. It is merely meant to motivate the introduction of the energy density spectrum of the Alfvén waves. The second equation is correct again.

If the energy density of waves travelling into the direction of the guide field $\boldsymbol{B}_0$ is the same as for waves travelling in the opposite direction, then $\nu_\varphi = \nu/\tan^2\theta_0$. This implies that there exists a critical pitch angle θ_c that determines wether gyrophase diffusion dominates pitch-angle diffusion or vice versa. Shalchi (2012) derives analogue results for the diffusion coefficients, in particular he arrives at the same relation between $D_{\theta\theta}$ and $D_{\varphi\varphi}$. He also shows that the diffusion coefficient $D_{\theta\varphi} = 0$. However, he employs a different definition of the diffusion coefficient.

The only diffusion coefficients unequal to zero thus are $D_{\theta\theta}$ and $D_{\varphi\varphi}$. Since the angles θ and φ encode the direction of motion of the particles, these diffusion coefficients imply repeated small changes in the particles' trajectories that eventually lead to an isotropisation of the particle distribution, see Shalchi (2012, Fig. 6). We derived the diffusion coefficients considering the interaction of a particle with circularly polarised Alfvén waves of random phase and with small amplitudes whose direction of propagation were aligned with the guide field. We do not expect that these assumptions hold exactly at particle acceleration sites, see Blandford and Eichler (1987, p. 21). Nonetheless, we keep the idea that there exists a frame in which the energy of the particles is nearly unchanged when interacting with plasma waves and that the net effect of the wave-particle interactions is the isotropisation of the particle distribution. This motivates the following expression for the scattering operator

$$\begin{aligned}\left(\frac{\delta f}{\delta t}\right)_c &= \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\nu}{2}\frac{\partial f}{\partial\theta}\right) + \left(\frac{\nu_\varphi}{2}\frac{\partial f}{\partial\varphi}\right) \\ &\approx \frac{\nu}{2}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\varphi^2}\right] \\ &=: \frac{\nu}{2}\Delta_{\theta,\varphi}\,,\end{aligned} \tag{3.37}$$

cf. the general expression for the diffusion coefficients in spherical coordinates given in eq. (3.32). The key approximations are, firstly, that the scattering frequency is independent of the pitch angle θ and the gyro phase φ and, secondly, that we replaced $\nu_\varphi = \nu/\tan^2\theta$ with $\nu_\varphi = \nu/\sin^2\theta$. Since $\cos^2\theta \leq 1$, the latter approximation overestimates the diffusion of the gyro phase for low values of θ. We note that this scattering operator is said to model *isotropic scattering* and that $\Delta_{\theta,\varphi}$ is the angular part of the Laplacian operator in spherical coordinates.

At the end of this section we comment that we derived the scattering operator for the single particle distribution function $f(t, \boldsymbol{x}, \boldsymbol{p})$. However, we would like to work with the fully relativistic Vlasov–Fokker–Planck equation (3.15). In agreement with eq. (3.16) we define

$$\left(\frac{\delta\mathcal{F}}{\delta\tau}\right)_c := 2mH(E)\gamma\left(\frac{\delta f}{\delta t}\right)_c \delta(p_\mu p^\mu - m^2c^2)\,, \tag{3.38}$$

where γ is the Lorentz factor of the particle in the rest frame of the wave and

where we exploited that the wave-particle interaction does not change the particle's energy (see Achterberg & Norman, 2018, text after eq. 31 and eq. 35).

3.2.3 MIXED COORDINATES

On our way to a solution of the relativistic VFP equation (3.15) for the case of a parallel shock, we derived an explicit expression for the scattering operator. However, the expression is only correct in the rest frame of the Alfvén wave whereas the space and momentum variables (x^μ, p_μ) are defined in a laboratory system, e.g. the rest frame of the shock wave. In the view of the fact that interactions only change the momentum variables of the involved entities, it makes sense to only transform the momentum variables into the rest frame of the waves. We refer to the coordinates whose spatial components are defined in the laboratory system and its momentum components in a different frame as *mixed coordinates*.

A difficulty in defining the mixed coordinates is that the rest frame of the waves is not well defined, since there maybe waves propagating in all directions and if not only Alfvén waves scatter the particles, then the Alfvén and the sound speed need to be taken into account. If the velocity of the background plasma, i.e. the plasma in which the shock wave propagates, and the Alfvén speed are small compared to the speed of light, the rest frame of the background plasma is approximately the rest frame of the waves. The reason is, firstly, that relativistic particles ($\gamma \gg 1$) travel almost at the speed of light and transforming into a slowly moving reference frame does not change their velocity significantly and, secondly, for a particle moving at such a high speed a slowly propagating and oscillating wave appears to be stationary. We defined the Alfvén speed in the rest frame of the background plasma and, for the sake of the argument, we now make the additional assumption that the background plasma is the interstellar medium (ISM). Typical values characterising the ISM are a number density n of 0.1 to $100\,\mathrm{cm}^{-3}$ and a magnetic field strength B_0 between 0.3 to 10 nT. Assuming that the ISM consists completely out of protons and electrons and neglecting the mass of the electrons, yields an Alfvén speed $c_A = B_0/\sqrt{\rho_0\mu_0} \approx 20.5 - 21.8\,\mathrm{km\,s}^{-1}$, see Thorne and Blandford (2017, Sec. 19.7.3) for a similar estimate. Furthermore, typical speeds of a shock wave of a supernova remnant are in the range of $3000\,\mathrm{km\,s}^{-1}$ to $10\,000\,\mathrm{km\,s}^{-1}$ (Blandford & Eichler, 1987, p. 12). Hence, we choose the rest frame of the background plasma as an approximation to the rest frame of the waves.

We now transform the momentum coordinates into the rest frame of the background plasma. We note that the following computations summarise the Sections 2–4 of Achterberg and Norman (2018). The transformation of the

momentum variables is done with a *restricted Lorentz transformation*[16], namely

$$\boldsymbol{\Lambda}(\boldsymbol{U}) = \begin{pmatrix} \Gamma & -\Gamma \boldsymbol{U}^{\mathsf{T}}/c \\ -\Gamma \boldsymbol{U}/c & \mathbf{1} + (\Gamma - 1)/U^2 \boldsymbol{U} \otimes \boldsymbol{U} \end{pmatrix}. \tag{3.39}$$

The velocity $\boldsymbol{U}$ is the plasma velocity and $\Gamma = (1 - (U/c)^2)^{-1/2}$ is the corresponding Lorentz factor. The inverse transformation is $\boldsymbol{\Lambda}^{-1}(\boldsymbol{U}) = \boldsymbol{\Lambda}(-\boldsymbol{U})$. Henceforth, we use the index notation and introduce the convention that Greek indices refer to quantities defined in the laboratory frame and Roman letters with a hat are indices of quantities that are given in the rest frame of the background plasma, for example, $p^{\hat{a}} = \Lambda^{\hat{a}}_{\mu} p^{\mu}$. We note that Lorentz transformations leave the *Minkowski metric*[17] invariant, i.e.

$$x_{\mu} x^{\mu} = \eta_{\mu\nu} x^{\mu} x^{\nu} = \eta_{\mu\nu} \Lambda^{\mu}_{\hat{a}} x^{\hat{a}} \Lambda^{\nu}_{\hat{b}} x^{\hat{b}} = \eta_{\hat{a}\hat{b}} x^{\hat{a}} x^{\hat{b}} = x_{\hat{a}} x^{\hat{a}}, \tag{3.40}$$

where $\eta_{\mu\nu}$ are the components of the Minkowski metric. We note that we adopt the *signature* convention $(+,-,-,-)$. Eq. (3.40) implies that $\eta_{\hat{a}\hat{b}} = \eta_{\mu\nu} \Lambda^{\mu}_{\hat{a}} \Lambda^{\nu}_{\hat{b}}$. The fact that the inverse Lorentz formation is obtained by inverting the sign of the velocity is in tensor notation expressed as $\Lambda^{\mu}_{\hat{a}} \Lambda^{\hat{a}}_{\nu} = \delta^{\mu}_{\nu}$. Note that exchanging the Greek index with the Roman index is equivalent to two contractions with the Minkowski metric; these contractions flip the sign of the velocity.

Before we transform the relativistic VFP equation into the rest frame of the background plasma, we compute the *fictitious force* that must appear because the velocity of the background plasma $\boldsymbol{U}$ changes in time and space; we thus transform into a *non-inertial*, i.e. an accelerated, reference frame. An expression for the fictitious force is derived by applying Newton's second law to a free particle, i.e.

$$\frac{\mathrm{d}p_{\mu}}{\mathrm{d}\tau} = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\Lambda^{\hat{a}}_{\mu} p_{\hat{a}}\right) = \Lambda^{\hat{a}}_{\mu} \frac{\mathrm{d}p_{\hat{a}}}{\mathrm{d}\tau} + \frac{\mathrm{d}\Lambda^{\hat{a}}_{\mu}}{\mathrm{d}\tau} p_{\hat{a}} = 0,$$

where we applied the product rule. Since $\boldsymbol{U}$ depends on t and $\boldsymbol{x}$, the derivative $\mathrm{d}/\mathrm{d}\tau$ has to be interpreted as the proper time derivative along the particle's orbit, i.e.

$$\frac{\mathrm{d}\Lambda^{\hat{a}}_{\mu}}{\mathrm{d}\tau} = \frac{p^{\nu}}{m} \frac{\partial \Lambda^{\hat{a}}_{\mu}}{\partial x^{\nu}} = \Lambda^{\nu}_{\hat{b}} \frac{p^{\hat{b}}}{m} \frac{\partial \Lambda^{\hat{a}}_{\mu}}{\partial x^{\nu}}.$$

Contracting the first equation with $\Lambda^{\mu}_{\hat{c}}$ and inserting the second equation yields a definition for a fictitious Minkowski force, namely

$$\frac{\mathrm{d}p_{\hat{c}}}{\mathrm{d}\tau} = -\Lambda^{\mu}_{\hat{c}} \Lambda^{\nu}_{\hat{b}} \frac{\partial \Lambda^{\hat{a}}_{\mu}}{\partial x^{\nu}} \frac{p^{\hat{b}} p_{\hat{a}}}{m} =: K_{\hat{c}}, \tag{3.41}$$

cf. Achterberg and Norman (2018, eq. 9).

[16] This means that the orientation of time and space are preserved. (Jeevanjee, 2011, Ex. 4.17.)

[17] The term 'metric' might be misleading. We define $\eta(\boldsymbol{v}, \boldsymbol{w}) = v^0 w^0 - \boldsymbol{v} \cdot \boldsymbol{w}$ with $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^4$. This implies that η is a non-degenerate, symmetric bilinear form. However, η is not positive-definite and, hence, it does not induce a norm.

Transforming the momentum variables in the relativistic VFP equation (3.15) results in

$$\begin{aligned}
&\frac{p^\mu}{m}\frac{\partial \mathcal{F}(x^\mu, p_{\hat{a}})}{\partial x^\mu} + \frac{q}{m}F_\mu{}^\nu p_\nu \frac{\partial \mathcal{F}(x^\mu, p_{\hat{a}})}{\partial p_\mu} \\
&\quad = \frac{p^\mu}{m}\frac{\partial \mathcal{F}}{\partial x^\mu} + \frac{p^\mu}{m}\frac{\partial \mathcal{F}}{\partial p_{\hat{a}}}\frac{\partial p_{\hat{a}}}{\partial x^\mu} + \frac{q}{m}F_\mu{}^\nu p_\nu \frac{\partial \mathcal{F}}{\partial p_{\hat{a}}}\frac{\partial p_{\hat{a}}}{\partial p_\mu} \\
&\quad = \frac{p^\mu}{m}\frac{\partial \mathcal{F}}{\partial x^\mu} + \frac{p^\mu p_\nu}{m}\frac{\partial \Lambda^\nu_{\hat{a}}}{\partial x^\mu}\frac{\partial \mathcal{F}}{\partial p_{\hat{a}}} + \frac{q}{m}F_\mu{}^\nu p_\nu \Lambda^\mu_{\hat{a}}\frac{\partial \mathcal{F}}{\partial p_{\hat{a}}} \\
&\quad = \Lambda^\mu_{\hat{a}}\frac{p^{\hat{a}}}{m}\frac{\partial \mathcal{F}}{\partial x^\mu} + \Lambda^\mu_{\hat{c}}\Lambda^{\hat{b}}_\nu \frac{\partial \Lambda^\nu_{\hat{a}}}{\partial x^\mu}\frac{p^{\hat{c}} p_{\hat{b}}}{m}\frac{\partial \mathcal{F}}{\partial p_{\hat{a}}} + \frac{q}{m}\Lambda^{\hat{b}}_\nu F_\mu{}^\nu \Lambda^\mu_{\hat{a}} p_{\hat{b}}\frac{\partial \mathcal{F}}{\partial p_{\hat{a}}} \\
&\quad = \left(\frac{\delta \mathcal{F}}{\delta \tau}\right)_c ,
\end{aligned} \tag{3.42}$$

where we applied the chain rule in the first equation and evaluated the corresponding derivatives in the second equation, namely

$$\frac{\partial p_{\hat{a}}}{\partial x^\mu} = \frac{\partial \Lambda^\nu_{\hat{a}}}{\partial x^\mu} p_\nu \quad \text{and} \quad \frac{\partial p_{\hat{a}}}{\partial p_\mu} = \frac{\partial}{\partial p_\mu}(\Lambda^\nu_{\hat{a}} p_\nu) = \Lambda^\mu_{\hat{a}} .$$

In the the third equation we transformed all momentum variables, e.g. $p^\mu = \Lambda^\mu_{\hat{a}} p^{\hat{a}}$. The scattering operator on the right-hand side of eq. (3.42) is the one that we derived for the wave-particle interaction in the rest frame of the wave, see eq. (3.38). Taking advantage of the fact that $\Lambda^\nu_{\hat{a}}\Lambda^{\hat{b}}_\nu = \delta^{\hat{b}}_{\hat{a}}$ implies

$$\frac{\partial \Lambda^\nu_{\hat{a}}}{\partial x^\mu}\Lambda^{\hat{b}}_\nu = -\Lambda^\nu_{\hat{a}}\frac{\partial \Lambda^{\hat{b}}_\nu}{\partial x^\mu} , \tag{3.43}$$

we obtain

$$\Lambda^\mu_{\hat{a}}\frac{p^{\hat{a}}}{m}\frac{\partial \mathcal{F}}{\partial x^\mu} + \left(K_{\hat{a}} + \frac{q}{m}F_{\hat{a}}{}^{\hat{b}} p_{\hat{b}}\right)\frac{\partial \mathcal{F}}{\partial p_{\hat{a}}} = \left(\frac{\delta \mathcal{F}}{\delta \tau}\right)_c , \tag{3.44}$$

where we also used that $F_{\hat{a}}{}^{\hat{b}} = \Lambda^{\hat{b}}_\nu F_\mu{}^\nu \Lambda^\mu_{\hat{a}}$ is the electromagnetic field tensor in the rest frame of the background plasma. (Achterberg & Norman, 2018, eq. 30). This the *fully relativistic VFP equation in mixed coordinates*.

The single particle distribution function $\mathcal{F}$ does contain the Heaviside function $H(E') = H(p^{\hat{0}})$ and the delta distribution $\delta(p_{\hat{a}}p^{\hat{a}} - m^2c^2)$ that ensure that the energies of the particles are positive and that they are on *mass shell*, i.e. their energy is determined by the energy-momentum relation, see eq. (3.16). Notice that we reintroduced the prime to refer to quantities in the rest frame of the background plasma as an equivalent notation to hatted Roman indices. In a next step we integrate eq. (3.44) over $\int \mathrm{d}p^{\hat{0}} p^{\hat{0}}$ to reduce the number of independent variables from eight to seven and, in agreement with this, to derive an equation for $f(t, \boldsymbol{x}, \boldsymbol{p}')$.[18] We split the integration into three parts: First, we integrate the

[18] I thank Prof. John Kirk who gave me the decisive hints to do the integration.

spatial advection term, secondly we proceed with the *momentum advection term* and, thirdly, we integrate the collision operator.

The integration of the spatial advection term yields

$$\begin{aligned}
&\int_{-\infty}^{\infty} 2p^{\hat{0}}\Lambda^{\mu}_{\hat{a}}p^{\hat{a}}\frac{\partial f(t,\boldsymbol{x},\boldsymbol{p}')}{\partial x^{\mu}}H(p^{\hat{0}})\delta(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad=\int_{-\infty}^{\infty}\frac{p^{\hat{0}}}{(p'^2+m^2c^2)^{1/2}}\Lambda^{\mu}_{\hat{a}}p^{\hat{a}}\frac{\partial f(t,\boldsymbol{x},\boldsymbol{p}')}{\partial x^{\mu}}\\
&\qquad\qquad\times H(p^{\hat{0}})\left[\delta\left(p^{\hat{0}}-(p'^2+m^2c^2)^{1/2}\right)+\delta\left(p^{\hat{0}}+(p'^2+m^2c^2)^{1/2}\right)\right]\mathrm{d}p^{\hat{0}}\\
&\quad=\Lambda^{\mu}_{\hat{a}}p^{\hat{a}}\frac{\partial f(t,\boldsymbol{x},\boldsymbol{p}')}{\partial x^{\mu}}\,,
\end{aligned}\tag{3.45}$$

where we used that the composition of the delta distribution with a function, say $g(x)$, is

$$\delta(g(x))=\sum_i\frac{\delta(x-x_i)}{|g'(x_i)|}\,.$$

The sum extends over all roots x_i of $g(x)$.

We begin the integration of the momentum advection term focusing on the part corresponding to the Lorentz force, i.e.

$$\begin{aligned}
\frac{q}{m}\int_{-\infty}^{\infty}p^{\hat{0}}F_{\hat{a}}{}^{\hat{b}}p_{\hat{b}}\frac{\partial\mathcal{F}}{\partial p_{\hat{a}}}\,\mathrm{d}p^{\hat{0}}&=\frac{q}{m}\int_{-\infty}^{\infty}p^{\hat{0}}F^{\hat{a}\hat{b}}p_{\hat{b}}\frac{\partial\mathcal{F}}{\partial p^{\hat{a}}}\,\mathrm{d}p^{\hat{0}}\\
&=\frac{q}{m}\int_{-\infty}^{\infty}p^{\hat{0}}\left(F^{\hat{0}\hat{b}}p_{\hat{b}}\frac{\partial\mathcal{F}}{\partial p^{\hat{0}}}+F^{\hat{i}\hat{b}}p_{\hat{b}}\frac{\partial\mathcal{F}}{\partial p^{\hat{i}}}\right)\mathrm{d}p^{\hat{0}}\,,
\end{aligned}\tag{3.46}$$

where we split the sum over $\hat{a}$ into its time and space components. Note the ambiguous notation; the Roman indices $\hat{i}$, $\hat{j}$ and $\hat{k}$ run from one to three.

The integral of the time component yields

$$\begin{aligned}
&-2q\int_{-\infty}^{\infty}f\frac{\partial}{\partial p^{\hat{0}}}\left(p^{\hat{0}}F^{\hat{0}\hat{b}}p_{\hat{b}}\right)H(p^{\hat{0}})\delta(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad=-2qf\int_{-\infty}^{\infty}\left(F^{\hat{0}\hat{b}}p_{\hat{b}}+F^{\hat{0}\hat{b}}\eta_{\hat{b}\hat{0}}\right)H(p^{\hat{0}})\delta(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad=-\frac{q}{p^{\hat{0}}}fF^{\hat{0}\hat{b}}p_{\hat{b}}=-f\frac{q}{p^{\hat{0}}}F^{\hat{0}\hat{i}}p_{\hat{i}}\,.
\end{aligned}\tag{3.47}$$

Note that under the integral the energy $p^{\hat{0}}$ and the momentum components $p^{\hat{i}}$ are independent variables. However, after integrating $p^{\hat{0}}=(p'^2+m^2c^2)^{1/2}$. Moreover, the anti-symmetry of the electromagnetic field tensor implies that $F^{\hat{0}}{}_{\hat{0}}=0$ and $f(t,\boldsymbol{x},\boldsymbol{p}')$ is independent of the energy.

The integral of the space components gives

$$2q\int_{-\infty}^{\infty} p^{\hat{0}}F^{\hat{i}\hat{b}}p_{\hat{b}}\frac{\partial}{\partial p^{\hat{i}}}\left(fH(p^{\hat{0}})\delta(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\right)\mathrm{d}p^{\hat{0}} \tag{3.48}$$
$$= 2q\int_{-\infty}^{\infty}\mathrm{d}p^{\hat{0}}\,p^{\hat{0}}F^{\hat{i}\hat{b}}p_{\hat{b}}fH(p^{\hat{0}})\frac{\partial}{\partial p^{\hat{i}}}\delta(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}+qF^{\hat{i}\hat{b}}p_{\hat{b}}\frac{\partial f}{\partial p^{\hat{i}}}\,,$$

where we applied the product rule and evaluated one of the integrals as before.

To simplify the handling of the derivative of the delta distribution we replace it with a continuously differentiable nascent delta function, e.g. $\delta_\varepsilon(x) \coloneqq \eta(x/\varepsilon) = \exp(-x^2/2\varepsilon^2)/\sqrt{2\pi}\varepsilon$. Restricting our attention to the corresponding integral, we find

$$\begin{aligned}
&2q\lim_{\varepsilon\to 0}\int_{-\infty}^{\infty}\mathrm{d}p^{\hat{0}}\,p^{\hat{0}}F^{\hat{i}\hat{b}}p_{\hat{b}}fH(p^{\hat{0}})\frac{\partial}{\partial p^{\hat{i}}}\delta_\varepsilon(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad= 2qf\lim_{\varepsilon\to 0}\int_{-\infty}^{\infty}F^{\hat{i}\hat{b}}p_{\hat{b}}H(p^{\hat{0}})p_{\hat{i}}\frac{\partial}{\partial p^{\hat{0}}}\delta_\varepsilon(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad= -2qf\lim_{\varepsilon\to 0}\int_{0}^{\infty}\frac{\partial}{\partial p^{\hat{0}}}\left(F^{\hat{i}\hat{b}}p_{\hat{b}}p_{\hat{i}}\right)\delta_\varepsilon(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad= -2qf\int_{0}^{\infty}F^{\hat{i}}{}_{\hat{0}}p_{\hat{i}}\delta(p_{\hat{a}}p^{\hat{a}}-m^2c^2)\,\mathrm{d}p^{\hat{0}}\\
&\quad= -\frac{q}{p^{\hat{0}}}fF^{\hat{i}}{}_{\hat{0}}p_{\hat{i}} = \frac{q}{p^{\hat{0}}}fF^{\hat{0}\hat{i}}p_{\hat{i}}\,.
\end{aligned} \tag{3.49}$$

The first equality is a consequence of

$$\begin{aligned}
&\frac{p_{\hat{i}}}{p^{\hat{0}}}\frac{\partial}{\partial p^{\hat{0}}}\delta_\varepsilon(p^{\hat{0}}p_{\hat{0}}+p_{\hat{j}}p^{\hat{j}}-m^2c^2) = 2p_{\hat{i}}\delta'_\varepsilon(p^{\hat{0}}p_{\hat{0}}+p_{\hat{j}}p^{\hat{j}}-m^2c^2)\\
&= \frac{\partial}{\partial p^{\hat{i}}}\delta_\varepsilon(p^{\hat{0}}p_{\hat{0}}+p_{\hat{j}}p^{\hat{j}}-m^2c^2)\,,
\end{aligned}$$

where δ'_ε denotes the derivative of the nascent delta function. After having integrated by parts we replace the nascent delta function with the delta distribution, see eq. (3.35). Eventually we used that $F^{\hat{i}}{}_{\hat{0}} = F^{\hat{i}\hat{0}} = -F^{\hat{0}\hat{i}}$.

We collect the results of eqs. (3.47), (3.48) and (3.49) to arrive at

$$\begin{aligned}
\frac{q}{m}\int_{-\infty}^{\infty}p^{\hat{0}}F_{\hat{a}}{}^{\hat{b}}p_{\hat{b}}\frac{\partial\mathcal{F}}{\partial p_{\hat{a}}}\,\mathrm{d}p^{\hat{0}} &= -f\frac{q}{p^{\hat{0}}}F^{\hat{0}\hat{i}}p_{\hat{i}}+qF^{\hat{i}\hat{b}}p_{\hat{b}}\frac{\partial f}{\partial p^{\hat{i}}}+\frac{q}{p^{\hat{0}}}fF^{\hat{0}\hat{i}}p_{\hat{i}}\\
&= qF^{\hat{i}\hat{b}}p_{\hat{b}}\frac{\partial f}{\partial p^{\hat{i}}}\\
&= \gamma' mq\,(\boldsymbol{E}'+\boldsymbol{v}'\times\boldsymbol{B}')\cdot\nabla_{p'}f\,.
\end{aligned} \tag{3.50}$$

Notice that the obtained reduction to the space components of the Lorentz force is in agreement with requiring that the particles are on mass shell; the energy changes only if the momentum of the particles change.

We compute in an analogue manner the integral of the fictitious force term. Since we exploited the anti-symmetry of $F^{\hat{a}\hat{b}}$, we define

$$K^{\hat{a}} = \eta^{\hat{a}\hat{c}} K_{\hat{c}} = -\frac{1}{m}\eta^{\hat{a}\hat{d}} \Lambda^{\mu}_{\hat{d}} \frac{\partial \Lambda^{\hat{b}}_{\mu}}{\partial x^{\nu}} \Lambda^{\nu}_{\hat{c}} p_{\hat{b}} p^{\hat{c}} =: \frac{1}{m} T^{\hat{a}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \,, \tag{3.51}$$

and note that eq. (3.43) implies that the newly defined tensor is anti-symmetric in its first two arguments, i.e. $T^{\hat{a}\hat{b}}{}_{\hat{c}} = -T^{\hat{b}\hat{a}}{}_{\hat{c}}$.[19] The integral of the fictitious force term can thus be written as

$$\int_{-\infty}^{\infty} p^{\hat{0}} K^{\hat{a}} \frac{\partial \mathcal{F}}{\partial p^{\hat{a}}} \,\mathrm{d}p^{\hat{0}} = \frac{1}{m} \int_{-\infty}^{\infty} p^{\hat{0}} \left(T^{\hat{0}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial \mathcal{F}}{\partial p^{\hat{0}}} + T^{\hat{i}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial \mathcal{F}}{\partial p^{\hat{i}}} \right) \mathrm{d}p^{\hat{0}} \,,$$

where we again separated the time component from the space components.

The evaluation of the integral of the time component yields

$$\frac{1}{m} \int_{-\infty}^{\infty} p^{\hat{0}} T^{\hat{0}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial \mathcal{F}}{\partial p^{\hat{0}}} \,\mathrm{d}p^{\hat{0}} = -\frac{f}{p^{\hat{0}}} \left(T^{\hat{0}\hat{i}}{}_{\hat{c}\hat{i}} p_{\hat{i}} p^{\hat{c}} + T^{\hat{0}\hat{i}}{}_{\hat{0}\hat{i}} p_{\hat{i}} \right) . \tag{3.52}$$

The integration of the space components gives

$$\begin{aligned} \int_{-\infty}^{\infty} p^{\hat{0}} T^{\hat{i}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial \mathcal{F}}{\partial p^{\hat{i}}} \,\mathrm{d}p^{\hat{0}} &= T^{\hat{i}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial f}{\partial p^{\hat{i}}} - \frac{f}{p^{\hat{0}}} \left(T^{\hat{i}}{}_{\hat{0}\hat{c}} p_{\hat{i}} p^{\hat{c}} + T^{\hat{i}\hat{b}}{}_{\hat{0}\hat{b}} p_{\hat{i}} p_{\hat{b}} \right) \\ &= T^{\hat{i}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial f}{\partial p^{\hat{i}}} + \frac{f}{p^{\hat{0}}} \left(T^{\hat{0}\hat{i}}{}_{\hat{c}\hat{i}} p_{\hat{i}} p^{\hat{c}} + T^{\hat{0}\hat{i}}{}_{\hat{0}\hat{i}} p_{\hat{i}} p_{\hat{0}} \right) , \end{aligned} \tag{3.53}$$

where we used that $T^{\hat{i}}{}_{\hat{0}\hat{c}} = T^{\hat{i}\hat{0}}{}_{\hat{c}} = -T^{\hat{0}\hat{i}}{}_{\hat{c}}$. Additionally, $T^{\hat{i}\hat{j}}{}_{\hat{0}} p_{\hat{i}} p_{\hat{j}} = 0$, also because of the anti-symmetry.

In total, we are left with the space components of the fictitious force, i.e.

$$\int_{-\infty}^{\infty} p^{\hat{0}} K^{\hat{a}} \frac{\partial \mathcal{F}}{\partial p^{\hat{a}}} \,\mathrm{d}p^{\hat{0}} = T^{\hat{i}\hat{b}}{}_{\hat{c}\hat{b}} p_{\hat{b}} p^{\hat{c}} \frac{\partial f}{\partial p^{\hat{i}}} = m\boldsymbol{K}' \cdot \nabla_{p'} f \tag{3.54}$$

Finally, we integrate the collision operator, namely

$$2m \int_{-\infty}^{\infty} p^{\hat{0}} \gamma' \left(\frac{\delta f}{\delta t} \right)_c H(p^{\hat{0}}) \delta(p_{\hat{a}} p^{\hat{a}} - m^2 c^2) = \gamma' m \left(\frac{\delta f}{\delta t} \right)_c = \frac{\gamma' m \nu'}{2} \Delta_{\theta' \varphi'} f \,. \tag{3.55}$$

[19] Dr. Reville remarked that the anti-symmetry is a necessity, because

$$0 = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\tau} p_{\hat{a}} p^{\hat{a}} = p_{\hat{a}} \frac{\mathrm{d}p^{\hat{a}}}{\mathrm{d}\tau} = p_{\hat{a}} p_{\hat{b}} T^{\hat{a}\hat{b}}{}_{\hat{c}} p^{\hat{c}}$$

for arbitrary four-momenta $p^{\hat{a}}$ and the product $p_{\hat{a}} p_{\hat{b}}$ is symmetric; the contraction of a symmetric and an anti-symmetric tensor yields zero.

Putting the spatial advection term of eq. (3.45), the momentum advection term as presented in eqs. (3.50) and (3.54), and the collision operator together gives the *relativistic VFP equation in mixed coordinates*:

$$\Lambda^{\mu}_{\hat{a}} p^{\hat{a}} \frac{\partial f}{\partial x^{\mu}} + \gamma' m \left[\frac{1}{\gamma'} \boldsymbol{K}' + q(\boldsymbol{E}' + \boldsymbol{v}' \times \boldsymbol{B}') \right] \cdot \nabla_{p'} f = \gamma' m \left(\frac{\delta f}{\delta t} \right)_c . \tag{3.56}$$

In the next section we further simplify the relativistic VFP equation. Therefore, we make the assumption that the speed of the background plasma is much less than the speed of light, as for example is the case for a supernova remnant.

3.2.4 TRANSPORT EQUATION

Under this assumptions it is possible to do a Taylor expansion of the Lorentz transformation given in eq. (3.39) in U/c. The expansion requires us to rescale the time components of the involved four-vectors, i.e.

$$\begin{aligned}
&\begin{pmatrix} 1/c & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} E'/c \\ \boldsymbol{p}' \end{pmatrix} \\
&\quad = \begin{pmatrix} 1/c & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \Gamma & -\Gamma \boldsymbol{U}^{\mathsf{T}}/c \\ -\Gamma \boldsymbol{U}/c & \mathbf{1} + (\Gamma - 1)/U^2 \boldsymbol{U} \otimes \boldsymbol{U} \end{pmatrix} \begin{pmatrix} c & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} 1/c & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} E/c \\ \boldsymbol{p} \end{pmatrix} \\
&\quad = \begin{pmatrix} \Gamma & -\Gamma \boldsymbol{U}^{\mathsf{T}}/c^2 \\ -\Gamma \boldsymbol{U} & \mathbf{1} + (\Gamma - 1)/U^2 \boldsymbol{U} \otimes \boldsymbol{U} \end{pmatrix} \begin{pmatrix} E/c^2 \\ \boldsymbol{p} \end{pmatrix} \\
&\quad = \begin{pmatrix} 1 & \mathbf{0}^{\mathsf{T}} \\ -\boldsymbol{U} & \mathbf{1} \end{pmatrix} \begin{pmatrix} E/c^2 \\ \boldsymbol{p} \end{pmatrix} + \mathcal{O}(U/c) ,
\end{aligned} \tag{3.57}$$

where we inserted an identity matrix and only kept zeroth order terms, i.e. we dropped all terms of order U/c. The rescaling was necessary to correctly identify the order of the matrix elements.

Employing this approximation to the first term of the relativistic VFP equation (3.56) yields

$$\begin{aligned}
\Lambda^{\mu}_{\hat{a}} p^{\hat{a}} \frac{\partial f}{\partial x^{\mu}} &= \begin{pmatrix} \Gamma & \Gamma \boldsymbol{U}^{\mathsf{T}}/c \\ \Gamma \boldsymbol{U}/c & \mathbf{1} + (\Gamma - 1)/U^2 \boldsymbol{U} \otimes \boldsymbol{U} \end{pmatrix} \begin{pmatrix} E'/c \\ \boldsymbol{p}' \end{pmatrix} \cdot \begin{pmatrix} 1/c \partial_t \\ \nabla_x \end{pmatrix} f \\
&= \begin{pmatrix} 1 & \mathbf{0}^{\mathsf{T}} \\ \boldsymbol{U} & \mathbf{1} \end{pmatrix} \begin{pmatrix} E'/c^2 \\ \boldsymbol{p}' \end{pmatrix} \cdot \begin{pmatrix} \partial_t \\ \nabla_x \end{pmatrix} f + \mathcal{O}(U/c) \\
&= \gamma' m \left[\partial_t + (\boldsymbol{U} + \boldsymbol{v}') \cdot \nabla_x \right] f + \mathcal{O}(U/c) ,
\end{aligned} \tag{3.58}$$

where we inserted two identity matrices and used the property of the scalar product that for an arbitrary matrix $\boldsymbol{A}$, the scalar product $\boldsymbol{x} \cdot (\boldsymbol{A}\boldsymbol{y})$ equals $(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{x}) \cdot \boldsymbol{y}$.

The fictitious force term can be simplified in a similar manner. Starting with

eq. (3.51)[20], we get

$$\begin{aligned}
&\begin{pmatrix} 1/c & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} K^{\hat{0}} \\ \boldsymbol{K}' \end{pmatrix} \\
&\quad = -\frac{1}{m} \begin{pmatrix} 1/c & \mathbf{0}^\mathsf{T} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \boldsymbol{\Lambda} \left[\left(\boldsymbol{\Lambda}^{-1} \begin{pmatrix} E'/c \\ \boldsymbol{p}' \end{pmatrix} \right) \cdot \begin{pmatrix} 1/c\partial_t \\ \nabla_x \end{pmatrix} \right] \boldsymbol{\Lambda}^{-1} \begin{pmatrix} E'/c \\ \boldsymbol{p}' \end{pmatrix} \\
&\quad = -\frac{1}{m} \begin{pmatrix} 1/c & \mathbf{0}^\mathsf{T} \\ -\boldsymbol{U} & \mathbf{1} \end{pmatrix} \left[\begin{pmatrix} 1 & \mathbf{0}^\mathsf{T} \\ \boldsymbol{U} & \mathbf{1} \end{pmatrix} \begin{pmatrix} E'/c^2 \\ \boldsymbol{p}' \end{pmatrix} \cdot \begin{pmatrix} \partial_t \\ \nabla_x \end{pmatrix} \right] \begin{pmatrix} 1 & \mathbf{0}^\mathsf{T} \\ \boldsymbol{U} & \mathbf{1} \end{pmatrix} \begin{pmatrix} E'/c^2 \\ \boldsymbol{p}' \end{pmatrix} + \mathcal{O}(U/c) \\
&\quad = -\gamma' \begin{pmatrix} 0 \\ \gamma' m \, \mathrm{D}\boldsymbol{U}/\mathrm{D}t + \boldsymbol{p}' \cdot \nabla_x \boldsymbol{U} \end{pmatrix} + \mathcal{O}(U/c) \,. \qquad (3.59)
\end{aligned}$$

Notice that E' and $\boldsymbol{p}'$ are independent variables, i.e. they do not depend on t and $\boldsymbol{x}$. Inserting the approximate expressions of eqs. (3.58) and (3.59) into the relativistic VFP equation (3.56) results in

$$\begin{aligned}
&\frac{\partial f}{\partial t} + (\boldsymbol{U} + \boldsymbol{v}') \cdot \nabla_x f \\
&\quad - \left(\gamma' m \frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t} + \boldsymbol{p}' \cdot \nabla_x \boldsymbol{U} \right) \cdot \nabla_{p'} f + q \boldsymbol{v}' \times \boldsymbol{B}' \cdot \nabla_{p'} f = \left(\frac{\delta f}{\delta t} \right)_c \,,
\end{aligned} \qquad (3.60)$$

where we also assumed that the ideal MHD approximation holds for the background plasma, i.e. $\boldsymbol{E}' = 0$, see eq. (2.84).

Keeping only the zeroth order terms of the Lorentz transformation can be interpreted as the assumption that the background plasma moves slow enough to justify a Galilean transformation, i.e. all relativistic changes related to the transformation into the rest frame of the plasma can be neglected. However, the motion of the particles in the rest frame of the plasma is treated relativistically, whence the appearance of γ'. We thus may dub eq. (3.60) the *semi-relativistic VFP equation*. Notice that this name complements the non-relativistic limit in which the thermal plasma *and* the particles are modelled non-relativistically. The semi-relativistic VFP equation is the starting point of all the chapters containing our research, namely Chapters 4–6.

3.3 THE COSMIC-RAY TRANSPORT EQUATION

With an explicit expression for the collision operator at hand and the reduction of the fully relativistic to the semi-relativistic VFP equation, we are in a position to apply the kinetic equation (3.60) to the parallel shock scenario. The result is the cosmic-ray transport equation.

[20] Note that

$$\eta^{\hat{a}\hat{d}} \Lambda^\mu_{\hat{a}} \frac{\partial \Lambda^{\hat{b}}_\mu}{\partial x^\nu} \Lambda^\nu_{\hat{c}} p_{\hat{b}} p^{\hat{c}} = \Lambda^{\hat{a}}_\mu \frac{\partial \Lambda^\mu_{\hat{b}}}{\partial x^\nu} \Lambda^\nu_{\hat{c}} p^{\hat{b}} p^{\hat{c}} \,.$$

In this scenario the velocity $\boldsymbol{U}$ of the background plasma is determined by the the propagation of the shock wave. In the rest frame of the shock wave, which we choose to be the laboratory frame, i.e. the frame in which $\boldsymbol{x}$ and t are defined, the velocity profile is

$$U(x) = \begin{cases} U_1 = U_\mathrm{s} & \text{for } x < 0 \\ U_2 = U_\mathrm{s}/r & \text{for } x \geq 0 \end{cases}, \tag{3.61}$$

where r, as before, is the compression ratio of the shock, see Fig. (3.1.2).

The discontinuous velocity profile suggests to split the spatial domain of the single particle distribution function f into two parts, namely a function f_1 and f_2 defined in the up- and downstream respectively. This avoids the need to handle the discontinuity at the shock wave, in particular the derivative of $\boldsymbol{U}$ is not continuously differentiable. Notice that because of the mixed coordinate system $f_1(t, \boldsymbol{x}, \boldsymbol{p}_1)$ and $f_2(t, \boldsymbol{x}, \boldsymbol{p}_2)$ are functions of the momentum as defined in the respective rest frames of the background plasma, i.e. $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are the momenta in the rest frame of the upstream and downstream plasma respectively. The splitting of the domain into two parts allows us to solve eq. (3.60) independently in each of them. However, this simplification comes at the cost of the necessity to match the obtained solutions at the shock wave.

In a first step, we focus on finding solutions in the rest frames of the up- and downstream plasmas. The discontinuous velocity profile reduces the semi-relativistic VFP equation to

$$\frac{\partial f_i}{\partial t} + (U_i \boldsymbol{e}_x + \boldsymbol{v}_i) \cdot \nabla_x f_i + q \boldsymbol{v}_i \times \boldsymbol{B} \cdot \nabla_{p_i} f_i = \frac{\nu_i}{2} \Delta_{\theta_i, \varphi_i} f_i \,, \tag{3.62}$$

where the index $i \in \{1, 2\}$ distinguishes between up- and downstream quantities respectively. We also used that the magnetic field in the laboratory frame and the rest frame of the background plasma are about equal, more precisely $\boldsymbol{B}_i' = \boldsymbol{B}_i + \mathcal{O}((U/c)^2)$ and that the $\boldsymbol{B}$-field is the same in the up- and downstream of a parallel shock, i.e. $\boldsymbol{B} = \boldsymbol{B}_1 = \boldsymbol{B}_2$, see eq. (3.4). Moreover, we plugged in the expression that we derived for the collision operator, cf. eq. (3.37).

A key assumption when we derived the energy spectrum of the charged particles in Sec. 3.1.2 was that their distribution function is isotropic in the rest frame of the background plasma, i.e. in the up- and downstream rest frame. This assumption in combination with the motion of the shock wave ineluctably leads to a net particle flux through the shock, i.e. more particles leave the upstream than enter it from downstream.[21] A steady-state is established because charged particles are assumed to be constantly injected. Notice that in Sec. 3.1.2 we did *not* use mixed coordinates whereas in eq. (3.62) we do. In effect, the shock front is located at $x = 0$ and does not move. As particles that leave the upstream leave

[21] This becomes apparent when the net number flux through the shock is computed, i.e. when the limits of the integral in eq. (3.6) are changed to [-1,1] which results in $\mathrm{d}j_s^x/\mathrm{d}p = 4\pi f(p) p^2 U_1$. Also see the corresponding formula for the imaginary surface in the downstream.

it, no matter which coordinate system we choose, it is a necessity to recover the particle flux in the mixed coordinate setting. The fact that the relative motion between the particles and the shock went into the spatial advection term, i.e. the VFP equation contains the additional term $U_i \partial_x f$, may allow for an anisotropic distribution function as a solution to it. The anisotropy could account for the necessary number flux. On top of that, the computation of the net flux through the shock shows that $\mathrm{d}j_s^x/\mathrm{d}p \propto f(p)U_1$, where $f(p)$ is the isotropic distribution function in the upstream rest frame. This implies that the anisotropic part must be proportional to U_1/v_1 to yield a comparable flux. Considering that we are interested in relativistic particles, i.e. $v_1 \approx c$, it may be that the anisotropic part can be regarded as a comparatively small deviation from an isotropic particle distribution.

These arguments motivate an ansatz for the single particle distribution function f that splits it into an isotropic and an anisotropic part, i.e.

$$f_i(t, \boldsymbol{x}, \boldsymbol{p}_i) = f_i^0(t, \boldsymbol{x}, p_i) + \delta f_i(t, \boldsymbol{x}, \boldsymbol{p}_i) = f_i^0(t, \boldsymbol{x}, p_i) + \boldsymbol{e}_{p_i} \cdot \boldsymbol{f}_i^1(t, \boldsymbol{x}, p_i)\,, \quad (3.63)$$

where we separated out the angular dependence of f; notice that we keep using spherical coordinates for the momentum variables, i.e. $\boldsymbol{p}_i = p_i \boldsymbol{e}_{p_i}$. We, additionally, presuppose that the anisotropy δf_i is of the order $\mathcal{O}(\lambda_i/\mathcal{L} f_i^0)$ with $\lambda_i/\mathcal{L} \ll 1$, where $\lambda_i \coloneqq v_i/\nu_i$ is the mean free path of the particles and $\mathcal{L}$ is the characteristic length on which the isotropic part f_i^0 changes, i.e. $\partial f_i^0/\partial x \sim f_i^0/\mathcal{L}$. The ansatz together with the above presupposition, which is typical for diffusion processes, is called the *diffusion approximation*. Furthermore, the anisotropy δf_i is referred to as the *dipole anisotropy*.

We remark that eq. (3.62) informs us that the evolution of the distribution function f is determined by the particle velocity v_i, the velocity of the plasma U_i and the scattering frequency ν_i; as pointed out in the introduction of Sec. (3.1.2) the magnetic field $\boldsymbol{B}$ is irrelevant for the acceleration at a *parallel* shock. We can combine these three parameters to obtain a possible candidate for the characteristic length, namely $\mathcal{L} = v_i^2/(U_i \nu_i)$. This choice will prove to be justified in retrospect. Taking it for granted, implies that $\delta f_i \in \mathcal{O}(\lambda_i/\mathcal{L} f_0^i) = \mathcal{O}(U_i/v_i f_0^i)$.

We furthermore would like to point out that

$$\varrho_i \coloneqq \int_{\mathbb{R}^3} f_i \,\mathrm{d}^3 p = \int_0^\infty \int_0^\pi \int_0^{2\pi} f_i p^2 \,\mathrm{d}p \,\mathrm{d}\Omega = 4\pi \int_0^\infty f_i^0 p^2 \,\mathrm{d}p\,,$$

i.e. the integral of the anisotropic part over the sphere vanishes. Note that we also defined the configuration-space density ϱ of the energetic and charged particles. This means that the 'particle content' is in the isotropic part only and if there is an anisotropy in the distribution, e.g. if there are more particles moving to the right than to the left, then this is realised by subtracting particles that move to the left from the isotropic part and add them to the particles that move right. For the case described in the previous sentence, the distribution function would be $f_i = f_i^0 + \cos\theta_i f_{i,1}^1$ where $f_{i,1}^1$ is the first component of $\boldsymbol{f}^1$ and

assuming that $f_{i,1}^1 > 0$, the anisotropy does exactly this, i.e. it removes particles from the isotropic part that move to the left, because $\cos\theta_i < 0$ for $\theta_i > \pi/2$, and adds them to the particles that move right. It is in this specific sense that the anisotropy in the diffusion approximation describes a deviation from an isotropic distribution. (Kirk, 1994, Sec. 4.2)

Plugging the diffusion approximation into the VFP equation (3.62) that models the parallel shock scenario gives

$$\begin{aligned}\frac{\partial f_i^0}{\partial t} + \boldsymbol{e}_{p_i} \cdot \frac{\partial \boldsymbol{f}_i^1}{\partial t} + U_i \frac{\partial f_i^0}{\partial x} + U_i \boldsymbol{e}_{p_i} \cdot \frac{\partial \boldsymbol{f}_i^1}{\partial x} \\ + \upsilon_i \boldsymbol{e}_{p_i} \cdot \nabla_x f_i^0 + \upsilon_i \boldsymbol{e}_{p_i} \cdot \nabla_x (\boldsymbol{e}_{p_i} \cdot \boldsymbol{f}_i^1) + (\boldsymbol{e}_{p_i} \times \boldsymbol{\omega}_g) \cdot \boldsymbol{f}_i^1 = -\nu_i \boldsymbol{e}_{p_i} \cdot \boldsymbol{f}_i^1 \,,\end{aligned} \tag{3.64}$$

where we introduced the vector $\boldsymbol{\omega}_g \coloneqq q\boldsymbol{B}/\gamma m$ and used that

$$\begin{aligned}\nabla_{p_i} f_i &= \boldsymbol{e}_{p_i} \frac{\partial f_i^0}{\partial p} + \frac{\boldsymbol{f}_i^1}{p_i} + \boldsymbol{e}_{p_i} \boldsymbol{p}_i \cdot \frac{\partial}{\partial p} \left(\frac{\boldsymbol{f}_i^1}{p_i} \right) \quad \text{and} \\ \Delta_{\theta_i, \varphi_i} f_i &= -2 \boldsymbol{e}_{p_i} \cdot \boldsymbol{f}_i^1 \,.\end{aligned}$$

The first equality is a result of expressing the Nabla operator in spherical coordinates and the fact that $\boldsymbol{p}_i / p_i = \boldsymbol{e}_{p_i}$.

To arrive at an equation for the isotropic part f_i^0, we exploit that the anisotropy vanishes when we integrate over the sphere. We, thus, integrate eq. (3.64) over $\mathrm{d}\Omega$. This results in

$$\frac{\partial f_i^0}{\partial t} + U_i \frac{\partial f_i^0}{\partial x} + \frac{\upsilon_i}{3} \nabla_x \cdot \boldsymbol{f}_i^1 = 0 \,, \tag{3.65}$$

where we divided by 4π and used that

$$\int_0^\pi \int_0^{2\pi} e_{p_i,k} \, \mathrm{d}\Omega = 0 \quad \text{and} \quad \int_0^\pi \int_0^{2\pi} e_{p_i,k} e_{p_i,l} \, \mathrm{d}\Omega = 4\pi \frac{\delta_{kl}}{3} \,. \tag{3.66}$$

The equation for the kth component of the anisotropy $\boldsymbol{f}^1$ is obtained through a multiplication of eq. (3.64) with $e_{p_i,k}$ and a subsequent integration over $\mathrm{d}\Omega$. The equations for the three components are

$$\frac{\partial f_{i,1}^1}{\partial t} + U_i \frac{\partial f_{i,1}^1}{\partial x} + \nu_i f_{i,1}^1 = -\upsilon_i \frac{\partial f_i^0}{\partial x} \,, \tag{3.67}$$

$$\frac{\partial f_{i,2}^1}{\partial t} + U_i \frac{\partial f_{i,2}^1}{\partial x} - \omega_g f_{i,3}^1 + \nu_i f_{i,2}^1 = -\upsilon_i \frac{\partial f_i^0}{\partial y} \,, \tag{3.68}$$

$$\frac{\partial f_{i,3}^1}{\partial t} + U_i \frac{\partial f_{i,3}^1}{\partial x} + \omega_g f_{i,2}^1 + \nu_i f_{i,3}^1 = -\upsilon_i \frac{\partial f_i^0}{\partial z} \,. \tag{3.69}$$

We again took advantage of the integrals in eq. (3.66) and, additionally, we used that an integral of a product of three components of $\boldsymbol{e}_{p_i}$ over the sphere vanishes. Note that $\boldsymbol{B} = B_0 \boldsymbol{e}_x$. Eqs. (3.67)–(3.69) are arranged such that the source of

the anisotropy is emphasised, namely gradients in the isotropic part of the distribution function. If we assume that the particles are injected isotropically and homogeneously in the y–z plane, the components $f^1_{i,2}$ and $f^1_{i,3}$ are identically zero; the equation (3.65) for the isotropic part tells us that no gradients in the y- and/or z-direction will evolve.

Dividing the equation (3.67) for the first component $f^1_{i,1}$ of the anistropy by v_i, reveals that the term $U_i/v_i \partial f^1_{i,1}/\partial x$ is of the order $\mathcal{O}((U_i/v_i)(\lambda_i/\mathcal{L})\partial f^0_i/\partial x)$ and can be dropped, because we assumed that U_i/v_i and $\lambda_i/\mathcal{L}$ are much smaller than one. Let the characteristic time scale on which the isotropic part of the distribution function f^0_i changes be τ and if $\tau \gg 1/\nu$, i.e. if the acceleration of particles takes much longer than it takes to isotropise the distribution function, then we may drop the time derivative of $f^1_{i,1}$ in eq. (3.67) as well.[22] Hence, the anisotropy is

$$f^1_{i,1} = -\frac{v_i}{\nu_i}\frac{\partial f_{i,0}}{\partial x} = -\lambda_i \frac{\partial f_{i,0}}{\partial x}\,, \tag{3.70}$$

cf. Drury (1983, eq 2.35). Plugging the anisotropy into the equation (3.65) for the isotropic part results in the *cosmic-ray transport equation* for a parallel shock, namely

$$\frac{\partial f^0_i}{\partial t} + U_i \frac{\partial f^0_i}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\lambda_i v_i}{3}\frac{\partial f^0_i}{\partial x}\right) =: \frac{\partial}{\partial x}\left(\kappa_i \frac{\partial f^0_i}{\partial x}\right), \tag{3.71}$$

where we defined the *diffusion coefficient* $\kappa_i = \lambda_i v_i/3$.

3.3.1 SOLUTION OF THE COSMIC-RAY TRANSPORT EQUATION

We solve the cosmic-ray transport equation (3.71) for a parallel shock assuming that a steady state is reached, i.e. assuming that the single particle distribution function f does not change anymore with time. This is in agreement with Sec. 3.1.2 in which we computed the steady-state spectrum of the particles.

The steady-state cosmic-ray transport equation is an ordinary differential equation in x, namely

$$U_i \frac{\partial f^0_i}{\partial x} = \frac{\partial}{\partial x}\left(\kappa_i \frac{\partial f^0_i}{\partial x}\right), \tag{3.72}$$

and physically it can be interpreted as a flux of particles due to advection that is balanced by a diffusive flux.

We allow for the possibility that the scattering frequency ν_i and, thus, the diffusion coefficient κ_i, may depend on x and the magnitude of the momentum

[22] If $\tau \gg 1/\nu_i$, then f^0_i in eq. (3.67) can be assumed to be independent of time, because it evolves comparatively slow. Neglecting $U_i \partial f^1_{i,1}/\partial x$, the solution of eq. (3.67) then is

$$f^1_{i,1} = -v_i/\nu_i \frac{\partial f^0_i}{\partial x}\left(1 - \mathrm{e}^{-t\nu_i}\right),$$

which for $1/\nu_i \ll t$ gives eq. (3.70).

p. Integrating eq. (3.72) over x yields

$$U_i f_i^0 = \kappa_i(x, p_i)\frac{\partial f_i^0}{\partial x} + \tilde{A}_i(p_i)\,, \tag{3.73}$$

where $\tilde{A}_i$ is the constant of integration. The homogeneous solution of the above ODE is $f_{i,h}^0 = B_i(p_i)\exp(\int U_i/\kappa_i\,\mathrm{d}x)$. We vary the constant B_i to construct a particular solution, namely $f_{i,p}^0 = B_i(x, p_i)\exp(\int U_i/\kappa_i\,\mathrm{d}x)$. Plugging the particular solution into eq. (3.73) and integrating it over x determines the constant B_i, i.e.

$$\begin{aligned} B_i(x, p_i) &= -\tilde{A}_i(p_i)\int \exp\left(-\int \frac{U_i}{\kappa_i(x, p_i)}\,\mathrm{d}x\right)/\kappa_i(x, p_i)\,\mathrm{d}x \\ &= A_i(p_i)\exp\left(-\int \frac{U_i}{\kappa_i(x, p_i)}\,\mathrm{d}x\right) + \tilde{C}_i(p_i)\,, \end{aligned}$$

where we set $A_i \coloneqq \tilde{A}_i/U_i$ and integrated by substitution using that $H(x) \coloneqq \int U_i/\kappa_i\,\mathrm{d}x$ has the derivative $H'(x) = U_i/\kappa_i$.

The general solution of the cosmic-ray transport equation is the sum of the homogeneous and the particular solution, namely

$$f_i^0(x, p_i) = A_i(p_i) + C_i(p_i)\exp\left(\int \frac{U_i}{\kappa_i(x, p_i)}\,\mathrm{d}x\right)\,, \tag{3.74}$$

where $C_i \coloneqq B_i + \tilde{C}_i$. (Drury, 1983; Kirk, 1994, eq. 2.33 and eq. 81 respectively)

The distribution function f is a number density and so is its isotropic part, we thus require that it remains finite downstream, i.e. $f_2^0(x, p_2) < \infty$ when $x \to \infty$. If the x dependence of the diffusion coefficient κ_i is such that

$$\lim_{x\to\pm\infty}\int 1/\kappa_i\,\mathrm{d}x \to \pm\infty\,,$$

then the requirement that f_2^0 is bounded implies that $C_2(p_2)$ must equal zero and thus the distribution function is homogeneous in the downstream. The corresponding phase-space density $C_1(p_1)$ in the upstream decays exponentially with decreasing distance from the shock. Moreover, we can think of $A_1(p_1)$ as an isotropic population of particles that 'inhabits' the upstream irrespective of the shock wave. Keeping in mind that we are not interested in how pre-existing particles are accelerated but in how constantly injected particles change their energy, we may set $A_1(p_1) = 0$. Taking the outlined restrictions into consideration, the physically relevant solutions of the cosmic-ray transport equation are

$$\begin{aligned} f_1^0(x, p_1) &= C_1(p_1)\exp\left(\int_0^x \frac{U_1}{\kappa_1(x, p_1)}\,\mathrm{d}x\right) && \text{for } x < 0 \\ f_2^0(x, p_2) &= A_2(p_2) && \text{for } x \geq 0\,. \end{aligned} \tag{3.75}$$

Notice that the limits of integration ensure that $f_1^0(0, p_1) = C_1(p_1)$ at the shock.

We emphasise once more that the isotropic part of the particle distribution function in the downstream is independent of x. An immediate consequence is that there is no anisotropy in the downstream, i.e. $f^1_{2,1}(x, p_2) = -\lambda_2 \partial f^0_2/\partial x = 0$, see eq. (3.70), whereas in the upstream a dipole anisotropy exists, namely $f^1_{1,1}(x, p_1) = 3U_1/v_1 f^0_1(x, p_1)$.[23] Note that the dipole anisotropy is of the order $\mathcal{O}(U_1/v_1 f^0_1)$, as speculated when we introduced the diffusion approximation. At this point we pick a up a loose thread, namely our choice for the characteristic length $\mathcal{L}$ scale on which the isotropic part of distribution function varies. The exponential decay of the isotropic part suggests to use an e-folding, i.e.

$$\begin{aligned} -1 \overset{!}{=} \int_0^x \frac{U_i}{\kappa_i(x, p_i)}\,\mathrm{d}x = \frac{3U_1}{v_1^2}\int_0^x \nu_i(x, p_i)\,\mathrm{d}x = \frac{3U_1}{v_1^2}\langle\nu_i(p_i)\rangle x \\ \Longrightarrow\ \mathcal{L} = \frac{v_1^2}{3U_1\langle\nu_i(p_i)\rangle}\,, \end{aligned} \tag{3.76}$$

which justifies our choice and where we applied the mean value theorem for definite integrals, i.e. $\int_0^x \nu_i\,\mathrm{d}x' = \langle\nu_i\rangle x$.

With eqs. (3.63), (3.70) and (3.75), we are ready to write down the complete single particle distribution function in the up- and downstream of the shock wave, namely

$$\begin{aligned} f_1(x, p_1) &= C_1(p_1)\left(1 - 3\cos\theta_1\frac{U_1}{v_1}\right)\exp\left(\int_0^x \frac{U_1}{\kappa_1(x, p_1)}\,\mathrm{d}x\right) && \text{for } x < 0 \\ f_2(x, p_2) &= A_2(p_2) && \text{for } x \geq 0\,. \end{aligned} \tag{3.77}$$

We expect the single particle distribution function to be continuous at a parallel shock front, because the $\boldsymbol{B}$-field is unchanged and the particles continue to follow their trajectories undisturbed. However, we use a mixed coordinate system and the momentum variable in the upstream is defined with respect U_1 whereas it is defined with respect to U_2 in the downstream. This implies that the distribution functions f_1 and f_2 are discontinuous at the shock.

For a continuous distribution function f, we thus need a frame in which the definition of the momentum variable does not change across the shock: We follow Drury (1983, p. 983) in transforming the momentum variables in up-*and* downstream to the shock rest frame. Using the transformation given in

[23] In footnote 21 we computed the net number flux through the shock in the upstream rest frame. It must equal the number flux induced by the anisotropy in the mixed-coordinate system and that is the case, because

$$\frac{\mathrm{d}j^x_s}{\mathrm{d}p_1} = \int v_1\cos\theta_1 f_1 p_1^2\,\mathrm{d}\Omega = \int v_1\cos^2\theta_1 f^1_{1,1}\,\mathrm{d}\Omega = \int 3U_1\cos^2\theta_1 f^0_1 p_1^2\,\mathrm{d}\Omega = 4\pi f^0_1 p_1^2 U_1\,.$$

eq. (3.57) yields for the magnitude of the momentum

$$\begin{aligned} p_i = \sqrt{\boldsymbol{p}_i \cdot \boldsymbol{p}_i} &= \left[(\boldsymbol{p} - \gamma m U_i \boldsymbol{e}_x) \cdot (\boldsymbol{p} - \gamma m U_i \boldsymbol{e}_x)\right]^{1/2} \\ &= p\left(1 - 2U_i/v\cos\theta + (U_i/v)^2\right)^{1/2} \\ &= p(1 - U_i/v\cos\theta) + \mathcal{O}\left((U_i/v)^2\right), \end{aligned} \tag{3.78}$$

where $p = \gamma m v$, without the subscript i, is the magnitude of the momentum in the shock rest frame. The last line shows the truncated Taylor series of p_i in U_i/v. The transformation of the momentum variables also changes the angles θ_i, namely

$$\begin{aligned} \cos\theta_i = \boldsymbol{p}_i \cdot \boldsymbol{e}_x / p_i &= (\boldsymbol{p} - \gamma m U_i \boldsymbol{e}_x) \cdot \boldsymbol{e}_x / p_i \\ &= \frac{\cos\theta - U_i/v}{(1 - 2U_i/v\cos\theta + (U_i/v)^2)^{1/2}} \\ &= \cos\theta - (1 - \cos^2\theta)U_i/v + \mathcal{O}\left((U_i/v)^2\right). \end{aligned} \tag{3.79}$$

To match the distribution functions $f_1(x, p_1, \theta_1)$ and $f_2(p_2)$ in the up- and downstream at the shock wave, we use the general expression for the distribution functions given in eq. (3.63) and the formulae of eq. (3.78) and (3.79) to express p_i and θ_i in terms of p and θ. We then Taylor expand in U_i/v and evaluate the distribution functions at $x = 0$. This gives

$$\begin{aligned} f_1(0, p, \theta) &= C_1\left(p_1(p,\theta)\right) + \cos\left(\theta_1(\theta)\right) f^1_{1,1}\left(0, p_1(p,\theta)\right) \\ &= C_1(p) - p\cos\theta \frac{\partial C_1}{\partial p}\frac{U_1}{v} \\ &\quad + \left(\cos\theta - (1-\cos^2\theta)\frac{U_1}{v}\right)\left(f^1_{1,1}(0,p) - p\cos\theta\frac{\partial f^1_{1,1}}{\partial p}\frac{U_1}{v}\right) + \mathcal{O}\left(\frac{U_1^2}{v^2}\right) \\ &= C_1(p) - \cos\theta\left(p\frac{\partial C_1}{\partial p}\frac{U_1}{v} - f^1_{1,1}(0,p)\right) + \mathcal{O}\left(\frac{U_1^2}{v^2}\right) \\ &= C_1(p) - \cos\theta\left(p\frac{\partial C_1}{\partial p}\frac{U_1}{v} + \lambda_1(0,p)\frac{\partial f_1^0}{\partial x}(0,p)\right) + \mathcal{O}\left(\frac{U_1^2}{v^2}\right) \end{aligned} \tag{3.80}$$

where we took advantage of the fact that $f^1_{1,1}(x, p_1)$ is of the order $\mathcal{O}(U_1/v_1 f_1^0)$ and used eq. (3.70). An analogue computation for the downstream distribution function yields

$$f_2(p) = A_2(p) - p\cos\theta\frac{\partial A_2}{\partial p}\frac{U_2}{v} + \mathcal{O}\left(\frac{U_2^2}{v^2}\right). \tag{3.81}$$

Because of the continuity of the distribution function in the rest frame of the shock, $f(0,p) = f_1(0,p) = f_2(p)$ and equating the coefficients leads to the following two matching conditions

$$A(p) \coloneqq C_1(p) = A_2(p), \tag{3.82}$$

$$p\frac{\partial A(p)}{\partial p}\frac{U_1}{v} + \lambda_1(0,p)\frac{\partial f_1^0}{\partial x} = p\frac{\partial A(p)}{\partial p}\frac{U_2}{v}. \tag{3.83}$$

where we introduced the phase-space number density $A(p)$ and note that it is the isotropic part of the distribution function. Equation (3.82) can be interpreted as a confirmation of the requirement that the number of particles with a specific momentum must not change, i.e. the isotropic part of the distribution function must be continuous, across the shock if the momentum variables are defined in the same reference frame. Equation (3.83) states that the particle number flux is continuous across the shock. This flux is called the *particle streaming*. (Drury, 1983, cf. eq. 2.39 and 2.40)

If we plug into eq. (3.83) the value of the derivative of the isotropic part f_1^0, as given in eq. (3.75), at $x = 0$, then eq. (3.83) becomes an ordinary differential equation that determines $A(p)$, namely

$$\frac{p}{3}(U_1 - U_2)\frac{\mathrm{d}A}{\mathrm{d}p} + U_1 A(p) = 0\,. \tag{3.84}$$

Separation of variables yields

$$A(p) = C p^{-3U_1/(U_1-U_2)} = C p^{-3r/(r-1)}\,. \tag{3.85}$$

Drury (1983, Sec. 3.2) solves the time-dependent cosmic-ray transport equation (3.71) for a constant, isotropic and monoenergetic injection of particles with momentum p_{inj} at the shock front. He finds that in the limit $t \to \infty$ the value of the isotropic part of the distribution function at the injection momentum is $A(p_{\mathrm{inj}}) = 3Q/(U_1 - U_2)p_{\mathrm{inj}}$, where Q is the constant rate at which particles are injected, see Drury (1983, eq. 3.25).[24] This result provides us with a initial condition that determines the constant of integration C, i.e.

$$A(p) = \frac{Q}{U_1 p_{\mathrm{inj}}}\frac{3r}{r-1}\left(\frac{p}{p_{\mathrm{inj}}}\right)^{-3r/r-1} \tag{3.86}$$

The energy spectrum of the charged particles is given by

$$\frac{\mathrm{d}N}{\mathrm{d}p} = \int f(0,p)p^2\,\mathrm{d}\Omega = 4\pi A(p)p^2 = \frac{4\pi Q}{U_1 p_{\mathrm{inj}}}\frac{3r}{(r-1)}\left(\frac{p}{p_{\mathrm{inj}}}\right)^{-3r/(r-1)+2}\,. \tag{3.87}$$

Assuming a compression ratio of $r = 4$ and using that for highly relativistic particles $E \approx pc$, we recover what we derived in Sec. 3.1.2, namely that the spectral index of the energetic particles at the shock is $\mathrm{d}N/\mathrm{d}E \propto E^{-2}$, see eq. (3.13).

[24] Eq. 3.25 differs by the factor $1/p_{\mathrm{inj}}$. There is a typo in eq. 3.22: the factor $1/p_{\mathrm{inj}}$ is missing, the solution to eq. 3.21 has to include it. Also see Kirk (1994, eq. 126).

CHAPTER 4

THE CARTESIAN TENSOR AND THE SPHERICAL HARMONIC EXPANSION

The acceleration of particles at a parallel shock is an example of the transport of charged particles in tenuous astrophysical plasmas. The semi-relativistic VFP equation (3.60) is a much more general transport equation and such applicable to a wide range of physical scenarios. However, the fact that the single particle distribution function f depends on six variables plus time, namely $\boldsymbol{x}$ and $\boldsymbol{p}$, renders a naive application of numerical algorithms to solve the PDE computationally infeasible. The dimensionality of the distribution function and, thus, the computational cost can be reduced when we separate out the angular dependence of the momentum variable as we did in the diffusion approximation, see eq. (3.63). Concretely, we are looking for an expansion of f that has the following general form

$$f(t, \boldsymbol{x}, |\boldsymbol{p}|, \theta, \varphi) = \sum_{i=0}^{\infty} c_i(t, \boldsymbol{x}, |\boldsymbol{p}|) g_i(\theta, \varphi) ,$$

where the functions g_i are known. From this perspective the diffusion approximation is a series of the above kind that is truncated after the first four terms with the additional assumption that the expansion coefficient c_i with $i > 0$ are small compared to c_0. If the semi-relativistic VFP equation is used to model more generic scenarios, it is necessary to go beyond the diffusion approximation, i.e. to include more terms of the series and to not restrict the magnitude of the expansion coefficients c_i. Notice that replacing f with the suggested expansion implies that we have to numerically compute multiple expansion coefficients c_i instead of f, though they only depend on four variables plus time.

Two expansions have been widely used, firstly, the Cartesian tensor expansion and, secondly, the spherical harmonic expansion. Both are known from the multipole expansion of an electrostatic (or gravitational) potential. This is also the reason for calling the anisotropy of the distribution function in the diffusion approximation the dipole anisotropy. Generalising the diffusion approximation in this sense gives

$$f(t, \boldsymbol{x}, \boldsymbol{p}) = \sum_{l=0}^{\infty} F^{(l)}_{i_1 \cdots i_l}(t, \boldsymbol{x}, |\boldsymbol{p}|) \frac{p^{i_1} \cdots p^{i_l}}{|\boldsymbol{p}|^l} , \tag{4.1}$$

where p^1, p^2 and p^3 are the components of $\boldsymbol{p}$, cf. Epperlein and Haines (1986), Johnston (1960), Shkarofsky et al. (1966) and Thomas et al. (2012). The objects $F^{(l)}$ are called *Cartesian tensors* and their components $F^{(l)}_{i_1\cdots i_l}$ are the expansion coefficients. Note that we are implicitly summing over all repeated indices and that the tensor $F^{(1)}_i$ is the dipole anisotropy and $F^{(2)}_{ij}$ may be referred to as the *quadrupole anisotropy*. Moreover, the product of the components of $\boldsymbol{p}$ can be considered a (Cartesian) tensor, namely $\otimes\boldsymbol{p}^{(l),i_1\cdots i_l} \coloneqq p^{i_1}\cdots p^{i_l}$. Hence the summation over repeated indices may be seen as a contraction of two tensors (formerly called a Cartesian tensor scalar product, see Johnston, 1960). Bearing in mind that we use spherical coordinates in momentum space, dividing this tensor by $|\boldsymbol{p}|^l$ gives an expression that only depends on θ and φ.[1]

The spherical harmonic expansion of f is

$$f(t,\boldsymbol{x},\boldsymbol{p}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_l^m(t,\boldsymbol{x},|\boldsymbol{p}|)Y_l^m(\theta,\varphi)\,, \tag{4.2}$$

cf., for example, Allis (1956), Bell et al. (2006), Reville and Bell (2013) and Tzoufras et al. (2013). The function Y_l^m is a *(Laplace's) spherical harmonic of degree l and order m*, i.e.

$$Y_l^m(\theta,\varphi) \coloneqq \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{\mathrm{i}m\varphi} \coloneqq N_l^m P_l^m(\cos\theta)e^{\mathrm{i}m\varphi}\,. \tag{4.3}$$

N_l^m is the normalisation and the functions P_l^m are the *associated Legendre polynomials*, see Jackson (1998, p. 108, eq. 3.53).

Both expansions can be substituted into the semi-relativistic VFP equation (3.60) to derive a system of equations for the expansion coefficients $F^{(l)}_{i_1\cdots i_l}$ and f_l^m. In the next chapter, we introduce a novel method to do this for the spherical harmonic expansion. In this chapter we explore the relation between the Cartesian tensor and spherical harmonic expansion and, on the way, we acquaint ourselves with the vector space structure of the set of spherical harmonics. We note that a large part of the chapter's content is a re-print of Schween and Reville (2022b).

Because both expansions of f represent the same distribution function, there must exist a relation between their expansion coefficients $F^{(l)}_{i_1\cdots i_l}$ and f_l^m, see Courant and Hilbert (1989). Johnston (1960) investigated this relation and derived a way to express the components of the Cartesian tensors $F^{(l)}_{i_1\cdots i_l}$ as a sum of the f_l^ms for values of $l \leq 4$. We revisit the problem of relating $F^{(l)}_{i_1\cdots i_l}$ and f_l^m and derive general formulae for converting between them (in both directions).

[1] In this chapter we use the more explicit $|\boldsymbol{p}|$ instead of p to denote the magnitude of the momentum, because working with expressions of the form $p^{i_1}\cdots p^{i_l}/|\boldsymbol{p}|^l$ may make it more difficult to discriminate between the cases in which we refer to a particular component of $\boldsymbol{p}$ and the ones in which we mean a power of $|\boldsymbol{p}|$.

At the heart of our derivation are two ideas: Firstly, the problem is also familiar in multipole expansions of an electrostatic (or gravitational) potential. In this context the coefficients of the Cartesian tensor expansion of the potential are called (Cartesian) multipole moments, i.e. monopole, dipole, quadrupole and high-order moments, and the ones of the spherical harmonic expansion are called spherical multipole moments. Secondly, we interpret the spherical harmonic expansion, as presented in eq. (4.2), as a linear combination of basis vectors that represents a function f. The function then is an element of the vector space of spherical harmonics

$$\mathcal{S} \coloneqq \operatorname{span}\{Y_0^0, Y_1^1, Y_1^0, Y_1^{-1}, \dots\}.$$

This perspective suggests that the Cartesian tensor expansion is a representation of f in a different basis of $\mathcal{S}$. This in turn implies that the relation between the expansion coefficients can be given as a basis transformation.

Both ideas gain plausibility in a direct comparison of the respective expansions, i.e.

$$4\pi\epsilon_0\phi(\boldsymbol{x}) = \sum_{l=0}^{\infty} \frac{1}{l!} Q^{(l)}_{i_1\cdots i_l} \frac{x^{i_1}\cdots x^{i_l}}{|\boldsymbol{x}|^{2l+1}} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{2l+1}\frac{q_l^m}{|\boldsymbol{x}|^{2l+1}} |\boldsymbol{x}|^l Y_l^m(\theta,\varphi)$$

$$f(t,\boldsymbol{x},\boldsymbol{p}) = \sum_{l=0}^{\infty} F^{(l)}_{i_1\cdots i_l}(t,\boldsymbol{x},|\boldsymbol{p}|)\frac{p^{i_1}\cdots p^{i_l}}{|\boldsymbol{p}|^l} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_l^m(t,\boldsymbol{x},|\boldsymbol{p}|) Y_l^m(\theta,\varphi).$$

If we set $|\boldsymbol{x}| = 1$ and include the factor $1/l!$ and $4\pi/(2l+1)$ in the definition of $Q^{(l)}$ and q_l^m respectively, then the expansions of the potential ϕ can be interpreted as linear combinations of functions that are defined on the sphere, in particular, $4\pi\epsilon_0\phi = \sum \tilde{q}_l^m Y_l^m$, where $\tilde{q}_l^m \coloneqq 4\pi/(2l+1) q_l^m$. The same is true of the expansions of the single particle distribution function f, if we evaluate f at a fix but arbitrary triple $t_0, \boldsymbol{x}_0$ and $|\boldsymbol{p}|_0$. This implies, firstly, that the relation between the Cartesian and spherical multipole moments is the same as the relation between the expansion coefficients $F^{(l)}_{i_1\cdots i_l}$ and f_l^m and, secondly, varying the magnitude of $\boldsymbol{x}$ or the values of the triple $t_0, \boldsymbol{x}_0$ and $|\boldsymbol{p}|_0$ changes the function ϕ or f respectively, but it does not change the relation between the coefficients of the linear combinations representing them.

In this chapter we derive a way to convert between the Cartesian multipole moments and the spherical multipole moments and, subsequently, apply it to convert between $F^{(l)}_{i_1\cdots i_l}$ and f_l^m. While the equivalence of the two multipole moments is long known, e.g. Courant and Hilbert (1989, p. 517), systematic methods that relate them are not readily found in the literature.

Notice that we decided to work with the multipole moments, because it is possible to define them explicitly whereas the expansion coefficients $F^{(l)}_{i_1\cdots i_l}$ and f_l^m are solutions to systems of PDEs. Furthermore, multipole moments play an important role in multiple branches of physics, such as general relativity (Thorne, 1980), molecular physics (Cipriani & Silvi, 1982), plasma physics (Epperlein

& Haines, 1986; Johnston, 1960) and other areas like computational physics (Ludwig, 1991) and we hope that this fact broadens the applicability of the derived formulae for the relation between the expansion coefficients of the Cartesian tensor and spherical harmonic expansion.

We implemented the new way to convert between the Cartesian and spherical multipole moments in a free and open-source command-line tool called `multipole-conv` (Schween & Reville, 2022a).

4.1 DEFINITION OF THE MULTIPOLE MOMENTS

In electrodynamics (or equivalently, on exchanging constants, in gravitational theory), the multipole expansion is an expansion of the solution to Poisson's equation, namely

$$4\pi\epsilon_0\phi(\boldsymbol{x}) = \int \frac{\rho(\boldsymbol{x})}{|\boldsymbol{x}-\boldsymbol{x}'|}\,\mathrm{d}^3x' \,. \tag{4.4}$$

Either the denominator of the integrand is Taylor expanded in x'/x or it is interpreted as the generating function of the Legendre Polynomials. The former leads to the (Cartesian) multipole expansion with the Cartesian multipole moments and the latter to the spherical multipole expansion with the spherical multipole moments.

In the first part of this section we derive a new definition of the Cartesian multipole moments that is based on the *Kelvin transform*. The second part dedicates itself to the spherical multipole expansion and we present the standard definition of the spherical multipole moments.

4.1.1 DEFINITION OF THE CARTESIAN MULTIPOLE MOMENTS

In textbooks (e.g. Jackson, 1998, p. 146) no general definition of the Cartesian multipole moments is given. Instead the Taylor expansion of the denominator in the integrand of eq. (4.4) is computed up to, for example, second order, i.e.

$$\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|} = \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} \left(\boldsymbol{x}'\cdot\nabla\right)^l \frac{1}{|\boldsymbol{x}|} = \frac{1}{|\boldsymbol{x}|} - x'^i\partial_i\frac{1}{|\boldsymbol{x}|} + x'^i x'^j\partial_i\partial_j\frac{1}{|\boldsymbol{x}|} - \cdots \tag{4.5}$$

In that case the potential is

$$\begin{aligned} 4\pi\epsilon_0\phi(\boldsymbol{x}) = &\int \rho(\boldsymbol{x}')\,\mathrm{d}^3x'\,\frac{1}{|\boldsymbol{x}|} + \int \rho(\boldsymbol{x}')x'_i\,\mathrm{d}^3x'\,\frac{x^i}{|\boldsymbol{x}|^3} \\ &+ \frac{1}{2}\int \rho(\boldsymbol{x}')x'_i x'_j\,\mathrm{d}^3x'\,\frac{(3x^ix^j - x^2\delta^{ij})}{|\boldsymbol{x}|^5} + \cdots \end{aligned}$$

It is common to interchange the primed and unprimed components. Consider, for example, the second order term, i.e.

$$\frac{1}{2}\int \rho(\boldsymbol{x}')(3x'_ix'_j - |\boldsymbol{x}'|^2\,\delta_{ij})\,\mathrm{d}^3x'\,\frac{x^ix^j}{|\boldsymbol{x}|^5}\,,$$

where we used that $|\boldsymbol{x}|^2 \delta^{ij} x'_i x'_j = |\boldsymbol{x}'|^2 \delta_{ij} x^i x^j$. This transforms the potential into

$$4\pi\epsilon_0\phi(\boldsymbol{x}) = \frac{q}{|\boldsymbol{x}|} + d_i \frac{x^i}{|\boldsymbol{x}|^3} + \frac{1}{2}\boldsymbol{Q}_{ij}\frac{x^i x^j}{|\boldsymbol{x}|^5} + \cdots ,$$

with the usual definitions of the monopole and the dipole. The components of the quadrupole moment are defined as

$$\boldsymbol{Q}_{ij} \coloneqq \int \rho(\boldsymbol{x}')(3x'_i x'_j - |\boldsymbol{x}'|^2 \delta_{ij})\ \mathrm{d}^3x' .$$

Hence, a general method to define the components of the Cartesian multipole moment of an arbitrary order l consists in taking l partial derivatives of $1/|\boldsymbol{x}|$ and interchanging the components of $\boldsymbol{x}$ and $\boldsymbol{x}'$ as in the above example. But note that the interchange did not change the (functional) form of the numerator of $\partial_i\partial_j 1/|\boldsymbol{x}|$, it merely replaced the components of $\boldsymbol{x}$ with the components of $\boldsymbol{x}'$. This implies that we could alternatively have obtained the expression in parenthesis in the quadrupole moment's definition by computing $\partial'_i\partial'_j 1/|\boldsymbol{x}'|$ and "removing" the denominator. The outlined idea leads to our definition of the components of the *Cartesian multipole moment of rank l*, namely

$$\begin{aligned} Q^{(l)}_{i_1\cdots i_l} &\coloneqq \int \rho(\boldsymbol{x}')\mathcal{K}\left[(-1)^l \partial'_{i_1}\cdots\partial'_{i_l}\frac{1}{|\boldsymbol{x}|'}\right]\mathrm{d}^3x' \\ &\coloneqq \int \rho(\boldsymbol{x}')M^{(l)}_{i_1\cdots i_l}(\boldsymbol{x}')\ \mathrm{d}^3x' . \end{aligned} \tag{4.6}$$

We call the functions $M_{i_1\cdots i_l}$ *multipole functions*. The indices reflect that the functions are different for different partial derivatives.

Here the calligraphic $\mathcal{K}$ denotes the *Kelvin transform* that is defined as

$$\mathcal{K}[f](\boldsymbol{x}) \coloneqq \frac{1}{|\boldsymbol{x}|} f\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|^2}\right) \tag{4.7}$$

see Axler et al. (2001, p. 61). An application of the definition given in eq. (4.6) for $l = 2$ recovers the quadrupole moment and demonstrates in which sense the Kelvin transform removes the denominator. With this definition at hand, the *(Cartesian) multipole expansion* is

$$4\pi\epsilon_0\phi(\boldsymbol{x}) = \sum_{l=0}^{\infty}\frac{1}{l!}Q^{(l)}_{i_1\cdots i_l}\frac{x^{i_1}\cdots x^{i_l}}{|\boldsymbol{x}|^{2l+1}} , \tag{4.8}$$

Of course, the first three multipole moments are the monopole $Q^{(0)} = q$, the dipole moment $Q^{(1)}_i = d_i$ and the quadrupole moment $Q^{(2)}_{ij} = \boldsymbol{Q}_{ij}$.

We show now that the definition in eq. (4.6) gives the known Cartesian multipole moments for an arbitrary rank l. The definition is correct if, as in the example of the quadrupole moment, for all l the interchange of $\boldsymbol{x}$ and $\boldsymbol{x}'$ does

not change the form of the numerator of the l-th partial derivative of $1/|\boldsymbol{x}|$ and if the Kelvin transform removes its denominator $|\boldsymbol{x}|^{2l+1}$.

A sufficient condition for the Kelvin transform to remove the denominator is that the numerator of the l-th partial derivative of $1/|\boldsymbol{x}|$ is a *homogeneous polynomial*. We say that a polynomial p of degree l is homogeneous if and only if

$$p(\lambda \boldsymbol{x}) = \lambda^l p(\boldsymbol{x}) \text{ for all } \lambda \in \mathbb{R}\,. \tag{4.9}$$

That this is a sufficient condition can be seen by noting that

$$(-1)^l \partial_{i_1} \cdots \partial_{i_l} \frac{1}{|\boldsymbol{x}|} = \frac{M^{(l)}_{i_1 \cdots i_l}(\boldsymbol{x})}{|\boldsymbol{x}|^{2l+1}}\,, \tag{4.10}$$

where $M^{(l)}_{i_1 \ldots i_l}$ is an as yet undetermined function (the above equality can be proven by induction). Moreover, Axler et al. (2001, cf. Lemma 5.15 on p. 86) show that $M^{(l)}_{i_1 \cdots i_l}$ is a homogeneous polynomial of degree l. Taking the Kelvin transform then yields

$$\mathcal{K}\left[(-1)^l \partial_{i_1} \cdots \partial_{i_l} \frac{1}{|\boldsymbol{x}|}\right] = |\boldsymbol{x}|^{2l+1} \frac{M^{(l)}_{i_1 \cdots i_l}(\boldsymbol{x})}{|\boldsymbol{x}|^{2l+1}} = M^{(l)}_{i_1 \cdots i_l}(\boldsymbol{x})\,,$$

which shows that it suffices that $M^{(l)}_{i_1 \cdots i_l}$ is a homogeneous polynomial to remove the denominator and that the undetermined functions are the multipole functions.

In the example of the quadrupole term, the interchange of $\boldsymbol{x}$ and $\boldsymbol{x}'$ did not change the (functional) form of $M^{(2)}_{ij}$ because this form was "somehow" special. We now know that the multipole functions are homogeneous polynomials. We can further restrict their form. If we keep in mind that $M^{(l)}_{i_1 \cdots i_l}$ is the numerator of $g(\boldsymbol{x}) \coloneqq \partial_{i_1} \cdots \partial_{i_l}(1/|\boldsymbol{x}|)$, the following two considerations fix its form:

Firstly, the form of the function g does not change if it is defined in a rotated coordinate system. Assume that the relation between the rotated coordinates $\hat{\boldsymbol{x}}$ and the original coordinates $\boldsymbol{x}$ is $\hat{\boldsymbol{x}} = \boldsymbol{R}\boldsymbol{x}$. The former statement means that $\hat{g}(\hat{\boldsymbol{x}}) = g(\hat{\boldsymbol{x}})$ where $\hat{g}(\hat{\boldsymbol{x}}) \coloneqq \hat{\partial}_{i_1} \cdots \hat{\partial}_{i_l}(1/|\hat{\boldsymbol{x}}|)$. A comparison of g's and $\hat{g}$'s definitions proves this. This restricts the form of the numerator of g, as the following example for $l = 2$ illustrates

$$\begin{aligned}\hat{g}(\hat{\boldsymbol{x}}) = \hat{\partial}_i \hat{\partial}_j \frac{1}{|\hat{\boldsymbol{x}}|} = (\boldsymbol{R})_i{}^k \partial_k (\boldsymbol{R})_j{}^l \partial_l \frac{1}{|\boldsymbol{x}|} &= (\boldsymbol{R})_i{}^k (\boldsymbol{R})_j{}^l \frac{3x_k x_l - |\boldsymbol{x}|^2 \delta_{kl}}{|\boldsymbol{x}|^5} \\ &= \frac{3\hat{x}_i \hat{x}_j - \hat{x}^2 \delta_{ij}}{|\hat{\boldsymbol{x}}|^5} = g(\hat{\boldsymbol{x}})\,,\end{aligned}$$

where we used that $|\hat{\boldsymbol{x}}| = |\boldsymbol{x}|$ and $\hat{\nabla} = \boldsymbol{R}\nabla$. Note to fulfil the condition $\hat{g}(\hat{\boldsymbol{x}}) = g(\hat{\boldsymbol{x}})$, the rotation matrices must either transform a component of $\boldsymbol{x}$ or they must

be multiplied with each other, i.e. $(\boldsymbol{R})_i{}^k(\boldsymbol{R})_j{}^l\delta_{kl} = (\boldsymbol{R})_i{}^k(\boldsymbol{R}^\mathsf{T})_{kj} = \delta_{ij}$. Hence, all the terms of $M^{(l)}_{i_1\cdots i_l}$ must be products of components of $\boldsymbol{x}$ and Kronecker deltas to "cope" with the rotation matrices.

Secondly, the exchange of any two of the indices in g's definition does not change its numerator, because it is possible to exchange the partial derivatives with each other (Schwarz's theorem). Hence, the object $M^{(l)}_{i_1\cdots i_l}$ is symmetric in all its indices. This implies that a sum of the mentioned products of components of $\boldsymbol{x}$ and Kronecker deltas is needed. For example, consider the $l = 3$ case

$$M^{(3)}_{ijk} = c_{3,0}x_ix_jx_k + c_{3,1}\left|\boldsymbol{x}\right|^2\left(x_i\delta_{jk} + x_j\delta_{ik} + x_k\delta_{ij}\right),$$

where $c_{3,0}, c_{3,1} \in \mathbb{R}$ are coefficients. Note that the sum is over pairs of indices and that this guarantees that any two of the indices can be exchanged without changing $M^{(3)}_{ijk}$. The factor $|\boldsymbol{x}|^2$ reflects that $M^{(3)}_{ijk}$ must be a homogeneous polynomial of degree three.

We summarise: The fact that $M^{(l)}_{i_1\cdots i_l}$ is a homogeneous polynomial of degree l, that can be invariably defined in a rotated coordinate system and that is symmetric in its indices implies that its functional form must be

$$M^{(l)}_{i_1\cdots i_l} = \sum_{k=0}^{\lfloor l/2\rfloor} c_{l,k}\left|\boldsymbol{x}\right|^{2k} R_{l,2k}\,, \tag{4.11}$$

where $\lfloor l/2\rfloor$ is the floor function and where we introduced

$$R_{l,2k} \coloneqq \hat{P}\left(\delta_{i_1i_2}\cdots\delta_{i_{2k-1}i_{2k}}x_{i_{2k+1}}\cdots x_{i_l}\right).$$

The operator $\hat{P}$ produces the *sum* over the pairs of indices needed to assure symmetry (cf. with the previous example). This operator hides a lot of the complexity, but is explained in detail in Appendix A.1.2. A closed-form expression for the coefficients $c_{l,k}$ is derived in Appendix A.1.3. Similar arguments are used by Efimov (1979, p. 426) to fix the functional form of the multipole functions.

This is the "somehow" special (functional) form of the multipole functions which permits that the interchange of $\boldsymbol{x}$ and $\boldsymbol{x}'$ does not change their form but only replaces the components of $\boldsymbol{x}$ with the components of $\boldsymbol{x}'$. To see this notice that, in the light of eq. (4.10) and eq. (4.11), a term in the Taylor expansion of eq. (4.5) of arbitrary order l contains

$$\left|\boldsymbol{x}\right|^{2k}x'^{i_1}\cdots x'^{i_l}\hat{P}\left(\delta_{i_1i_2}\cdots\delta_{i_{2k-1}i_{2k}}x_{i_{2k+1}}\cdots x_{i_l}\right).$$

Using that the exponent of $|\boldsymbol{x}|^{2k}$ is even and remembering that $|\boldsymbol{x}|^2\,\delta^{ij}x'_ix'_j = |\boldsymbol{x}'|^2\,\delta_{ij}x^ix^j$, we can replace the components of $\boldsymbol{x}$ and $\boldsymbol{x}'$.

We conclude that the definition of the Cartesian multipole moments, given in eq. (4.6), is correct. Eventually, we remark that this definition allows to directly prove that the $Q^{(l)}$ are tensors, i.e. they transform as required under coordinate transformations, see Appendix A.1.1.

Before we close this section, we emphasise that the multipole functions $M^{(l)}_{i_1\cdots i_l}$ are *harmonic functions*, i.e.

$$\Delta M^{(l)}_{i_1\cdots i_l} = 0\,.$$

This is also proven in (the already quoted) Lemma 5.15 in Axler et al. (2001, p. 86). Moreover, we denote the space of homogeneous and harmonic polynomials of degree l (in three variables) with $\mathcal{H}^l(\mathbb{R}^3)$.

In what follows, it will prove convenient to use a different notation for the indices of $M^{(l)}$, namely

$$M^{(l)}_{pqr} \coloneqq \mathcal{K}\left[(-1)^l \partial_x^p \partial_y^q \partial_z^r \frac{1}{|\boldsymbol{x}|}\right] \text{ with } p+q+r = l\,. \tag{4.12}$$

This notation implicitly takes advantage of the symmetry, cf. Johnston (1960, eq. 5a). For example, $M^{(3)}_{112} = M^{(3)}_{121}$ in the $M^{(l)}_{i_1\ldots i_l}$ notation and both components are $M^{(3)}_{210}$ in the newly introduced pqr-notation.

And last, but not least Axler et al. (2001, p. 92) show in Theorem 5.25 that

$$\mathcal{H}^l(\mathbb{R}^3) = \operatorname{span}\left\{M^{(l)}_{pqr} \mid p \leq 1 \text{ and } p+q+r = l\right\}. \tag{4.13}$$

In words, the functions $M^{(l)}_{pqr}$ are homogeneous and harmonic polynomials of degree l and a subset of them is a basis of the space of these polynomials. We call this subset the *multipole basis functions*. We note that restricting the domain of the harmonic and homogeneous polynomials to the unit sphere yields the space of spherical harmonics of degree l, i.e. $\mathcal{H}^l(S^2)$, see Definition 1 in Müller (1966).

4.1.2 DEFINITION OF THE SPHERICAL MULTIPOLE MOMENTS

After introducing a new definition of the Cartesian multipole moments, we turn our attention to the standard definition of the spherical multipole moments. Instead of Taylor expanding the denominator of the integrand in eq. (4.4), we interpret it as the generating function of the Legendre polynomials, i.e

$$\frac{1}{|\boldsymbol{x}-\boldsymbol{x}'|} = \sum_{l=0}^{\infty} \frac{|\boldsymbol{x}'|^l}{|\boldsymbol{x}|^{l+1}} P_l(\cos\alpha)\,, \tag{4.14}$$

where α is the angle between $\boldsymbol{x}$ and $\boldsymbol{x}'$, see Jackson (1998, p. 102, eq. 3.38). Together with the addition theorem for Laplace's spherical harmonics (cf. Jackson, 1998, p. 110, eq. 3.62), which is

$$P_l(\cos\alpha) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_l^{m*}(\theta',\varphi') Y_l^m(\theta,\varphi)\,, \tag{4.15}$$

we obtain the *spherical multipole expansion*

$$4\pi\epsilon_0\phi(\boldsymbol{x}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\frac{q_l^m}{|\boldsymbol{x}|^{2l+1}}\,|\boldsymbol{x}|^l\,Y_l^m(\theta,\varphi)\,. \tag{4.16}$$

The function $|\boldsymbol{x}|^l\,Y_l^m(\theta,\varphi)$ is called a *solid harmonic of degree l and order m*. Furthermore, the coefficients q_l^m are called *spherical multipole moments* and Jackson (1998, eq. 4.3) defines them as

$$q_l^m \coloneqq \int \rho(\boldsymbol{x}')\,|\boldsymbol{x}'|^l\,Y_l^{m*}(\theta',\varphi')\;\mathrm{d}^3x'\,. \tag{4.17}$$

The solid harmonics are homogeneous polynomials of degree l. This can be seen by using the following definition of the associated Legendre polynomials

$$P_l^m(\cos\theta) = (-1)^m\sin^m\theta\frac{\mathrm{d}^m}{\mathrm{d}\cos\theta^m}P_l(\cos\theta)\,, \tag{4.18}$$

see Jackson (cf. 1998, p. 108, eq. 3.49). Then, the solid harmonics with order greater than or equal to zero ($m \geq 0$) are

$$\begin{aligned}
&|\boldsymbol{x}|^l\,Y_l^m(\theta,\varphi)\\
&\quad= (-1)^m N_l^m\,|\boldsymbol{x}|^{l-m}\,\frac{\mathrm{d}^m P_l}{\mathrm{d}\cos\theta^m}(|\boldsymbol{x}|\sin\theta\cos\varphi + \mathrm{i}\,|\boldsymbol{x}|\sin\theta\sin\varphi)^m\\
&\quad= (-1)^m N_l^m\sum_{k=0}^{\left\lfloor\frac{l-m}{2}\right\rfloor} a_k\,|\boldsymbol{x}|^{2k}\,z^{l-m-2k}(x+\mathrm{i}y)^m\\
&\quad\coloneqq p(\boldsymbol{x})\,.
\end{aligned} \tag{4.19}$$

The $a_k \in \mathbb{R}$ are coefficients that collect the numerical factors. In the second equation we expressed the spherical coordinates in Cartesian coordinates and exploited that the m-th derivative of a degree l polynomial yields a degree $l-m$ polynomial. The Legendre polynomials, in particular, consist of either odd or even powers of $\cos\theta$. It can be checked that the definition of a homogeneous polynomial of degree l, that was given in eq. (4.9), applies to p as defined in eq. (4.19). The argument holds true for negative m as well. In this case $(x+\mathrm{i}y)^m$ becomes $(x-\mathrm{i}y)^m$.

Furthermore, the solid harmonics (as their name suggests) are harmonic functions. In spherical coordinates the Laplace operator can be split into a radial part and an angular part, i.e. $\Delta = \Delta_{|\boldsymbol{x}|} + \Delta_{\theta,\varphi}/|\boldsymbol{x}|^2$ and the spherical harmonics are eigenfunctions of the angular part, i.e.

$$\Delta_{\theta,\varphi}Y_l^m = -l(l+1)Y_l^m\,, \tag{4.20}$$

see Landau and Lifshitz (1977, p. 91, eq. 28.7). Hence,

$$\begin{aligned}
\Delta\left(|\boldsymbol{x}|^l\,Y_l^m\right) &= Y_l^m\left(\partial_{|\boldsymbol{x}|}^2\,|\boldsymbol{x}|^l + \frac{2}{|\boldsymbol{x}|}\partial_{|\boldsymbol{x}|}\,|\boldsymbol{x}|^l\right) + |\boldsymbol{x}|^{l-2}\,\Delta_{\theta,\varphi}Y_l^m\\
&= l(l+1)\,|\boldsymbol{x}|^{l-2}\,Y_l^m - l(l+1)\,|\boldsymbol{x}|^{l-2}\,Y_l^m\\
&= 0\,.
\end{aligned}$$

We conclude this section by proving that the solid harmonics are also a basis of the space of homogeneous and harmonic polynomials of degree l, namely

$$\mathcal{H}^l(\mathbb{R}^3) = \operatorname{span}\left\{|\boldsymbol{x}|^l\, Y_l^m \mid -l \leq m \leq l\right\}. \tag{4.21}$$

We proceed in two steps: First, we determine the dimension of $\mathcal{H}^l(\mathbb{R}^3)$ and, secondly, we show that the solid harmonics are independent. The dimension of $\mathcal{H}^l(\mathbb{R}^3)$ is $2l+1$. This can be derived in different ways, for example, as in Müller (1966, p. 11, eq. 11) or, alternatively, as done in Axler et al. (2001, p. 78, Proposition 5.8). Notice there are exactly $2l+1$ solid harmonics of degree l. The linear independence follows from the orthogonality of spherical harmonics, namely

$$\begin{aligned}\left(\, Y_{l'}^{m'} \,\middle|\, Y_l^m \,\right) &\coloneqq \int_{S^2} {Y_{l'}^{m'}}^* Y_l^m \,\mathrm{d}\Omega \\ &= \int_0^{2\pi}\int_0^{\pi} {Y_{l'}^{m'}}^* Y_l^m \,\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi = \delta_{l'l}\delta_{m'm}\,,\end{aligned} \tag{4.22}$$

where S^2 is the unit sphere, i.e. $S^2 \coloneqq \{\boldsymbol{x} \in \mathbb{R}^3 \mid |\boldsymbol{x}| = 1\}$, see Jackson (1998, p. 108, eq. 3.55). Hence the solid harmonics of degree l are a basis of the space of homogeneous and harmonic polynomials of degree l.

We gave explicit definitions of the Cartesian multipole moments and the spherical multipole moments. Both definitions contain homogeneous and harmonic polynomials: The multipole functions $M_{pqr}^{(l)}$ in the former case, solid harmonics $r^l Y_l^m$ in the latter. Moreover, a subset of the functions $M_{pqr}^{(l)}$, namely the multipole basis functions, are a basis of $\mathcal{H}^l(\mathbb{R}^3)$. The same is true of the solid harmonics. Thus, the relation between the Cartesian multipole moments and spherical multipole moments can be formalised as a basis transformation between the solid harmonics and the multipole basis functions. In the next section we introduce Efimov's ladder operator to derive the basis transformation.

4.2 EFIMOV'S LADDER OPERATOR

In this section we present an alternative definition of the Cartesian multipole moments and we show that this definition is equivalent to the one we introduced in Sec. (4.1.1). Thereby we prove the correctness of our definition in a more rigorous fashion. The reason to introduce a second definition is that mathematical results about both definitions play a role in our derivation of the relation between the expansion coefficients.

4.2.1 AN ALTERNATIVE DEFINITION OF THE CARTESIAN MULTIPOLE MOMENTS

Efimov (1979) gives an alternative definition of the Cartesian multipole moments. To derive it, he multiplies the Taylor expansion given eq. (4.5) with $|\boldsymbol{x}|$,

which yields

$$\frac{|\boldsymbol{x}|}{|\boldsymbol{x}-\boldsymbol{x}'|} = \sum_{l=0}^{\infty} |\boldsymbol{x}| \frac{(-\boldsymbol{x}'\cdot\nabla)^l}{l!}\frac{1}{|\boldsymbol{x}|} = \sum_{l=0}^{\infty} \frac{M^{(l)}_{i_1\cdots i_l}(\boldsymbol{x})}{l!\,|\boldsymbol{x}|^{2l}} x'^{i_1}\cdots x'^{i_l}\,.$$

Taking advantage of the fact that the functions $M^{(l)}_{i_1\cdots i_l}$ are homogeneous polynomials of degree l and setting $\tilde{\boldsymbol{x}} \coloneqq \boldsymbol{x}/|\boldsymbol{x}|^2$, leads to the equivalent equation

$$\sum_{l=0}^{\infty} \frac{1}{|\tilde{\boldsymbol{x}}|}\frac{(-\boldsymbol{x}'\cdot\nabla)^l}{l!}|\tilde{\boldsymbol{x}}| = \sum_{l=0}^{\infty}\frac{1}{l!}M^{(l)}_{i_1\cdots i_l}(\tilde{\boldsymbol{x}})x'_{i_1}\cdots x'_{i_l}\,.$$

Because the terms on both sides of the equation are sums of homogeneous polynomials of the same degrees, one finds

$$(-1)^l\frac{1}{|\tilde{\boldsymbol{x}}|}(\boldsymbol{x}'\cdot\nabla)^l|\tilde{\boldsymbol{x}}| = M^{(l)}_{i_1\cdots i_l}(\tilde{\boldsymbol{x}})x'^{i_1}\cdots x'^{i_l}\,. \tag{4.23}$$

Notice that the derivatives on the left-hand side of the equation are taken with respect to $\boldsymbol{x}$ (and not $\tilde{\boldsymbol{x}}$). We may rewrite the above to get a convenient grouping

$$\begin{aligned}(-1)^l\overbrace{\frac{1}{|\tilde{\boldsymbol{x}}|}(\boldsymbol{x}'\cdot\nabla)|\tilde{\boldsymbol{x}}|\underbrace{\frac{1}{|\tilde{\boldsymbol{x}}|}(\boldsymbol{x}'\cdot\nabla)|\tilde{\boldsymbol{x}}|\cdots\frac{1}{|\tilde{\boldsymbol{x}}|}(\boldsymbol{x}'\cdot\nabla)|\tilde{\boldsymbol{x}}|}_{\coloneqq f(\tilde{\boldsymbol{x}})}}^{l\text{ times}}\\ = (-1)^{l+1}x'^i\left(2\tilde{x}_i\tilde{x}^j\tilde{\partial}_j - |\tilde{\boldsymbol{x}}|^2\tilde{\partial}_i + \tilde{x}_i\right)f(\tilde{\boldsymbol{x}})\,.\end{aligned}$$

The expression in parenthesis is *Efimov's ladder operator*. We define

$$\mathrm{D}_i \coloneqq \left(2x_ix^j\partial_j - |\boldsymbol{x}|^2\partial_i + x_i\right) \tag{4.24}$$

and using this definition l times turns eq. (4.23) into

$$(\boldsymbol{x}'\cdot\tilde{\boldsymbol{D}})^l 1(\tilde{\boldsymbol{x}}) = M^{(l)}_{i_1\cdots i_l}(\tilde{\boldsymbol{x}})x'^{i_1}\cdots x'^{i_l}\,.$$

Here $1(\tilde{\boldsymbol{x}}) = 1$ is a constant function. This implies an alternative definition of the multipole functions

$$M^{(l)}_{i_1\cdots i_l}(\boldsymbol{x}) \coloneqq \mathrm{D}_{i_1}\cdots\mathrm{D}_{i_l}1(\boldsymbol{x})\,. \tag{4.25}$$

Hence, we can also use Efimov's ladder operator to define the Cartesian multipole moments, i.e.

$$Q^{(l)}_{i_1\cdots i_l} \coloneqq \int \rho(\boldsymbol{x}')\mathrm{D}'_{i_1}\cdots\mathrm{D}'_{i_l}1(\boldsymbol{x}')\;\mathrm{d}^3x'\,. \tag{4.26}$$

At the end of this section, we note that Efimov's ladder operators commute, i.e.

$$[\mathrm{D}_i,\mathrm{D}_j] = \mathrm{D}_i\mathrm{D}_j - \mathrm{D}_j\mathrm{D}_i = 0\,. \tag{4.27}$$

This must be the case, because $M^{(l)}_{i_1\cdots i_l}$ is symmetric. Moreover,

$$\mathrm{D}_i\mathrm{D}_i = |\boldsymbol{x}|^4 \Delta\,.$$

An immediate consequence of these two equations is that

$$M^{(l)}_{i_1\cdots i_k\cdots i_k\cdots i_l} = 0\,. \tag{4.28}$$

We say that the object $M^{(l)}_{i_1\cdots i_l}$ is *traceless*, i.e. the contraction of two arbitrary indices is zero, see Efimov (1979, eq. 1.5 - 1.7). This property is carried over to the Cartesian multipole moments. Thus, the Cartesian multipole moments $Q^{(l)}$ are symmetric and traceless tensors of rank l.

4.2.2 EQUIVALENCE OF THE DEFINITIONS

We explicitly prove that the two definitions of the multipole functions are equivalent, i.e.

$$\mathcal{K}\left[(-1)^l \partial_{i_1}\cdots\partial_{i_l}\frac{1}{|\boldsymbol{x}|}\right] = \mathrm{D}_{i_1}\cdots\mathrm{D}_{i_l}1\,.$$

The first step consists in showing that

$$\mathcal{K}\left[(-1)^l \partial_{i_1}\cdots\partial_{i_l}\frac{1}{|\boldsymbol{x}|}\right] = \prod_{k=1}^{l}\left((2l-(2k-1))\,x_{i_k} - x^2\partial_{i_k}\right)1\,. \tag{4.29}$$

This can be proven by induction. A quick computation shows that the statement holds true for the base case $l = 1$. From eq. (4.10), and noting that $\mathcal{K}[\mathcal{K}[f]] = f$, it follows that

$$(-1)^l \partial_{i_1}\cdots\partial_{i_l}\frac{1}{|\boldsymbol{x}|} = \mathcal{K}\left[M^{(l)}_{i_1\cdots i_l}\right](\boldsymbol{x}) = \frac{M^{(l)}_{i_1\cdots i_l}(\boldsymbol{x})}{|\boldsymbol{x}|^{2l+1}}\,.$$

Taking the derivative with respect to x_j yields

$$\partial_j\partial_{i_1}\cdots\partial_{i_l}\frac{1}{|\boldsymbol{x}|} = \frac{(-1)^{l+1}}{|\boldsymbol{x}|^{2l+3}}\left((2l+1)x_j - |\boldsymbol{x}|^2\,\partial_j\right)M^{(l)}_{i_1\cdots i_l}(\boldsymbol{x})\,.$$

Note that acting with the expression in parenthesis on $M^{(l)}_{i_1\cdots i_l}$ gives a homogeneous polynomial of degree $l+1$. We again take the Kelvin transform to remove $1/|\boldsymbol{x}|^{2l+3}$ and use the induction hypothesis to see that

$$\begin{aligned}
&\mathcal{K}\left[(-1)^{l+1}\partial_j\partial_{i_1}\cdots\partial_{i_l}\frac{1}{|\boldsymbol{x}|}\right]\\
&\quad= \left((2l+1)x_j - |\boldsymbol{x}|^2\,\partial_j\right)M^{(l)}_{i_1\cdots i_l}(\boldsymbol{x})\\
&\quad= \left((2l+1)x_j - |\boldsymbol{x}|^2\,\partial_j\right)\prod_{k=1}^{l}\left((2l-(2k-1))\,x_{i_k} - |\boldsymbol{x}|^2\,\partial_{i_k}\right)1\\
&\quad= \prod_{k=0}^{l}\left((2l-(2k-1))\,x_{i_k} - |\boldsymbol{x}|^2\,\partial_{i_k}\right)1\,,
\end{aligned}$$

where we in the last line relabelled the index $j \to i_0$. Shifting the index of the product by one yields the conclusion.

In a second step, we prove that each factor in the product in eq. (4.29) is equivalent to Efimov's ladder operator. We choose an arbitrary factor with index k equal to n and express the right-hand side of eq. (4.29) as

$$\left((2l-1)x_{i_1} - |\boldsymbol{x}|^2 \partial_{i_1}\right) \times \cdots \times \left(2(l-n)x_{i_n} - |\boldsymbol{x}|^2 \partial_{i_n} + x_{i_n}\right) M^{(l-n)}_{i_{n+1}\cdots i_l}(\boldsymbol{x}), \quad (4.30)$$

where we used that

$$\begin{aligned}
&\prod_{k=n+1}^{l} \left((2l-(2k-1))x_{i_k} - |\boldsymbol{x}|^2 \partial_{i_k}\right) 1 \\
&\qquad = \prod_{k=1}^{l-n} \left((2(l-n)-(2k-1))x_{i_{k+n}} - |\boldsymbol{x}|^2 \partial_{i_{k+n}}\right) 1 \\
&\qquad = M^{(l-n)}_{i_{n+1}\cdots i_l}(\boldsymbol{x}) .
\end{aligned}$$

To show that the factor in eq. (4.30) corresponding to the index $k = n$ and D_{i_n} are equivalent, we have to prove that

$$2(l-n)x_{i_n} M^{(l-n)}_{i_{n+1}\cdots i_l} = 2x_{i_n} x^m \partial_m M^{(l-n)}_{i_{n+1}\cdots i_l} .$$

This equality holds because the function $u \coloneqq M^{(l-n)}_{i_n\cdots i_l}$ is a homogeneous polynomial of degree $l-n$. Note that every homogeneous polynomial of an arbitrary degree l can be written as a linear combination of monomials $x^{j_1}\cdots x^{j_l}$, i.e. $p(\boldsymbol{x}) = \alpha_{j_1\cdots j_l} x^{j_1}\cdots x^{j_l}$, where α is a collection of coefficients. There are $(l+2)(l+1)/2$ different monomials. If the object α is symmetric in all its indices, it contains an equal amount of independent coefficients. Thus, there exists a symmetric α such that $u(\boldsymbol{x}) = \alpha_{j_1\cdots j_{l-n}} x^{j_1}\cdots x^{j_{l-n}}$. This implies that

$$\begin{aligned}
x^m \partial_m u(\boldsymbol{x}) &= x^m \alpha_{j_1\cdots j_{l-n}} \partial_m \left(x^{j_1}\cdots x^{j_{l-n}}\right) = (l-n) x^m \alpha_{j_1\cdots j_{l-n}} \delta^{j_1}_m \cdots x^{j_{l-n}} \\
&= (l-n)\alpha_{j_1\cdots j_{l-n}} x^{j_1}\cdots x^{j_{l-n}} = (l-n)u(\boldsymbol{x}) .
\end{aligned}$$

In the second line we used that α is symmetric. Since n was arbitrary, we conclude that all factors in the product in eq. (4.29) are equivalent to a corresponding component of the ladder operator $\boldsymbol{D}$. Before we end this section, we use the alternative definition of the Cartesian multipole moments (and the multipole functions) to learn more about them.

We take advantage of the previous computation to make a statement about the Cartesian mutlipole expansion: We just stated that a homogeneous polynomial of degree l can be written as a linear combination of monomials with the help of a symmetric collection of coefficients α. Moreover, it can be shown that if the polynomial is also harmonic, α must be traceless (see Jeevanjee, 2011, ex.3.26). At the end of Sec. (4.26) we showed that the Cartesian multipole

moments are traceless, hence not only the multipole functions $M^{(l)}_{i_1\cdots i_l}$ are homogeneous and harmonic polynomials but also the functions in the Cartesian multipole expansion, namely

$$Q^{(l)}_{i_1\cdots i_l} x^{i_1}\cdots x^{i_l}\,,$$

are homogeneous and harmonic polynomials of degree l.

4.2.3 PRELIMINARY BASIS TRANSFORMATION

We showed that the definitions of the Cartesian multipole moments and the spherical multipole moments are based on homogeneous and harmonic polynomials, namely the Cartesian multipole moments use the multipole functions and the spherical harmonics use the solid harmonics. With the tools derived up to now, we are in a position to relate the two multipole moments. This can be done by expressing the solid harmonics as a sum of multipole functions $M^{(l)}_{i_1\cdots i_l}$.

We start again with the denominator of the integrand in eq. (4.4) and fix $\boldsymbol{x}'$ to be the unit vector pointing into the z-direction. This leads to

$$\frac{1}{|\boldsymbol{x}-\boldsymbol{e}_z|} = \sum_{l=0}^{\infty} \frac{(-\boldsymbol{e}_z\cdot\nabla)^l}{l!}\frac{1}{|\boldsymbol{x}|} = \sum_{l=0}^{\infty}\frac{1}{|\boldsymbol{x}|^{l+1}} P_l(\cos\theta)\,,$$

where we used the Taylor expansion given in eq. (4.5) for the left-hand side and eq. (4.14) for the right-hand side. Since $\boldsymbol{x}' = \boldsymbol{e}_z$, the angle α is the polar angle θ. In a next step, we take the Kelvin transform of the above equation, i.e.

$$\sum_{l=0}^{\infty}\frac{1}{l!}\mathcal{K}\left[(-1)^l\partial_z^l\frac{1}{|\boldsymbol{x}|}\right] = \sum_{l=0}^{\infty}\mathcal{K}\left[\frac{1}{|\boldsymbol{x}|^{2l+1}}\,|\boldsymbol{x}|^l P_l(\cos\theta)\right].$$

Notice that the Kelvin transform is linear, i.e. $\mathcal{K}[f+\lambda g] = \mathcal{K}[f] + \lambda\mathcal{K}[g]$, see eq. (4.7). We now use that $N_l^0\,|\boldsymbol{x}|^l P_l = |\boldsymbol{x}|^l\, Y_l^0$. Furthermore, since a solid harmonic of degree l is a homogeneous polynomial of degree l, the Kelvin transform removes the denominator $|\boldsymbol{x}|^{2l+1}$ of the right-hand side, cf. Section 4.1.1. Together with the equivalence of the Kelvin transform of the partial derivatives of $1/|\boldsymbol{x}|$ and Efimov's ladder operator, we obtain the equivalent statement

$$\sum_{l=0}^{\infty}|\boldsymbol{x}|^l P_l(\cos\theta) = \sum_{l=0}^{\infty}\frac{1}{l!}(\boldsymbol{e}_z\cdot\boldsymbol{D})^l 1\,.$$

Since all term on both sides of above's equation are polynomials of increasing degree l, the equation implies that

$$|\boldsymbol{x}|^l\, Y_l^0 = \frac{1}{l!}\sqrt{\frac{2l+1}{4\pi}}(\boldsymbol{e}_z\cdot\mathrm{D})^l 1 = \frac{1}{l!}\sqrt{\frac{2l+1}{4\pi}}M^{(l)}_{00l}\,. \tag{4.31}$$

Compare with Efimov (1979, p. 428, eq. 2.9) and notice that we use the notation introduced in eq. (4.12) for the indices of the multipole functions.

To express a solid harmonic of order m greater than zero as a sum of multipole functions, we use that the spherical harmonics Y_l^m are the common eigenfunctions of the square of the *angular momentum operator* (known from Quantum Mechanics) and one of its components. The angular momentum operator is defined as

$$\mathrm{L} \coloneqq -\mathrm{i}\boldsymbol{x} \times \nabla \,, \tag{4.32}$$

see Landau and Lifshitz (cf. 1977, p. 83, eq. 26.2). We are particularly interested in the *angular momentum ladder operator*, i.e.

$$\begin{aligned} \mathrm{L}_\pm = \mathrm{L}_x \pm \mathrm{i}\mathrm{L}_y &= -\mathrm{i}\left(y\partial_z - z\partial_y \pm \mathrm{i}(z\partial_x - x\partial_z)\right) \\ &= e^{\pm\mathrm{i}\varphi}\left(\pm\partial_\theta + \mathrm{i}\cot\theta\partial_\varphi\right) . \end{aligned} \tag{4.33}$$

The last line shows the ladder operators in spherical coordinates, see Landau and Lifshitz (1977, p. 85, eq. 26.15). The ladder operators are used to increase (or decrease) the order m of a spherical harmonic, e.g.

$$Y_l^m = ((l+m)(l-m+1))^{-1/2}\,\mathrm{L}_+ Y_l^{m-1} \,, \tag{4.34}$$

see Landau and Lifshitz (1977, p. 89, eq. 27.12). We can apply this operator m times to eq. (4.31) and we get

$$|\boldsymbol{x}|^l\, Y_l^m = \sqrt{\frac{(l-m)!}{(l+m)!}}\,|\boldsymbol{x}|^l\,\mathrm{L}_+^m Y_l^0 = \frac{1}{l!}\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}\mathrm{L}_+^m \mathrm{D}_z^l 1 \,, \tag{4.35}$$

where we used the fact that L_+ does not contain a derivative with respect to $|\boldsymbol{x}|$ as can be seen in eq. (4.33). We can compute the commutator of L_+ and D_z. This yields

$$[L_+, \mathrm{D}_z] = -\left(\mathrm{D}_x + \mathrm{i}\mathrm{D}_y\right) . \tag{4.36}$$

This computation is done in detail in Appendix A.2.

One can prove by induction that

$$\mathrm{L}_+^m \mathrm{D}_z^l 1 = (-1)^m \frac{l!}{(l-m)!}\mathrm{D}_z^{l-m}(\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^m 1 \,. \tag{4.37}$$

For the base case $m = 1$ we find

$$\begin{aligned} \mathrm{L}_+ \mathrm{D}_z^l 1 &= \left(\mathrm{D}_z \mathrm{L}_+ - (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)\right)\mathrm{D}_z^{l-1} 1 \\ &= \mathrm{D}_z^l \mathrm{L}_+ 1 - l\mathrm{D}_z^{l-1}(\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)1 = -\frac{l!}{(l-1)!}\mathrm{D}_z^{l-1}(\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)1 \,, \end{aligned}$$

where we repeatedly used the commutator given in eq. (4.36) and that Efimov's ladder operators commute, see eq. (4.27). The last equality is a consequence of

$\mathrm{L}_+ 1 = 0$. The induction step, $\mathrm{L}_+ \mathrm{L}_+^m \mathrm{D}_z^l 1$, repeats the above computation, i.e.

$$\begin{aligned}
\mathrm{L}_+ \mathrm{L}_+^m \mathrm{D}_z^l 1 &= (-1)^m \frac{l!}{(l-m)!} \mathrm{L}_+ \mathrm{D}_z^{l-m} (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^m 1 \\
&= (-1)^m \frac{l!}{(l-m)!} \\
&\quad \times \left(\mathrm{D}_z^{l-m} \mathrm{L}_+ (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^m 1 - (l-m) \mathrm{D}_z^{l-(m+1)} (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^{m+1} 1 \right) \\
&= (-1)^{m+1} \frac{l!}{(l-(m+1))!} \mathrm{D}_z^{l-(m+1)} (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^{m+1} .
\end{aligned}$$

Notice that $\mathrm{L}_+ (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^m = \mathrm{L}_+ \mathrm{D}_z^{m-m} (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^m \propto |\boldsymbol{x}|^l \mathrm{L}_+ Y_m^m = 0$, due to the induction hypothesis, formulated in eq.(4.37), and eq. (4.35).

Having proved that eq. (4.37) holds true, we substitute it into the previously derived expression for $|\boldsymbol{x}|^l Y_l^m$, i.e. into eq. (4.35). This yields the final result of this section: the *preliminary basis transformation*, namely

$$\begin{aligned}
|\boldsymbol{x}|^l Y_l^m &= a_l^m \mathrm{D}_z^{l-m} (\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)^m 1 \\
&= a_l^m \sum_{p=0}^{m} \mathrm{i}^{m-p} \binom{m}{p} \mathrm{D}_x^p \mathrm{D}_y^{m-p} \mathrm{D}_z^{l-m} 1 \\
&= a_l^m \sum_{p=0}^{m} \mathrm{i}^{m-p} \binom{m}{p} M^{(l)}_{p(m-p)(l-m)} .
\end{aligned} \tag{4.38}$$

Here $m \geq 0$ and we used the notation of eq. (4.12) for the indices of the multipole functions. The factor is

$$a_l^m \coloneqq (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{1}{(l+m)!(l-m)!}} .$$

We called this expression for the solid harmonics a preliminary basis transformation, because the sum on the right-hand side of eq. (4.38) contains multipole functions with $p > 1$, which are not multipole *basis* functions. The solid harmonics with negative order m are obtained with the help of $Y_l^{-m} = (-1)^m Y_l^{m*}$, see Jackson (1998, p. 146, eq. 4.3, 4.7). Efimov (1979, p. 429, eq. 2.10) derived eq. (4.38) with similar arguments.

The preliminary basis transformation already allows us to express the spherical multipole moments in terms of the Cartesian multipole moments. We can take the complex conjugate of eq. (4.38) and plug it into the definition of the spherical multipole moments as given in eq. (4.17). This results in

$$q_l^m = a_l^m \sum_{p=0}^{m} (-\mathrm{i})^{m-p} \binom{m}{p} Q^{(l)}_{p(m-p)(l-m)} .$$

In his textbook Classical Electrodynamics, Jackson (1998, p. 146) gives the formulae for the the spherical multipole moments with degree $l = 0, 1$ and 2.

The above formula allows a computation for arbitrary l and m. For example,

$$\begin{aligned} q_3^3 &= \frac{1}{24}\sqrt{\frac{7}{5\pi}}\left(\mathrm{i}Q^{(3)}_{030} - 3Q^{(3)}_{120} - 3\mathrm{i}Q^{(3)}_{210} + Q^{(3)}_{300}\right) \\ &= \frac{1}{24}\sqrt{\frac{7}{5\pi}}\left(\mathrm{i}Q^{(3)}_{222} - 3Q^{(3)}_{122} - 3\mathrm{i}Q^{(3)}_{112} + Q^{(3)}_{111}\right). \end{aligned}$$

In the last equation we switched from the pqr-notation to conventional tensor indices, namely $Q^{(3)}_{ijk}$.

4.3 BASIS TRANSFORMATION

As already pointed out, we called the expression derived in eq. (4.38) a preliminary basis transformation, because the sum on the right-hand side contained multipole functions with index $p > 1$. But only the multipole functions with $p \leq 1$ are a basis of the space of homogeneous and harmonic polynomials of degree l, see eq. (4.13). To turn eq. (4.38) into an actual basis transformation we have to express all $M^{(l)}_{pqr}$ with $p > 1$ as a linear combination of the multipole basis functions.

We can accomplish this by exploiting that the object $M^{(l)}_{i_1\cdots i_l}$ is symmetric and traceless, cf. eq. (4.28). Using the pqr-notation, which we introduced in eq. (4.12), the tracelessness of $M^{(l)}$ is expressed as

$$M^{(l)}_{pqr} + M^{(l)}_{(p-2)(q+2)r} + M^{(l)}_{(p-2)q(r+2)} = 0\,, \tag{4.39}$$

with $p > 1$ and, by definition, $p + q + r = l$. This becomes plausible, when an example is considered, e.g. $l = 5$ and $p = 2, q = 1$ and $r = 2$,

$$0 = M^{(5)}_{212} + M^{(5)}_{032} + M^{(5)}_{014} = M^{(5)}_{11233} + M^{(5)}_{22233} + M^{(5)}_{33233} = \sum_{i=1}^{3} M^{(5)}_{ii233}\,.$$

Note that we used the symmetry of $M^{(l)}_{i_1\cdots i_l}$. An immediate consequence of eq. (4.39) is that we can express a multipole function with index $p > 1$ as a sum of multipole functions with index $p - 2$. These, in turn, can again be expressed as a sum of multipole functions with index $p - 4$. Depending on whether p is even or odd, this chain ends when $p = 0$ or $p = 1$. This computation can be represented in the form of Pascal's triangle, see Fig. 4.1, or, more compactly, in the following formula

$$M^{(l)}_{pqr} = (-1)^{\lfloor p/2\rfloor} \sum_{k=0}^{\lfloor p/2\rfloor} \binom{\lfloor \frac{p}{2} \rfloor}{k} \begin{cases} M^{(l)}_{0(q+p-2k)(r+2k)} & p \text{ is even} \\ M^{(l)}_{1(q+p-(2k+1))(r+2k)} & p \text{ is odd} \end{cases} \tag{4.40}$$

We are able to write multipole functions with $p > 1$ as linear combinations of the multipole basis functions, though we must distinguish between even and

$$
\begin{array}{ccc}
 & M^{(l)}_{pqr} & \\
 & -M^{(l)}_{(p-2)(q+2)r} \quad -M^{(l)}_{(p-2)q(r+2)} & \\
M^{(l)}_{(p-4)(q+4)r} & 2M^{(l)}_{(p-4)(q+2)(r+2)} & M^{(l)}_{(p-4)q(r+4)} \\
 & \vdots &
\end{array}
$$

Figure 4.1: The dependence of the multipole functions with $p > 1$ on functions with smaller p is given by Pascal's triangle. Notice the alternating minus sign.

odd values of p. We, thus, split the sum in eq. (4.38) into a corresponding even and odd part and plug into it the expression for the multipole functions that we just derived. This yields

$$
\begin{aligned}
|\boldsymbol{x}|^l Y_l^m &= a_l^m \left[\sum_{n=0}^{a} \mathrm{i}^{m-2n} \binom{m}{2n} M^{(l)}_{2n(m-2n)(l-m)} \right. \\
&\qquad \left. + \sum_{n=0}^{b} \mathrm{i}^{m-(2n+1)} \binom{m}{2n+1} M^{(l)}_{(2n+1)(m-(2n+1))(l-m)} \right] \\
&= a_l^m \left[\sum_{n=0}^{a} \sum_{k=0}^{n} \mathrm{i}^m \binom{m}{2n} \binom{n}{k} M^{(l)}_{0(m-2k)(l-m+2k)} \right. \\
&\qquad \left. - \sum_{n=0}^{b} \sum_{k=0}^{n} \mathrm{i}^m \binom{m}{2n+1} \binom{n}{k} M^{(l)}_{1(m-(2k+1))(l-m+2k)} \right]
\end{aligned}
$$

The limits of the sums are

$$
a = \lfloor m/2 \rfloor \text{ and } b = \lfloor (m-1)/2 \rfloor . \tag{4.41}
$$

In a last step, we note that we can exchange the two sums. For example, for even p, we find

$$
\begin{aligned}
&\sum_{n=0}^{a} \sum_{k=0}^{n} \mathrm{i}^m \binom{m}{2n} \binom{n}{k} M^{(l)}_{0(m-2k)(l-m+2k)} \\
&\qquad = \sum_{k=0}^{a} \left(\mathrm{i}^m \sum_{n=k}^{a} \binom{m}{2n} \binom{n}{k} \right) M^{(l)}_{0(m-2k)(l-m+2k)} .
\end{aligned}
$$

The same holds true for odd p. We arrive at the final result of this section, namely

$$
|\boldsymbol{x}|^l Y_l^m = \sum_{k=0}^{a} \beta^m_{l,0,k} M^{(l)}_{0(m-2k)(l-m+2k)} + \sum_{k=0}^{b} \beta^m_{l,1,k} M^{(l)}_{1(m-(2k+1))(l-m+2k)} . \tag{4.42}
$$

For $m \geq 0$, the coefficients are

$$\begin{aligned} \beta^m_{l,0,k} &:= \mathrm{i}^m a^m_l \sum_{n=k}^{a} \binom{m}{2n}\binom{n}{k} \quad \text{and} \\ \beta^m_{l,1,k} &:= -\mathrm{i}^{m+1} a^m_l \sum_{n=k}^{b} \binom{m}{2n+1}\binom{n}{k}. \end{aligned} \tag{4.43}$$

The solid harmonics with $m < 0$ can be computed with the formula $Y_l^{-m} = (-1)^m Y_l^{m*}$.

We thus have a way to express an arbitrary solid harmonic of degree l and order m as a linear combination of the multipole *basis* functions with degree l. We call the coefficients $\beta^m_{l,0,k}$ and $\beta^m_{l,1,k}$ the *basis transformation* between the solid harmonics and the multipole basis functions in the space of homogeneous and harmonic polynomials of degree l.

Last, but not least, the complex conjugate of the basis transformation relates the spherical multipole moments with the Cartesian multipole moments. A practical implication of this section's considerations is that it is only necessary two compute $2l + 1$ components of the Cartesian multipole moments, namely the ones corresponding to the multipole basis functions given in eq. (4.13)) instead of the $(l + 2)(l + 1)/2$ components of the symmetric tensor $Q^{(l)}$. A way to reconstruct the *dependent components* of $M^{(l)}$ (or $Q^{(l)}$) will be presented in the next section while deriving the inverse basis transformation.

4.4 INVERSE TRANSFORMATION

We derive the inverse of the basis transformation given in eq. (4.42) by, firstly, collecting the coefficients $\beta^m_{l,0,k}$ and $\beta^m_{l,1,k}$ in a matrix, the *basis transformation matrix* $\boldsymbol{B}$ and, secondly, by inverting this matrix.

How difficult it is to invert a matrix depends on its structure, e.g. if it is a diagonal matrix computing the inverse consists in taking the reciprocal of its elements. The structure of the matrix $\boldsymbol{B}$ simplifies if we use the real spherical harmonics instead of the complex spherical harmonics.

The *real spherical harmonics* are defined as

$$\begin{aligned} \begin{pmatrix} Y_{lm0} & Y_{l00} & Y_{lm1} \end{pmatrix} &:= \begin{pmatrix} Y_l^m & Y_l^0 & Y_l^{-m} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & -\mathrm{i} \\ 0 & \sqrt{2} & 0 \\ (-1)^m & 0 & (-1)^m \mathrm{i} \end{pmatrix} \\ &=: \begin{pmatrix} Y_l^m & Y_l^0 & Y_l^{-m} \end{pmatrix} \boldsymbol{S}. \end{aligned} \tag{4.44}$$

Note that $\boldsymbol{S}$ is unitary, i.e. $\boldsymbol{S}^\dagger \boldsymbol{S} = \boldsymbol{1}$. An explicit expression for the real spherical harmonics is

$$Y_{lms} = N_{lm} P_l^m(\cos\theta)(\delta_{s0} \cos m\varphi + \delta_{s1} \sin m\varphi), \tag{4.45}$$

with $0 \leq m \leq l$. Furthermore, the above definition implies that the normalisation $N_{lm} \coloneqq \sqrt{2-\delta_{m0}} N_l^m$. For a list of the first few real spherical harmonics see Tab. C.1.

Applying the unitary transformation to eq. (4.42) results in the following pattern

$$
\begin{aligned}
&s = 0 : \\
&\text{even } m : \qquad |\boldsymbol{x}|^l Y_{lm0} = \sqrt{2} \sum_{k=0}^{a} \beta_{l,0,k}^m M_{0(m-2k)(l-m+2k)}^{(l)} \\
&\text{odd } m : \qquad |\boldsymbol{x}|^l Y_{lm0} = \sqrt{2} \sum_{k=0}^{b} \beta_{l,1,k}^m M_{1(m-(2k+1))(l-m+2k))}^{(l)} \\
&s = 1 : \\
&\text{even } m : \qquad |\boldsymbol{x}|^l Y_{lm1} = -\sqrt{2}\mathrm{i} \sum_{k=0}^{b} \beta_{l,1,k}^m M_{1(m-(2k+1))(l-m+2k)}^{(l)} \\
&\text{odd } m : \qquad |\boldsymbol{x}|^l Y_{lm1} = -\sqrt{2}\mathrm{i} \sum_{k=0}^{a} \beta_{l,0,k}^m M_{0(m-2k)(l-m+2k)}^{(l)}
\end{aligned}
\tag{4.46}
$$

This pattern has two implications for the structure of the basis transformation matrix $\boldsymbol{B}$. First, the unitary transformation $\boldsymbol{S}$ brings it into a form in which it consist of four blocks which correspond to the above four cases. Secondly, since the limits of the sums a and b decrease with decreasing m, the corresponding real solid harmonic depends on less multipole basis functions. For example, let us consider the case $s = 0$ and even m for $l = 5$. The even values of m are zero, two, and four. The corresponding limits a are zero, one and two respectively. Hence, the real solid harmonic of degree five and order zero depends on one multipole basis functions, the one with order two on two and the last one with order four on three multipole basis functions. This pattern translates into a triangular matrix block. Hence, the basis transformation matrix can be transformed into a matrix, which consists out of four triangular matrix blocks.

If the triangular matrices show up also depends on the ordering of the solid harmonics and the multipole basis functions. For the solid harmonics we pick the ordering that is implicitly assumed in eq. (4.44), namely

$$
|\boldsymbol{x}|^l \boldsymbol{Y}^l \coloneqq |\boldsymbol{x}|^l \begin{pmatrix} Y_l^l & Y_l^{l-1} & \cdots & Y_l^{-l+1} & Y_l^{-l} \end{pmatrix}^{\mathsf{T}} ,
$$

and a possible choice for the ordering of the multipole basis functions is

$$
\boldsymbol{M}^l \coloneqq \begin{pmatrix} M_{0l0}^{(l)} & M_{0(l-1)1}^{(l)} & \cdots & M_{1(l-2)1}^{(l)} & M_{1(l-1)0}^{(l)} \end{pmatrix}^T .
$$

Assembling the matrix $\boldsymbol{B}$ in accordance with these orderings brings eq. (4.42) into the compact form

$$
|\boldsymbol{x}|^l \boldsymbol{Y}^l(\theta, \varphi) = \boldsymbol{B}\boldsymbol{M}^l(\boldsymbol{x}) . \tag{4.47}
$$

This equation can be multiplied with the matrix $\boldsymbol{S}^\mathsf{T}$. On its left-hand side the real solid harmonics will show up and on its right-hand side a matrix with the four blocks that we found in eq. (4.46). If we now re-order the real solid harmonics (and the multipole basis functions) such that the harmonics with even order m (and odd order m) are grouped together, four triangular matrices result. If we additionally reverse the order of the real solid harmonics with $s = 1$, we get four *upper* triangular matrices. Let $\boldsymbol{P}$ denote the corresponding permutation matrix, then

$$|\boldsymbol{x}|^l \boldsymbol{P}\boldsymbol{S}^\mathsf{T}\boldsymbol{Y}^l = \boldsymbol{P}\boldsymbol{S}^\mathsf{T}\boldsymbol{B}\boldsymbol{P}^\mathsf{T}\boldsymbol{P}\boldsymbol{M}^l\,.$$

An actual computation for even l shows the following matrix structure

$$\tilde{\boldsymbol{B}} \coloneqq \boldsymbol{P}\boldsymbol{S}^\mathsf{T}\boldsymbol{B}\boldsymbol{P}^\mathsf{T} = \begin{pmatrix} \boldsymbol{U}_1 & & & \boldsymbol{U}_2 \\ & & \boldsymbol{U}_3 & \\ & \boldsymbol{U}_4 & & \end{pmatrix}. \tag{4.48}$$

For odd l, the four upper triangular matrices are at other positions in the matrix.

The inversion of this matrix is equivalent to inverting four upper triangular matrices. This can, for example, be done by applying repeatedly *back-substitution* with unit vectors as right-hand sides, see Press et al. (2007, Sec.2.2.1). We define the *inverse basis transformation matrix*

$$\boldsymbol{A} \coloneqq \boldsymbol{B}^{-1} = \boldsymbol{P}^T\tilde{\boldsymbol{B}}^{-1}\boldsymbol{P}\boldsymbol{S}^\mathsf{T}\,. \tag{4.49}$$

Thus, $\boldsymbol{M}^l = |\boldsymbol{x}|^l \boldsymbol{A}\boldsymbol{Y}^l$ and with the help of a suitable mapping between the matrix indices and the p, q, r indices, we get

$$M^{(l)}_{0qr} = \sum_{m=-l}^{l} \alpha^m_{l,0,q,r} r^l Y^m_l \quad \text{and} \quad M^{(l)}_{1qr} = \sum_{m=-l}^{l} \alpha^m_{l,1,q,r} r^l Y^m_l\,. \tag{4.50}$$

We call the $\alpha^m_{l,0,q,r}$ and $\alpha^m_{l,1,q,r}$ the *inverse basis transformation*.

The *dependent multipole functions*, i.e. the functions with index $p > 1$, can be expressed as a linear combination of the multipole *basis* functions, see eq. (4.40). Since the multipole *basis* functions are $\boldsymbol{M}^l = |\boldsymbol{x}|^l \boldsymbol{A}\boldsymbol{Y}^l$, it is possible to obtain an expression for the dependent multipole functions in terms of the solid harmonics by adding the rows of the inverse basis transformation matrix while multiplying them with the factor given in eq. (4.40).

The definitions of the Cartesian and the spherical multipole moments, given in eq. (4.6) and (4.17), imply that the relation between them is given by the complex conjugate of the inverse basis transformation matrix, i.e.

$$\boldsymbol{q}^l = \boldsymbol{B}^*\boldsymbol{Q}^l \quad \text{and} \quad \boldsymbol{Q}^l = \boldsymbol{A}^*\boldsymbol{q}^l\,. \tag{4.51}$$

Last, but not least, instead of inverting the basis transformation matrix $\boldsymbol{B}$, it is possible to directly compute the inverse basis transformation matrix $\boldsymbol{A}$. A derivation of the formulae for the αs is provided in Appendix A.3.

4.5 APPLICATION TO THE VFP EQUATION

In this section we come back to the original problem, namely the relation between the expansion coefficients of the Cartesian tensor and spherical harmonic expansion of the distribution function f. As explained in the introduction to this chapter, the relation between the two kinds of multipole moments is the same as the relation between the expansion coefficients, i.e. a basis transformation. The reason was that in both cases, i.e. in the case of the multipole expansions and in the case of the expansions of f, for an arbitrary but fixed value of $|\boldsymbol{x}|$ or $t, \boldsymbol{x}, |\boldsymbol{p}|$ respectively, the Cartesian and the spherical harmonic expansion represent equivalent functions on the sphere.

However, in the Cartesian and spherical multipole expansion appear numerical factors that do not appear in the expansions of the distribution function f, see eq. (4.8) and eq. (4.16). That there are no factors in the expansions of f means that $4\pi/(2l+1)$ was included in the definition of f_l^m and $1/l!$ was included in the definition of $F^{(l)}_{i_1\cdots i_l}$.[2] Thus, the relation between the expansion coefficients is given by a modified version of eq. (4.51), i.e.

$$\boldsymbol{f}^l = \boldsymbol{D}_1 \boldsymbol{B}^* \boldsymbol{D}_2 \boldsymbol{F}^l \,, \tag{4.52}$$

where we replaced $\boldsymbol{q}^l$ with $\boldsymbol{f}^l$, $\boldsymbol{Q}^l$ with $\boldsymbol{F}^l$, and multiplied with $\boldsymbol{D}_1 \coloneqq 4\pi/(2l+1)\mathbf{1}$ and $\boldsymbol{D}_2 \coloneqq l!\mathbf{1}$ to include the factors in the definitions of the expansion coefficients. The inverse transformation is obtained by computing the inverse of the above matrix as explained in Section 4.4.

Johnston (1960) derives this inverse basis transformation for $l = 2$ and $l = 3$. But when expanding f, he uses the real spherical harmonics without normalisation N_{lm} and without including $(-1)^m$ in the definition of the associated Legendre Polynomials. Thus he implicitly starts with the following equation

$$\begin{aligned} f(t, \boldsymbol{x}, \boldsymbol{p}) &= \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{s=0}^{1} (-1)^m \frac{2}{1+\delta_{0m}} \frac{(l-m)!}{(l+m)!} \frac{\bar{f}_{lms}(t, \boldsymbol{x}, |\boldsymbol{p}|)}{N_{lm}} Y_{lms}(\theta, \varphi) \\ &\coloneqq \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{s=0}^{1} f_{lms}(t, \boldsymbol{x}, |\boldsymbol{p}|) Y_{lms}(\theta, \varphi) \,. \end{aligned} \tag{4.53}$$

The fraction with factorials appears because the numerical factor in the addition theorem changes if (real) spherical harmonics without normalisation are used to derive it, cf. eq. (4.15). Moreover, $1/N_{lm}$ removes the normalisation. And in the last definition of eq. (4.53) all these factors are included in the coefficients f_{lms}. Thus, we should be able to reproduce his results by inverting the matrix on the right-hand side of

$$\mathbf{f}_{\mathrm{J}}^l = \boldsymbol{C}\boldsymbol{D}_3 \boldsymbol{N}^{-1} \boldsymbol{S}^\dagger \boldsymbol{B}^* \boldsymbol{D}_2 \mathbf{F}^l \,,$$

[2] Maybe it is more correct to say that $4\pi/(2l+1)$ and $1/l!$ were included in the definitions of q_l^m and $Q^{(l)}$ respectively.

which yields the following matrix

$$\boldsymbol{A}_{\mathrm{J}} = \boldsymbol{D}_2^{-1}\boldsymbol{A}^*\boldsymbol{S}\boldsymbol{D}_3^{-1}\boldsymbol{C}\,.$$

Where $\boldsymbol{A}^* = \boldsymbol{B}^{*-1}$ and where we introduced the following definitions

$$\begin{aligned}
\mathbf{f}_{\mathrm{J}}^{l} &\coloneqq (f_{ll0}\, f_{l(l-1)0} \cdots f_{l(l-1)1}\, f_{ll1}) \\
\boldsymbol{N} &\coloneqq \operatorname{diag}(N_{ll}\, N_{l(l-1)} \cdots N_{l(l-1)}\, N_{ll}) \\
\boldsymbol{C} &\coloneqq \operatorname{diag}\left((-1)^l\,(-1)^{l-1} \cdots (-1)^{l-1}\,(-1)^l\right) \text{ and} \\
\boldsymbol{D}_3 &\coloneqq \operatorname{diag}\left(\frac{2}{(2l)!}\,\frac{2}{(2l-1)!} \cdots 1 \cdots \frac{2}{(2l-1)!}\,\frac{2}{(2l)!}\right).
\end{aligned}$$

Furthermore, $\boldsymbol{S}$ is the unitary transformation between the complex spherical harmonics and the real spherical harmonics, see eq. (4.45). The complex conjugate is necessary, because we are using the complex conjugate of eq. (4.47) (namely we are using Y_l^{m*} and compare with eq. (4.51)). The dependent components of $F^{(l)}_{i_1\cdots i_l}$ can be computed as described at the end of Section 4.4. The entries of the matrix $\boldsymbol{A}_{\mathrm{J}}$ can be compared with eq. (9a-d) in Johnston (1960). Moreover, the entries of $\boldsymbol{A}^*$ can be directly computed with formulae derived in Appendix A.3.

Last, but not least, the Vlasov–Fokker–Planck equation and the corresponding expansions of f can be considered as an example partial differential equation (PDE). We note that whenever (Laplace's) spherical harmonics are used to expand a solution to a PDE, it is possible to work with the multipole functions instead (or vice versa). If the multipole functions are used, it is always enough to compute the $2l + 1$ independent components, namely the multipole basis functions.

CHAPTER 5

PARTIAL DIFFERENTIAL EQUATIONS FOR THE EXPANSION COEFFICIENTS

We note that the material presented in this chapter is published in Schween and Reville (2024).

In the last chapter we saw that an expansion of the single particle distribution function reduces the number of dependent variables from six to four, though it requires a procedure to determine a transport equation for each expansion coefficient. While derivations of this system of equations can be found in the literature for different applications, both for the Cartesian tensor expansion (see Johnston, 1960; Thomas et al., 2012; Williams & Jokipii, 1991, truncated at low order) and the spherical harmonic expansion (e.g. Bell et al., 2006; Reville & Bell, 2013; Tzoufras et al., 2011), the resulting equations are cumbersome, and the physical meaning is obfuscated.

In this chapter we present a simplification of the derivation of the system of partial differential equations for the spherical harmonic expansion of f. The proposed method avoids lengthy algebraic manipulations when expanding to high order. We begin our derivation with plugging the spherical harmonic expansion, given in eq. (4.2), into the semi-relativistic VFP equation (3.60) that we derived at the end of Section 3.2.3. Furthermore, we use the isotropic scattering operator, see eq. (3.37), that we derived in Section 3.2.2. As pointed out earlier in eq. (4.20) the eigenfunctions of this scattering operator are the spherical harmonics. This motivates an approach which exploits operators acting on the space of spherical harmonics and their matrix representations.

5.1 OPERATORS AND THE SYSTEM OF PDES

In this section we identify operators that are present in the semi-relativistic VFP equation and show that the system of PDEs for the expansion coefficients can then be obtained by replacing the operators with their matrix representations. This has the advantage of retaining the original structure of the VFP equation. Additionally, we explicitly compute the matrix elements of the matrix representations.

To this end we convert the VFP equation (3.60) into a matrix equation by

inserting the spherical harmonic expansion (4.2) and projecting onto the space of spherical harmonics with the help of the scalar product defined in eq. (4.22). This yields

$$\begin{aligned}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}&\left(Y_{l'}^{m'}\,\middle|\,Y_l^m\right)\frac{\partial f_l^m}{\partial t}+\left(Y_{l'}^{m'}\,\middle|\,(\boldsymbol{U}+\boldsymbol{v})\,Y_l^m\right)\cdot\nabla_x f_l^m\\&-\left(Y_{l'}^{m'}\,\middle|\,\left(\gamma m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}+(\boldsymbol{p}\cdot\nabla_x)\boldsymbol{U}\right)\cdot\nabla_p(f_l^m Y_l^m)\right)\\&+q\left(Y_{l'}^{m'}\,\middle|\,\boldsymbol{v}\cdot\left(\boldsymbol{B}\times\nabla_p(f_l^m Y_l^m)\right)\right)=\frac{\nu}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(Y_{l'}^{m'}\,\middle|\,\Delta_{\theta,\varphi}Y_l^m\right)f_l^m\end{aligned}\tag{5.1}$$

for $l'\in\mathbb{N}_0$ and $|m'|\leq l'$. Notice that all quantities related to momentum, namely $\boldsymbol{v}$, $\boldsymbol{p}$, γ, ν and $\boldsymbol{B}$ are defined in the rest frame of the background plasma, because of the mixed-coordinate system in use. The reason that eq. (5.1) is a matrix equation is that the scalar products can be interpreted as elements of matrices that multiply the vector $(\boldsymbol{f})_{j(l,m)}\coloneqq f_l^m$ or its derivatives. The index j used here is a one-to-one function of the indices l and m. We call this function an *index map* and it determines how the expansion coefficients are ordered. An example order is given through $\boldsymbol{f}=(f_0^0,f_1^1,f_1^0,f_1^{-1},\dots)^T$. In this case the index map j is $j(l,m)\coloneqq l(l+1)-m$, with j starting at zero. Its (unique) inverse is given as $l=\lfloor\sqrt{j}\rfloor$ and $m=l(l+1)-j$.

5.1.1 THE IDENTITY OPERATOR AND THE COLLISION OPERATOR

The time derivative may serve as an example. It can be written as

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(Y_{l'}^{m'}\,\middle|\,Y_l^m\right)\frac{\partial f_l^m}{\partial t}=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(\mathbf{1})_{i(l',m')j(l,m)}\partial_t(\boldsymbol{f})_{j(l,m)}\,,\tag{5.2}$$

where $\mathbf{1}$ is the identity matrix. Note that we introduced a second index map $i(l',m')=l'(l'+1)-m'$ such that the notation resembles the ordinary matrix-vector product. In particular, we emphasise that summing over l and m can be interpreted as the multiplication of the identity matrix with the time derivative of the vector $\boldsymbol{f}$. We remark that each time additional matrix indices are required a 'copy' of the index map i (or j) will be used, see for example the matrix-matrix products in eq. (5.39).

The identity matrix can be considered to be the matrix representation of the identity operator $\hat{1}$ acting on the space of the spherical harmonics. We define the *matrix representation* of an operator $\hat{O}$ acting on an inner product space $\mathcal{H}=\operatorname{span}\{b_1,b_2,\dots\}$ to be

$$(\boldsymbol{O})_{ij}=\left(b_i\,\middle|\,\hat{O}b_j\right)\quad\text{where the } b_i\text{s are the basis vectors of }\mathcal{H}\,.\tag{5.3}$$

We note that the space of all spherical harmonics together with the inner product defined in eq. (4.22) is the Hilbert space $L^2(S^2)$ of all square-integrable

functions on the unit sphere. We denote this space by

$$\mathcal{S} \coloneqq \operatorname{span}\left\{Y_l^m \mid l \in \mathbb{N}_0 \text{ and } |m| \leq l\right\} = \bigoplus_{l=0}^{\infty} \mathcal{H}^l(S^2)\,, \tag{5.4}$$

where the direct sum highlights that a spherical harmonic of degree l is a homogeneous and harmonic polynomial restricted to the unit sphere, cf. (4.21). (Axler et al., 2001, Theorem 5.12, p. 81).

The collision operator is another example for the interpretation of the integrals as matrix elements. The right-hand side of eq. (5.1) is

$$\frac{\nu}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(Y_{l'}^{m'}\,\middle|\,\Delta_{\theta,\varphi}Y_l^m\right)f_l^m \coloneqq -\nu\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(\boldsymbol{C})_{i(l',m')j(l,m)}(\boldsymbol{f})_{j(l,m)}\,, \tag{5.5}$$

where $\boldsymbol{C}$ is the matrix representation of the collision operator in the spherical harmonic basis. It is a diagonal matrix and its elements are easily computed, since the spherical harmonics are eigenfunctions of the angular part of the Laplace operator, see eqs. (3.37) and (4.20). The matrix elements are

$$(\boldsymbol{C})_{i(l',m')j(l,m)} = -\frac{1}{2}\left(Y_{l'}^{m'}\,\middle|\,\Delta_{\theta,\varphi}Y_l^m\right) = \frac{l(l+1)}{2}\delta_{ll'}\delta_{m'm}\,. \tag{5.6}$$

These two examples suggest that the system of PDEs that determines the expansion coefficients f_l^m can be formulated as representation matrices of operators acting on the space of spherical harmonics. We thus focus on identifying other operators whose actions on the spherical harmonics are either known or easily derived.

5.1.2 THE ANGULAR MOMENTUM OPERATOR

The magnetic force term in the VFP equation (3.60) is the momentum-space counterpart of the *angular momentum operator* in configuration space that we presented in eq. (4.32), i.e.,

$$q\boldsymbol{v}\cdot\left(\boldsymbol{B}\times\nabla_p f\right) = -q\boldsymbol{B}\cdot\left(\boldsymbol{v}\times\nabla_p f\right) = -\frac{q\boldsymbol{B}}{\gamma m}\cdot\left(\boldsymbol{p}\times\nabla_p\right)f \eqqcolon -\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{L}f\,, \tag{5.7}$$

where $\boldsymbol{\omega} = q\boldsymbol{B}/\gamma m$ is the angular frequency vector, and $\boldsymbol{L} = -\mathrm{i}\boldsymbol{p}\times\nabla_p$. We choose to keep the name 'angular momentum operator' and the symbol $\boldsymbol{L}$, despite its action on the momentum space variables $\boldsymbol{p}$, because from the mathematical point of view there is no difference.

For example, in terms of the coordinates θ and ϕ,

$$\begin{aligned}&\hat{L}_x \coloneqq -\mathrm{i}\partial_\varphi \quad \text{and}\\ &\hat{L}_\pm \coloneqq \hat{L}_y \pm \mathrm{i}\hat{L}_z = e^{\pm\mathrm{i}\varphi}\left(\pm\frac{\partial}{\partial\theta} + \mathrm{i}\cot\theta\frac{\partial}{\partial\varphi}\right),\end{aligned} \tag{5.8}$$

where the expression for the angular momentum ladder operator $\hat{L}_\pm$ is the same as the one in eq. (5.9).

When we express $\hat{L}_y$ and $\hat{L}_z$ in terms of the ladder operators, we get

$$\boldsymbol{L} = \begin{pmatrix} \hat{L}_x \\ 1/2\left(\hat{L}_+ + \hat{L}_-\right) \\ 1/2\mathrm{i}\left(\hat{L}_+ - \hat{L}_-\right) \end{pmatrix}. \tag{5.9}$$

We emphasise that the angular momentum operator acts only on θ and ϕ and, hence, the magnetic force term in eq. (5.1) becomes

$$\begin{aligned} \int_{S^2} {Y_{l'}^{m'}}^* q\boldsymbol{v} \cdot \left(\boldsymbol{B} \times \nabla_p (f_l^m Y_l^m)\right) \mathrm{d}\Omega &= -\mathrm{i}\boldsymbol{\omega} \cdot \left(Y_{l'}^{m'} \,\middle|\, \boldsymbol{L} Y_l^m \right) f_l^m \\ &=: -\mathrm{i}\omega_a (\boldsymbol{\Omega}^a)_{i(l',m')j(l,m)} (\boldsymbol{f})_{j(l,m)}\,, \end{aligned} \tag{5.10}$$

where we the expansion coefficient f_l^m could be taken out of the integral, because it does only depend on t, $\boldsymbol{x}$ and p and *not* on θ and φ. Furthermore, we introduced the representation matrices $\boldsymbol{\Omega}^a$ of the angular momentum operators. Notice that we implicitly sum over the index $a \in \{x, y, z\}$. We will frequently write $\boldsymbol{\Omega}^a$ to refer to all three representation matrices at once.

The action of the angular momentum operators on the spherical harmonics is well-known from quantum mechanics. For example, in Landau and Lifshitz (1977, §27) their action is given to be

$$\begin{aligned} \hat{L}_x Y_l^m &= m Y_l^m\,, \\ \hat{L}_+ Y_l^m &= \sqrt{(l+m+1)(l-m)}\, Y_l^{m+1} \quad \text{and} \\ \hat{L}_- Y_l^m &= \sqrt{(l+m)(l-m+1)}\, Y_l^{m-1}\,. \end{aligned} \tag{5.11}$$

The matrix representations of the angular momentum operators can now be computed using eq. (5.11) and eq. (5.9). They are

$$(\boldsymbol{\Omega}^x)_{i(l',m')j(l,m)} = m\delta_{l'l}\delta_{m'm}\,, \tag{5.12}$$

$$\begin{aligned} (\boldsymbol{\Omega}^y)_{i(l',m')j(l,m)} = {} & \frac{1}{2}\sqrt{(l+m+1)(l-m)}\,\delta_{l'l}\delta_{m'(m+1)} \\ & + \frac{1}{2}\sqrt{(l+m)(l-m+1)}\,\delta_{l'l}\delta_{m'(m-1)}\,, \end{aligned} \tag{5.13}$$

$$\begin{aligned} (\boldsymbol{\Omega}^z)_{i(l',m')j(l,m)} = {} & -\frac{\mathrm{i}}{2}\sqrt{(l+m+1)(l-m)}\,\delta_{l'l}\delta_{m'(m+1)} \\ & + \frac{\mathrm{i}}{2}\sqrt{(l+m)(l-m+1)}\,\delta_{l'l}\delta_{m'(m-1)}\,. \end{aligned} \tag{5.14}$$

Alternatively, these matrix elements can also be found in Varshalovich et al. (1988, Chapter 13.2, eq. 42).

5.1.3 THE DIRECTION OPERATORS

In addition to the identity and angular momentum operators, another set of operators appears in the VFP equation. We call them *direction operators* because they 'point' in the direction of the coordinate axes:

$$\boldsymbol{v} = v \begin{pmatrix} \cos\theta \\ \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \end{pmatrix} =: v \begin{pmatrix} \hat{A}^x \\ \hat{A}^y \\ \hat{A}^z \end{pmatrix}. \tag{5.15}$$

These operators act on the space of spherical harmonics, i.e. $\hat{A}^a : \mathcal{S} \to \mathcal{S}$. They appear, for example, in the spatial advection term of the VFP equation. In the equations for the expansion coefficients (5.1), the spatial advection term is

$$\begin{aligned} &\int_{S^2} Y_{l'}^{m'\,*} \left(\boldsymbol{U} + \boldsymbol{v}\right) Y_l^m \,\mathrm{d}\Omega \cdot \nabla_x f_l^m \\ &\quad = \left(U^a \left(Y_{l'}^{m'} \,\middle|\, Y_l^m \right) + v \left(Y_{l'}^{m'} \,\middle|\, \hat{A}^a Y_l^m \right) \right) \frac{\partial f_l^m}{\partial x^a} \\ &\quad =: \left(U^a (\mathbf{1})_{i(l',m')j(l,m)} + v (\boldsymbol{A}^a)_{i(l',m')j(l,m)} \right) \partial_{x_a} (\boldsymbol{f})_{j(l,m)} . \end{aligned} \tag{5.16}$$

As in the case of the angular momentum operator, the computation of the matrix representations of the $\hat{A}^a$ operators leads us to investigate their action on the spherical harmonics. There are multiple ways to determine their action. For example, it is possible to use recurrence relations for the associated Legendre polynomials or to use ladder operators, which increase/decrease l and/or m and that be can be used to represent them.

We begin with $\hat{A}^x = \cos\theta$ and we use a recurrence relation to examine its action, namely

$$\cos\theta P_l^m(\cos\theta) = \frac{l-m+1}{2l+1} P_{l+1}^m(\cos\theta) + \frac{l+m}{2l+1} P_{l-1}^m(\cos\theta) . \tag{5.17}$$

With the help of this recurrence relation, we show that the action of $\hat{A}^x$ is

$$\begin{aligned} \hat{A}^x Y_l^m &= \frac{l-m+1}{2l+1} \frac{N_l^m}{N_{l+1}^m} N_{l+1}^m P_{l+1}^m e^{\mathrm{i}m\varphi} + \frac{l+m}{2l+1} \frac{N_l^m}{N_{l-1}^m} N_{l-1}^m P_{l-1}^m e^{\mathrm{i}m\varphi} \\ &= \sqrt{\frac{(l+m+1)(l-m+1)}{(2l+3)(2l+1)}} Y_{l+1}^m + \sqrt{\frac{(l+m)(l-m)}{(2l+1)(2l-1)}} Y_{l-1}^m . \end{aligned} \tag{5.18}$$

Instead of using a recurrence relation for the associated Legendre polynomials, we can work with ladder operators which increase (decrease) l. These operators are investigated in Fakhri (2016, eq. 12) and they are defined as

$$\hat{J}_\pm(l) := \pm \sin\theta + l\cos\theta . \tag{5.19}$$

With this definition at hand, a short computation shows that

$$\hat{A}^x = \frac{1}{2l+1} \left(\hat{J}_+(l+1) + \hat{J}_-(l) \right) . \tag{5.20}$$

The action of $\hat{A}^x$ is now given by the action of the ladder operators $\hat{J}_\pm$, which can as well be found in Fakhri (2016, eq. 14 a, b), and it is

$$\hat{J}_+(l+1)Y_l^m = \sqrt{\frac{2l+1}{2l+3}(l-m+1)(l+m+1)}Y_{l+1}^m \quad \text{and} \tag{5.21}$$

$$\hat{J}_-(l)Y_l^m = \sqrt{\frac{2l+1}{2l-1}(l-m)(l+m)}Y_{l+1}^m\,. \tag{5.22}$$

For the sake of completeness, we proceed analogously with the operator $\hat{A}^y$ and $\hat{A}^z$, i.e. we, first, compute their action using a suitable recurrence relation for the associated Legendre polynomials and, secondly, we express their action with the help of appropriate ladder operators. But already at this point, we highlight that an explicit computation of the representation matrices of $\hat{A}^y$ and $\hat{A}^z$ is not necessary, as they are connected via rotations (see section 5.2.2).

To compute the action of $\hat{A}^y$ and $\hat{A}^z$, we use the two recurrence relations

$$\begin{aligned}\sin\theta P_l^m &= \frac{1}{2l+1}\left[(l-m+1)(l-m+2)P_{l+1}^{m-1} - (l+m-1)(l+m)P_{l+1}^{m-1}\right]\\ \sin\theta P_l^m &= \frac{-1}{2l+1}\left[P_{l+1}^{m+1} - P_{l-1}^{m+1}\right]\,.\end{aligned} \tag{5.23}$$

With these relations at hand, we show that the action of $\hat{A}^y$ is

$$\begin{aligned}\hat{A}^y Y_l^m &= \sin\theta\cos\varphi Y_l^m = \frac{1}{2}\sin\theta\left(e^{\mathrm{i}\varphi} + e^{-\mathrm{i}\varphi}\right)Y_l^m\\ &= \frac{1}{2}N_l^m\left(\sin\theta P_l^m e^{\mathrm{i}(m+1)\varphi} + \sin\theta P_l^m e^{\mathrm{i}(m-1)\varphi}\right)\\ &= \frac{1}{2}\left(-a_{l+1}^{m+1}Y_{l+1}^{m+1} + a_l^{-m}Y_{l-1}^{m+1} + a_{l+1}^{-m+1}Y_{l+1}^{m-1} - a_l^m Y_{l-1}^{m-1}\right)\,,\end{aligned} \tag{5.24}$$

where the coefficient a_l^m is

$$a_l^m \coloneqq \sqrt{\frac{(l+m)(l+m-1)}{(2l+1)(2l-1)}}\,. \tag{5.25}$$

In a completely analogous manner the action of $\hat{A}^z$ can be shown to be

$$\begin{aligned}\hat{A}^z Y_l^m &= \sin\theta\sin\varphi Y_l^m = \frac{1}{2\mathrm{i}}\sin\theta\left(e^{\mathrm{i}\varphi} - e^{-\mathrm{i}\varphi}\right)Y_l^m\\ &= \frac{\mathrm{i}}{2}\left(a_{l+1}^{m+1}Y_{l+1}^{m+1} - a_l^{-m}Y_{l-1}^{m+1} + a_{l+1}^{-m+1}Y_{l+1}^{m-1} - a_l^m Y_{l-1}^{m-1}\right)\,.\end{aligned} \tag{5.26}$$

Again, the action of $\hat{A}^y$ and $\hat{A}^z$ can alternatively be derived with ladder operators that can be found in Fakhri (2016, eq. 25, eq. 34), i.e.

$$\hat{A}_\pm^\pm(l) \coloneqq \pm[\hat{L}_\pm, \hat{J}_\pm(l)] = e^{\pm\mathrm{i}\varphi}\left(\pm\cos\theta\frac{\partial}{\partial\theta} + \frac{\mathrm{i}}{\sin\theta}\frac{\partial}{\partial\varphi} - l\sin\theta\right)\,, \tag{5.27}$$

$$\hat{A}_\mp^\pm(l) \coloneqq \mp[\hat{L}_\pm, \hat{J}_\mp(l)] = e^{\pm\mathrm{i}\varphi}\left(\pm\cos\theta\frac{\partial}{\partial\theta} + \frac{\mathrm{i}}{\sin\theta}\frac{\partial}{\partial\varphi} + l\sin\theta\right)\,. \tag{5.28}$$

The square brackets denote the commutator of the two operators. These operators shift l and m either in the same 'direction' or in opposite 'directions'. Their action is described by Fakhri (2016, eq. 26 a, b, eq. 35 a, b) and it is

$$\begin{aligned}
\hat{A}_+^+(l+1)Y_l^m &= \sqrt{\frac{2l+1}{2l+3}(l+m+2)(l+m+1)}Y_{l+1}^{m+1} \\
\hat{A}_-^-(l)Y_l^m &= \sqrt{\frac{2l+1}{2l-1}(l+m)(l+m-1)}Y_{l-1}^{m-1} \\
\hat{A}_-^+(l)Y_l^m &= \sqrt{\frac{2l+1}{2l-1}(l-m)(l-m-1)}Y_{l-1}^{m+1} \\
\hat{A}_+^-(l+1)Y_l^m &= \sqrt{\frac{2l+1}{2l+3}(l-m+2)(l-m+1)}Y_{l+1}^{m-1}\,.
\end{aligned} \tag{5.29}$$

The operators $\hat{A}^y$ and $\hat{A}^z$ can be represented with the help of the above ladder operators, i.e.

$$\hat{A}^y = \frac{1}{2(2l+1)}\left(-\hat{A}_+^+(l+1) + \hat{A}_-^+(l) + \hat{A}_+^-(l) - \hat{A}_-^-(l+1)\right) \tag{5.30}$$

$$\hat{A}^z = \frac{i}{2(2l+1)}\left(\hat{A}_+^+(l+1) - \hat{A}_-^+(l) + \hat{A}_+^-(l) - \hat{A}_-^-(l+1)\right). \tag{5.31}$$

This can be verified by direct computation.

Knowing the action of the direction operators, see eq. (5.18), (5.24) and (5.26), we can compute their matrix representations, namely

$$\begin{aligned}
&(\boldsymbol{A}^x)_{i(l',m')j(l,m)} \\
&\quad= \sqrt{\frac{(l+m+1)(l-m+1)}{(2l+3)(2l+1)}}\delta_{l'(l+1)}\delta_{m'm} + \sqrt{\frac{(l+m)(l-m)}{(2l+1)(2l-1)}}\delta_{l'(l-1)}\delta_{m'm}
\end{aligned} \tag{5.32}$$

and

$$\begin{aligned}
&(\boldsymbol{A}^y)_{i(l',m')j(l,m)} \\
&\quad= \frac{1}{2}\left(-a_{l+1}^{m+1}\delta_{l'(l+1)}\delta_{m'(m+1)} + a_l^{-m}\delta_{l'(l-1)}\delta_{m'(m+1)}\right. \\
&\qquad\left. + a_{l+1}^{-m+1}\delta_{l'(l+1)}\delta_{m'(m-1)} - a_l^{m}\delta_{l'(l-1)}\delta_{m'(m-1)}\right)
\end{aligned} \tag{5.33}$$

$$\begin{aligned}
&(\boldsymbol{A}^z)_{i(l',m')j(l,m)} \\
&\quad= \frac{\mathrm{i}}{2}\left(+a_{l+1}^{m+1}\delta_{l'(l+1)}\delta_{m'(m+1)} - a_l^{-m}\delta_{l'(l-1)}\delta_{m'(m+1)}\right. \\
&\qquad\left. + a_{l+1}^{-m+1}\delta_{l'(l+1)}\delta_{m'(m-1)} - a_l^{m}\delta_{l'(l-1)}\delta_{m'(m-1)}\right).
\end{aligned} \tag{5.34}$$

These matrix elements can also be found in Varshalovich et al. (1988, Chapter 13.2, eq. 14-16).

5.1.4 PRODUCTS OF OPERATORS

The only term in the VFP equation that cannot immediately be formulated using the identity, angular momentum and direction operators is

$$-\left(\gamma m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}+(\boldsymbol{p}\cdot\nabla_x)\boldsymbol{U}\right)\cdot\nabla_p f\,. \tag{5.35}$$

This term is a consequence of the transformation of $\boldsymbol{p}$ to the rest frame of the background plasma which is why it is referred to as the *fictitious force term*, cf. eq. (3.59).

We show now that ∇_p can be expressed in terms of the direction operators $\hat{A}^a$ and the angular momentum operators $\hat{L}^a$:

$$\nabla_p=\boldsymbol{e}_p\frac{\partial}{\partial p}+\frac{1}{p}\left(\boldsymbol{e}_\theta\frac{\partial}{\partial\theta}+\boldsymbol{e}_\varphi\frac{1}{\sin\theta}\frac{\partial}{\partial\varphi}\right)=\boldsymbol{e}_p\frac{\partial}{\partial p}-\mathrm{i}\frac{1}{p}\boldsymbol{e}_p\times\boldsymbol{L}\,. \tag{5.36}$$

Here, $\boldsymbol{e}_p,\boldsymbol{e}_\theta$ and $\boldsymbol{e}_\varphi$, are as before the unit vectors of the spherical coordinate system, and we emphasise that the components of $\boldsymbol{e}_p$ are just the direction operators:

$$\boldsymbol{e}_p=(\cos\theta,\sin\theta\cos\varphi,\sin\theta\sin\varphi)^T=(\hat{A}^x,\hat{A}^y,\hat{A}^z)^{\mathsf{T}}\,. \tag{5.37}$$

Note that expressing ∇_p with the help of the direction operators and the angular momentum operators comes at the cost of introducing products of operators. Up to now, we saw that it was possible to derive the system of equations for the expansion coefficients by simply replacing the operators with their respective matrix representations. We now prove that this is still true for products of operators if some care is taken.

First of all, we show that the matrix representation of a product of operators is the product of the matrix representations of the operators. To see this, we point out that the identity operator in the space of the spherical harmonics can be written as

$$g=\hat{1}g=\sum_{n=0}^{\infty}\sum_{k=-n}^{n}Y_n^k\left(Y_n^k\,|\,g\right)\qquad\text{for all } g\in\mathcal{S}\,. \tag{5.38}$$

If $g\in\mathcal{S}$, then it can be written as a linear combination of spherical harmonics and the scalar products in the above sum yield the coefficients of this linear combination.[1]

For the matrix representation of the product of two operators, say $\hat{A}$ and $\hat{B}$,

[1] A more familiar notation for the identity operator might be the Bra-Ket notation of quantum mechanics, namely $\hat{1}=\sum_{n,k}|nk><nk|$.

we get

$$\begin{aligned}\left(Y_{l'}^{m'}\,\middle|\,\hat{A}\hat{B}Y_l^m\right) &= \left(Y_{l'}^{m'}\,\middle|\,\hat{A}\hat{1}\hat{B}Y_l^m\right)\\ &= \left(Y_{l'}^{m'}\,\middle|\,\hat{A}\sum_{n=0}^{\infty}\sum_{k=-n}^{n}Y_n^k\left(Y_n^k\,|\,\hat{B}Y_l^m\right)\right)\\ &= \sum_{n=0}^{\infty}\sum_{k=-n}^{n}\left(Y_{l'}^{m'}\,\middle|\,\hat{A}Y_n^k\right)\left(Y_n^k\,|\,\hat{B}Y_l^m\right)\\ &= \sum_{n=0}^{\infty}\sum_{k=-n}^{n}(\boldsymbol{A})_{i(l',m')h(n,k)}(\boldsymbol{B})_{h(n,k)j(l,m)}\,.\end{aligned} \tag{5.39}$$

Hence, the matrix representations of the two operators are multiplied.

We emphasise that the above statement does *not* hold if we truncate the expansion of the distribution function at a finite $l_{\max}$. In this case, we compute the matrix representations of the operators acting on the finite dimensional space $\mathcal{S}^{l_{\max}} \coloneqq \operatorname{span}\{Y_0^0, Y_1^1, Y_1^0, Y_1^{-1}, \dots, Y_{l_{\max}}^{-l_{\max}}\}$ and *not* on the infinite dimensional space $\mathcal{S}$. The $\hat{A}^a$ operators contain ladder operators which increase (or decrease) l, see eq. (5.20), (5.30) and (5.31). It is necessary to define what happens if these operators act on $Y_{l_{\max}}^m$. Since $Y_{l_{\max+1}}^m$ is not an element of the space $\mathcal{S}^{l_{\max}}$, it makes sense to define that the result is zero when they act on $Y_{l_{\max}}^m$. But if a product of operators appears whose first operator increases l, while the second operator decreases it again, then the above definition would imply that this product yields zero when it acts on $Y_{l_{\max}}^m$. This is not the correct result, because the joint action results in a constant times $Y_{l_{\max}}^m$, which can be represented in $\mathcal{S}^{l_{\max}}$. This difficulty can be avoided, if we compute the matrix representations of the operators for $L = l_{\max} + 1$, evaluate the products and, subsequently, reduce the matrices to $l_{\max}$.

Since we now know how to compute the matrix representations of a product of operators, we can now look at eq. (5.1) and see how the fictitious force term

contributes to the system of equations for the expansion coefficients, i.e.

$$
\begin{aligned}
&-\int_{S^2} Y_{l'}^{m'\,*}\left(\gamma m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}+(\boldsymbol{p}\cdot\nabla_x)\boldsymbol{U}\right)\cdot\nabla_p(f_l^m Y_l^m)\,\mathrm{d}\Omega \qquad (5.40)\\
&\quad=-\int_{S^2} Y_{l'}^{m'\,*}\left(\gamma m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}+(\boldsymbol{p}\cdot\nabla_x)\boldsymbol{U}\right)\cdot\left(\boldsymbol{e}_p\partial_p-\frac{\mathrm{i}}{p}\boldsymbol{e}_p\times\boldsymbol{L}\right)f_l^m Y_l^m\,\mathrm{d}\Omega\\
&\quad=-\gamma m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}\cdot\left(Y_{l'}^{m'}\,\middle|\,\boldsymbol{e}_p Y_l^m\right)\frac{\partial f_l^m}{\partial p}+\frac{\mathrm{i}}{\upsilon}\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}\cdot\left(Y_{l'}^{m'}\,\middle|\,\boldsymbol{e}_p\times\boldsymbol{L}Y_l^m\right)f_l^m\\
&\qquad-p\left(Y_{l'}^{m'}\,\middle|\,(\boldsymbol{e}_p\cdot\nabla_x)\boldsymbol{U}\cdot\boldsymbol{e}_p Y_l^m\right)\frac{\partial f_l^m}{\partial p}\\
&\qquad+\mathrm{i}\left(Y_{l'}^{m'}\,\middle|\,(\boldsymbol{e}_p\cdot\nabla_x)\boldsymbol{U}\cdot(\boldsymbol{e}_p\times\boldsymbol{L})Y_l^m\right)f_l^m\\
&\quad=-\gamma m\frac{\mathrm{D}U_a}{\mathrm{D}t}\left(Y_{l'}^{m'}\,\middle|\,\hat{A}^a Y_l^m\right)\frac{\partial f_l^m}{\partial p}+\frac{\mathrm{i}}{\upsilon}\epsilon_{abc}\frac{\mathrm{D}U^a}{\mathrm{D}t}\left(Y_{l'}^{m'}\,\middle|\,\hat{A}^b\hat{L}^c Y_l^m\right)f_l^m\\
&\qquad-p\frac{\partial U_b}{\partial x^a}\left(Y_{l'}^{m'}\,\middle|\,\hat{A}^a\hat{A}^b Y_l^m\right)\frac{\partial f_l^m}{\partial p}+\mathrm{i}\epsilon_{bcd}\frac{\partial U^b}{\partial x^a}\left(Y_{l'}^{m'}\,\middle|\,\hat{A}^a\hat{A}^c\hat{L}^d Y_l^m\right)f_l^m\\
&\quad=-\gamma m\frac{\mathrm{D}U_a}{\mathrm{D}t}(\boldsymbol{A}^a)_{i(l',m')j(l,m)}\partial_p\boldsymbol{f}_{j(l,m)}\\
&\qquad+\sum_{n,k}\frac{\mathrm{i}}{\upsilon}\epsilon_{abc}\frac{\mathrm{D}U^a}{\mathrm{D}t}(\boldsymbol{A}^b)_{i(l',m')h(n,k)}(\boldsymbol{\Omega}^c)_{h(n,k)j(l,m)}\boldsymbol{f}_{j(l,m)}\\
&\qquad-\sum_{n,k}p\frac{\partial U_b}{\partial x^a}(\boldsymbol{A})^a_{i(l',m')h(n,k)}(\boldsymbol{A})^b_{h(n,k)j(l,m)}\partial_p\boldsymbol{f}_{j(l,m)}\\
&\qquad+\sum_{n,k}\sum_{n',k'}\mathrm{i}\epsilon_{bcd}\frac{\partial U^b}{\partial x^a}(\boldsymbol{A})^a_{i(l',m')h(n,k)}(\boldsymbol{A})^c_{h(n,k)r(n',k')}(\boldsymbol{\Omega})^d_{r(n',k')j(l,m)}\boldsymbol{f}_{j(l,m)}\,.
\end{aligned}
$$

We stress that it is exactly this computation where the proposed operator method provides the most obvious benefit. The repeated application of recurrence relations for the associated Legendre polynomials is replaced with the evaluation of matrix products, which avoids lengthy and cumbersome calculations, see Reville and Bell (2013, Appendix A).

5.1.5 THE COMPLETE SYSTEM OF EQUATIONS

We conclude this section with a presentation of the complete system of PDEs that determines the expansion coefficients f_l^m and an explanation of its terms. The system of equations is

$$
\begin{aligned}
&\partial_t\boldsymbol{f}+\left(U^a\mathbf{1}+\upsilon\boldsymbol{A}^a\right)\partial_{x_a}\boldsymbol{f}-\left(\gamma m\frac{\mathrm{D}U_a}{\mathrm{D}t}\boldsymbol{A}^a+p\frac{\partial U_b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^b\right)\partial_p\boldsymbol{f} \qquad (5.41)\\
&\qquad+\left(\frac{\mathrm{i}}{\upsilon}\epsilon_{abc}\frac{\mathrm{D}U^a}{\mathrm{D}t}\boldsymbol{A}^b\boldsymbol{\Omega}^c+\mathrm{i}\epsilon_{bcd}\frac{\partial U^b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^c\boldsymbol{\Omega}^d\right)\boldsymbol{f}-\mathrm{i}\omega_a\boldsymbol{\Omega}^a\boldsymbol{f}+\nu\boldsymbol{C}\boldsymbol{f}=0\,.
\end{aligned}
$$

The first term is the time derivative of the expansion coefficients. The second models how the expansion coefficients change because of the motion

of the plasma, which is described by $\boldsymbol{U}$, and the motion of the particles given through $\upsilon \boldsymbol{A}^a$. As already stated above, we call this term the spatial advection term. The next term changes the energy contained in the expansion coefficients because of the fictitious force acting on the particles. We called it the momentum advection term, because it advects the coefficients in p-direction. The fourth term reflects that the fictitious force not only accelerates (or decelerates), but it also changes how the energy is distributed among the expansion coefficients. The fifth term represents the effect of the magnetic force, namely the rotation of the velocity of the particles. This rotation translates into a rotation of the expansion coefficients.[2] We refer to this term as the *rotation term*. The last term is modelling the interactions of the particles with the thermal plasma. $\boldsymbol{C}$ is a diagonal matrix, which causes the expansion coefficients with $l \geq 1$ to decay exponentially at the rate $\nu l(l+1)/2$. The coefficients with $l \geq 1$ model the anisotropies of the distribution function and the effect of the exponential decay is to drive the distribution towards isotropy. We call it *collision term*.

We, once more, would like to highlight the one-to-one correspondence between operators and representation matrices. The system of PDEs for the expansion coefficients can be derived by simply replacing the operators with their corresponding matrices.

5.2 ROTATIONS IN THE SPHERICAL HARMONIC SPACE

In this section we show that the operators' matrix representations can be considered 'rotated' versions of each other and thus it is possible to compute any one of them by rotating another. This result is based on the observation that our choice of the coordinate system in momentum space was arbitrary and that the angular momentum operators $\hat{L}^a$ and the direction operators $\hat{A}^a$ are defined with respect to the coordinate axes, see their definitions in eq. (5.8) and (5.15) respectively. Notice that the spherical harmonics are *also* defined with respect to the same coordinate system. Rotating them clockwise by $\pi/2$ about the x-axis is equivalent to defining them in an equally rotated coordinate system. In the rotated system the former y-axis is the z-axis and we expect the representation matrix $\boldsymbol{A}_y$ to have transformed to $\boldsymbol{A}_z$. Taking advantage of this requires us to investigate how to rotate the spherical harmonics and how the operators and their representation matrices change under such a rotation.

[2] For readers familiar with rotations in the space of spherical harmonics (or the theory of angular momentum in quantum mechanics): It can be shown, that $\partial_t f(\boldsymbol{x}, \boldsymbol{p}, t) - \boldsymbol{\omega} \cdot (\boldsymbol{p} \times \nabla_p) f(\boldsymbol{x}, \boldsymbol{p}, t) = 0$ has the solution $f(\boldsymbol{x}, \boldsymbol{p}, t) = e^{\mathrm{i}t\boldsymbol{\omega}\cdot\boldsymbol{L}} f(\boldsymbol{x}, \boldsymbol{p}, 0) = f(\boldsymbol{x}, \boldsymbol{R}(t)\boldsymbol{p}, 0)$. Here, $\boldsymbol{R}(t)$ is a rotation matrix, which rotates $\boldsymbol{p}$ about $\boldsymbol{\omega}/\omega$ at the angular frequency ω. We express f as a series of spherical harmonics and rotating spherical harmonics is equivalent to expressing a spherical harmonic in terms of other spherical harmonics of the same degree l. This is what the $\boldsymbol{\Omega}^a$ matrices, which are representation matrices of the angular momentum operator, do.

5.2.1 THE ROTATION OPERATOR AND ITS MATRIX REPRESENTATION

We begin our investigation with formally introducing a rotation operator in the space of spherical harmonics, i.e. $Y_l^{\tilde{m}}(\theta,\varphi) = \hat{R}Y_l^m(\theta,\varphi)$. $\tilde{m}$ reflects the idea that we can think of the rotated spherical harmonics as being defined with respect to a new polar direction given by a corresponding rotation of the original one. A direct consequence of this thought is that a rotation does not change the degree l of the spherical harmonic.

An explicit form of the rotation operator $\hat{R}$ can be derived by defining its action on the spherical harmonics and by studying an infinitesimal rotation. This is, in essence, an application of representation theory. We present the details here for completeness. In doing so, we follow Jeevanjee (2011), in particular Chapter 5 and example 5.9 therein.

We start the derivation by pointing out that

$$Y_l^m(\theta,\varphi) = N_l^m P_l^m(\cos\theta)e^{\mathrm{i}m\varphi} = N_l^m P_l^m(p^x/p)e^{\mathrm{i}m\arctan(p^z/p^y)} =: Y_l^m(\boldsymbol{p})$$

and define the action of $\hat{R}$ to be

$$Y_l^{\tilde{m}}(\boldsymbol{p}) = \hat{R}Y_l^m(\boldsymbol{p}) \coloneqq Y_l^m(\boldsymbol{R}^{-1}\boldsymbol{p}), \tag{5.42}$$

where $\boldsymbol{R}$ is a rotation matrix which rotates the vector $\boldsymbol{p}$.

We now look at an infinitesimal rotation to derive an explicit expression for the rotation operator. To simplify the discussion, we restrict ourselves to a rotation about the x-axis by an infinitesimally small angle $\mathrm{d}\alpha$, i.e. $\boldsymbol{R}_x(\mathrm{d}\alpha)\boldsymbol{p} = \boldsymbol{p} + \mathrm{d}\alpha\boldsymbol{e}_x \times \boldsymbol{p}$. Its inverse is $\boldsymbol{R}_x^{-1}(\mathrm{d}\alpha) = \boldsymbol{p} - \mathrm{d}\alpha\boldsymbol{e}_x \times \boldsymbol{p}$.

An explicit expression for the rotation operator is now constructed by plugging the infinitesimal rotation into the definition of its action (5.42). This yields

$$\begin{aligned} Y_l^{\tilde{m}}(\boldsymbol{p}) = Y_l^m(\boldsymbol{p} - \mathrm{d}\alpha\boldsymbol{e}_x \times \boldsymbol{p}) &= Y_l^m(\boldsymbol{p}) - \left.\frac{\partial Y_l^m}{\partial p^a}\right|_{\mathrm{d}\alpha=0} \epsilon_{abc}(\boldsymbol{e}_x)^b(\boldsymbol{p})^c\,\mathrm{d}\alpha + \mathcal{O}(\mathrm{d}\alpha^2) \\ &= Y_l^m(\boldsymbol{p}) - (\boldsymbol{p}\times\nabla_p)_x Y_l^m(\boldsymbol{p})\,\mathrm{d}\alpha + \mathcal{O}(\mathrm{d}\alpha^2) \\ &= \left(1 - \mathrm{i}\,\mathrm{d}\alpha\hat{L}^x\right) Y_l^m(\boldsymbol{p}) + \mathcal{O}(\mathrm{d}\alpha^2), \end{aligned}$$

where we Taylor expanded in $\mathrm{d}\alpha$ at zero.

In a next step, we set $\mathrm{d}\alpha = \alpha/n$ and perform n infinitesimal rotations, which add up to a rotation about the x-axis by an angle α. In the limit $n \to \infty$, we get

$$Y_l^{\tilde{m}}(\boldsymbol{p}) = \lim_{n\to\infty}\left(1 - \mathrm{i}\frac{\alpha}{n}\hat{L}^x\right)^n Y_l^m(\boldsymbol{p}) = e^{-\mathrm{i}\alpha\hat{L}^x} Y_l^m(\boldsymbol{p}). \tag{5.43}$$

We state, without further proof, that this generalises to $\hat{R} = e^{-\mathrm{i}\alpha\boldsymbol{r}\cdot\boldsymbol{L}}$ for an arbitrary rotation about the *unit* vector $\boldsymbol{r}$ by an angle α.

The matrix elements of the rotation operator $\hat{R}$ are the *Wigner-D functions* that are well-known in the quantum theory of angular momentum, see Varshalovich et al. (1988, Chapter 4.5, eq. 1). They are defined to be

$$\begin{aligned}(\boldsymbol{U})_{i(l',m')j(l,m)}(\alpha\boldsymbol{r}) &\coloneqq \left(Y_{l'}^{m'} \,\middle|\, e^{-\mathrm{i}\alpha\boldsymbol{r}\cdot\boldsymbol{L}} Y_l^m \right) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(Y_{l'}^{m'} \,\middle|\, (-\mathrm{i}\alpha\boldsymbol{r}\cdot\boldsymbol{L})^k \, Y_l^m \right) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} (-\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a)^k_{i(l',m')j(l,m)} = (e^{-\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a})_{i(l',m')j(l,m)} \,.\end{aligned} \tag{5.44}$$

The conjugate transpose of the Wigner-D matrix is

$$\begin{aligned}\boldsymbol{U}^\dagger = (e^{-\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a})^\dagger &= \sum_{k=0}^{\infty} \frac{1}{k!} (-\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a)^{k\dagger} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} (\mathrm{i}\alpha r_a \boldsymbol{\Omega}^{a\dagger})^k = \sum_{k=0}^{\infty} \frac{1}{k!} (\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a)^k = e^{\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a} \,,\end{aligned} \tag{5.45}$$

where we used that for arbitrary matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ the following two properties hold: Firstly, for $(\boldsymbol{X}\cdots\boldsymbol{X})^\dagger = \boldsymbol{X}^\dagger \cdots \boldsymbol{X}^\dagger$ the order of the matrices does not matter and, second, that $(\boldsymbol{X}+\boldsymbol{Y})^\dagger = \boldsymbol{X}^\dagger + \boldsymbol{Y}^\dagger$. Moreover, we took advantage of the fact that the representation matrices of the angular momentum operator are Hermitian matrices, namely $\boldsymbol{\Omega}^{a\dagger} = \boldsymbol{\Omega}^a$, see eq. (5.64) and (5.65).

A direct and useful consequence is that the Wigner-D matrix is unitary[3], i.e.

$$\boldsymbol{U}^\dagger \boldsymbol{U} = e^{\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a} e^{-\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a} = \mathbf{1} \,. \tag{5.46}$$

5.2.2 ROTATED OPERATORS AND THEIR REPRESENTATION MATRICES

Knowing how to rotate the spherical harmonics, we are in a position to compute the action of an operator $\hat{O}$ in the rotated coordinate frame.

We once more note that we can compute the spherical harmonics defined in the rotated frame simply by rotating the original spherical harmonics in the same manner as the coordinate system.

Before we derive the action of an operator in the rotated coordinate system, we would like to make a technical remark: Transformations of coordinates can be interpreted *actively* or *passively*. An active rotation, for example, yields the coordinates of a rotated vector, whereas a passive rotation gives the coordinates of the vector in a rotated coordinate frame. Active and passive rotations are equivalent if the sign of the rotation angle is changed, i.e. $\boldsymbol{R}^{\mathrm{pas}}(\alpha\boldsymbol{r}) = \boldsymbol{R}^{\mathrm{act}}(-\alpha\boldsymbol{r})$. Even though we write as if we rotated the coordinate system, we actually interpret the rotations as *active*. Instead of working with two different coordinate

[3] In quantum mechanics the unitarity of the rotation operator implies that the probability to measure a specific value of the angular momentum does not change if the coordinate system is rotated.

systems in which the spherical harmonics are defined, we work with two different bases of the space of spherical harmonics, namely the original spherical harmonics and the rotated spherical harmonics. Both are defined in the original coordinate system. The reason to take the active point of view is to allow for a direct application of representation theory without the difficulties from two different sets of coordinates, for example, $\tilde{\boldsymbol{p}}$ in the rotated and $\boldsymbol{p}$ in the original coordinate frame.

That said, let $\tilde{O}$ be the operator in the rotated coordinate system, then $\tilde{O}Y_l^{\tilde{m}}(\theta,\varphi) = \hat{R}\hat{O}\hat{R}^{-1}Y_l^{\tilde{m}}(\theta,\varphi)$; the inverse rotation $\hat{R}^{-1}$ 'brings' the spherical harmonic back into the unrotated coordinate system. The action of $\hat{O}$ on the unrotated spherical harmonic is known. Its result is rotated again and $\tilde{O}$ gives, as expected, results in the rotated coordinate system.

Eventually, we are interested in the representation matrices of operators and, thus, we would like to know how the representation matrix of an operator $\hat{O}$ changes under a rotation of the coordinate system. The representation matrix $\tilde{\boldsymbol{O}}$ in the rotated frame is given by the action of $\hat{O}$ on the rotated spherical harmonics, i.e.

$$\begin{aligned}(\tilde{\boldsymbol{O}})_{i(l',m')j(l,m)} &= \left(Y_{l'}^{\tilde{m}'} \,\middle|\, \hat{O} Y_l^{\tilde{m}} \right) \\ &= \left(\hat{R} Y_{l'}^{m'} \,\middle|\, \hat{O}\hat{R} Y_l^{m} \right) \\ &= \sum_{n,k} \sum_{n',k'} \left(\hat{R} Y_{l'}^{m'} \,\middle|\, Y_{n'}^{k'} \right) \left(Y_{n'}^{k'} \,\middle|\, \hat{O} Y_n^k \right) \left(Y_n^k \,\middle|\, \hat{R} Y_l^m \right) \\ &= \sum_{n,k} \sum_{n',k'} (\boldsymbol{U}^\dagger)_{i(l',m')h(n',k')} (\boldsymbol{O})_{h(n',k')r(n,k)} (\boldsymbol{U})_{r(n,k)j(l,m)} \,,\end{aligned} \tag{5.47}$$

or compactly $\tilde{\boldsymbol{O}} = \boldsymbol{U}^\dagger \boldsymbol{O} \boldsymbol{U}$.

5.2.3 ROTATIONS OF THE ANGULAR MOMENTUM AND DIRECTION OPERATORS' MATRICES

As explained in the introduction to this section, the operator $\hat{A}^y$ has to be $\hat{A}^z$ in a coordinate system which is rotated by $-\pi/2$ about the x-axis. If we know the representation matrix $\boldsymbol{A}^y$, we can compute how it looks like in the rotated frame using the formula in eq. (5.47) and we know that in this frame it has to equal $\boldsymbol{A}^z$. Hence,

$$\boldsymbol{A}^z = e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^x} \boldsymbol{A}^y e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^x} \tag{5.48}$$

Moreover, we compute the representation matrix $\boldsymbol{A}^y$ knowing $\boldsymbol{A}^x$ by rotating the original coordinate system about the z-axis by $-\pi/2$, i.e.

$$\boldsymbol{A}^y = e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z} \boldsymbol{A}^x e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z} \tag{5.49}$$

This leads us to the conclusion that it is enough to compute the representation matrix of $\hat{A}^x$. We highlight that we do *not* have to compute the representation matrices of the involved rotations ourselves. Because their matrix elements

are the Wigner-D functions. They can, for example, be found in Varshalovich et al. (1988, Section 4.5.6 eq. 29, eq. 30 and Section 4.3.1, eq. 2). The rotation about the x-axis and z-axis are $\boldsymbol{U}(-\pi/2\boldsymbol{e}_x)$ and $\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)$ respectively and there matrix elements are

$$\left(e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^x}\right)_{i(l',m')j(l,m)} = \delta_{l'l}\delta_{m'm}e^{\mathrm{i}m\frac{\pi}{2}} \tag{5.50}$$

$$\left(e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\right)_{i(l',m')j(l,m)} = \delta_{l'l}\frac{(-1)^{l-m'}}{2^l} \times \sum_{\substack{k=0 \\ m'+m+k\geq 0}}^{n}(-1)^k \frac{\left[(l+m')!(l-m')!(l+m)!(l-m)!\right]^{1/2}}{k!(l-m'-k)!(l-m-k)!(m'+m+k)!},$$

where the limit of the sum is $n = \min(l-m', l-m)$.

We remark that the representation matrices of the angular momentum operators can as well be obtained by rotating $\boldsymbol{\Omega}^x$, i.e.

$$\boldsymbol{\Omega}^y = e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\boldsymbol{\Omega}^x e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z} \quad \text{and} \quad \boldsymbol{\Omega}^z = e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^x}\boldsymbol{\Omega}^y e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^x}. \tag{5.51}$$

We emphasise that to construct the system of equations (5.41), we only need four matrices, namely $\boldsymbol{A}^x, \boldsymbol{\Omega}^x, \boldsymbol{U}(-\pi/2\boldsymbol{e}_x)$, and $\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)$ and two of them are diagonal matrices. This should considerably diminish the burden of its implementation in future numerical applications.

5.3 A REAL SYSTEM OF EQUATIONS

Having derived and simplified the computation of the system of PDEs for the expansion coefficients, we note that it contains redundant equations. The reason is the reality of the single particle distribution function, as we will show in the next paragraph. In this section we remedy this using the *real* spherical harmonics. The relation between the complex and real spherical harmonics can be expressed with the help of a basis transformation matrix. Multiplying the system of equations with this basis transformation matrix removes the superfluous equations.

The phase-space density f is a real function, i.e $f \in \mathbb{R}$. This implies the existence of a relation between the coefficients f_l^m of the spherical harmonic expansion of f as given in eq. (4.2), because the results of the sums must yield a real function. Since a scalar (function) f is real if and only if $f^* = f$, the relation between the coefficients can be shown to be

$$f_l^m = \left(Y_l^m \,|\, f\right) = \left(Y_l^m \,|\, f^*\right) = (-1)^m f_l^{-m*}, \tag{5.52}$$

where we used that $Y_l^{m*} = (-1)^m Y_l^{-m}$. This means that the system of equations (5.41) contains too much information, i.e. too many equations, because

it determines the expansion coefficients f_l^{-m} as well. And these, as we have just proven, can be directly obtained from the ones with positive m. A way to remove the redundant equations is to use real spherical harmonics instead of complex spherical harmonics. We note that the above derivations are greatly simplified by working with the complex spherical harmonics, and this is why we choose to transform to the real spherical harmonics only after the matrix representations have been determined.

5.3.1 REAL SPHERICAL HARMONICS

We begin with a definition of the *real spherical harmonics*. One way to arrive at their definition is to rewrite the spherical harmonic expansion (4.2) as

$$\begin{aligned} f &= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_l^m Y_l^m = \sum_{l=0}^{\infty} f_l^0 Y_l^0 + \sum_{m=1}^{l} \left(f_l^m Y_l^m + f_l^{-m} Y_l^{-m} \right) \\ &= \sum_{l=0}^{\infty} f_l^0 Y_l^0 + \sum_{m=1}^{l} \left(f_l^m Y_l^m + f_l^{m*} Y_l^{m*} \right) \\ &= \sum_{l=0}^{\infty} f_l^0 Y_l^0 + \sum_{m=1}^{l} \left(\frac{f_l^m + f_l^{m*}}{\sqrt{2}} \frac{Y_l^m + Y_l^{m*}}{\sqrt{2}} - \frac{(f_l^m - f_l^{m*})}{\sqrt{2}\mathrm{i}} \frac{(Y_l^m - Y_l^{m*})}{\sqrt{2}\mathrm{i}} \right) \\ &=: \sum_{l=0}^{\infty} f_{l00} Y_{l00} + \sum_{m=1}^{l} \left(f_{lm0} Y_{lm0} + f_{lm1} Y_{lm1} \right) = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{s=0}^{1} f_{lms} Y_{lms} \,. \end{aligned} \tag{5.53}$$

In the last line we defined the real spherical harmonics Y_{lms}. We remark that this definition is equivalent to the one presented in eq. (4.44). The expansion coefficients f_{lms} are real as well. Furthermore, $f_{l01} = 0$ and $Y_{l01} = 0$.

We note that the real spherical harmonics are also orthonormal functions, namely $(Y_{l'm's'} \mid Y_{lms}) = \delta_{l'l}\delta_{m'm}\delta_{s's}$, which can be verified by a direct computation of the scalar product.

The context of the definition of the real spherical harmonics in eq. (5.53) allows us to interpret them as a real basis of the space of spherical harmonics $\mathcal{S}$. The basis transformation between the real and complex spherical harmonics can be computed with the help of

$$Y_{l'm's'} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m \left(Y_l^m \mid Y_{l'm's'} \right) =: \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m (\boldsymbol{S})_{i(l,m)j(l',m',s')} \,, \tag{5.54}$$

where we introduced a new index map $j(l', m', s')$ that again encodes the ordering of the real spherical harmonics Y_{lms} and their coefficients f_{lms}. For example, if the real spherical harmonics are ordered as

$$\mathcal{S} = \mathrm{span}\{Y_{000}, Y_{110}, Y_{100}, Y_{111}, Y_{220}, Y_{210}, Y_{200}, Y_{211}, Y_{221}, \dots\},$$

the index map is $j(l', m', s') = l'(l'+1) + (-1)^{s'+1} m'$[4]. Note the index j starts at zero. The matrix elements of the basis transformation matrix $\boldsymbol{S}$ can be computed to be

$$(\boldsymbol{S})_{i(l,m)j(l',m',s')} \tag{5.55}$$
$$= \delta_{l'l}\left(\frac{\left(\delta_{m'm} + (-1)^{m'}\delta_{-m'm}\right)}{\sqrt{2(1+\delta_{m'0})}}\delta_{s'0} - \mathrm{i}\frac{\left(\delta_{m'm} - (-1)^{m'}\delta_{-m'm}\right)}{\sqrt{2}}\delta_{s'1}\right)$$

where we used the definition of the real spherical harmonics given in eq. (5.53) and, once more, the relation $Y_l^{m*} = (-1)^m Y_l^{-m}$. An explicit expression of $\boldsymbol{S}$ is presented in eq. (4.44).

As noted before, the basis transformation matrix has the useful property that it is a unitary matrix. This can now be understood as a consequence of the fact that basis transformation matrices relating two orthogonal bases are unitary. We include a proof of this general statement for the case at hand, i.e.

$$\begin{aligned}(\mathbf{1})_{i(l',m',s')j(l,m,s)} &= \delta_{l'l}\delta_{m'm}\delta_{s's} \\ &= \left(Y_{l'm's'} \,|\, Y_{lms} \right) \\ &= \sum_{n,k} \left(Y_n^k \,|\, Y_{l'm's'} \right)^* \left(Y_n^k \,|\, Y_{lms} \right) \\ &= \sum_{n,k} (\boldsymbol{S}^\dagger)_{i(l',m',s')h(n,k)} (\boldsymbol{S})_{h(n,k)j(l,m,s)} \,.\end{aligned} \tag{5.56}$$

In matrix form this equation reads $\mathbf{1} = \boldsymbol{S}^\dagger \boldsymbol{S}$ and the inverse basis transformation matrix is $\boldsymbol{S}^{-1} = \boldsymbol{S}^\dagger$.

5.3.2 TURNING THE SYSTEM OF PDES INTO A REAL SYSTEM

With the basis transformation and its inverse at hand, we can turn the complex system of equations (5.41) into a real system thereby removing the superfluous equations. To this end, we compute how the matrix representation of an operator changes under the basis transformation. We find for an arbitrary operator, say $\hat{O}$, that

$$\begin{aligned}(\boldsymbol{O}_R)_{i(l',m',s')j(l,m,s)} &= \left(Y_{l'm's'} \,|\, \hat{O} Y_{lms} \right) \\ &= \sum_{n,k} \sum_{n',k'} \left(Y_n^k \,|\, Y_{l'm's'} \right)^* \left(Y_n^k \,|\, \hat{O} Y_{n'}^{k'} \right) \left(Y_{n'}^{k'} \,|\, Y_{lms} \right) \\ &= \sum_{n,k} \sum_{n',k'} (\boldsymbol{S}^\dagger)_{i(l',m',s')h(n,k)} (\boldsymbol{O})_{h(n,k)r(n',k')} (\boldsymbol{S})_{r(n',k')j(l,m,s)}\end{aligned} \tag{5.57}$$

The subscript R means that $\boldsymbol{O}_R$ is the matrix representation of $\hat{O}$ with respect to the real spherical harmonic basis. Eq. (5.57) in matrix form reads $\boldsymbol{O}_R = \boldsymbol{S}^\dagger \boldsymbol{O} \boldsymbol{S}$.

[4]Note since there is no real spherical harmonic Y_{l01}, we define that the index map $j(l', m', s')$ is valid for $m' = 0$ and $s' = 0$ and *exclude* $m' = 0$ and $s' = 1$.

We now transform the complex system of equations (5.41) by multiplying it with $\boldsymbol{S}^\dagger$ from the left and inserting the identity matrix $\mathbf{1} = \boldsymbol{S}^\dagger\boldsymbol{S}$ where necessary. This yields

$$\begin{aligned}
&\partial_t \boldsymbol{S}^\dagger \boldsymbol{f} + \left(U^a \mathbf{1} + \upsilon \boldsymbol{S}^\dagger \boldsymbol{A}^a \boldsymbol{S}\right) \partial_{x_a} \boldsymbol{S}^\dagger \boldsymbol{f} \\
&\quad - \left(\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t} \boldsymbol{S}^\dagger \boldsymbol{A}^a \boldsymbol{S} + p \frac{\partial U_b}{\partial x^a} \boldsymbol{S}^\dagger \boldsymbol{A}^a \boldsymbol{S}\boldsymbol{S}^\dagger \boldsymbol{A}^b \boldsymbol{S}\right) \partial_p \boldsymbol{S}^\dagger \boldsymbol{f} \\
&\quad + \left(\frac{\mathrm{i}}{\upsilon} \epsilon_{abc} \frac{\mathrm{D}U^a}{\mathrm{D}t} \boldsymbol{S}^\dagger \boldsymbol{A}^b \boldsymbol{S}\boldsymbol{S}^\dagger \boldsymbol{\Omega}^c \boldsymbol{S} + \mathrm{i}\epsilon_{bcd} \frac{\partial U^b}{\partial x^a} \boldsymbol{S}^\dagger \boldsymbol{A}^a \boldsymbol{S}\boldsymbol{S}^\dagger \boldsymbol{A}^c \boldsymbol{S}\boldsymbol{S}^\dagger \boldsymbol{\Omega}^d \boldsymbol{S}\right) \boldsymbol{S}^\dagger \boldsymbol{f} \\
&\quad - \mathrm{i}\omega_a \boldsymbol{S}^\dagger \boldsymbol{\Omega}^a \boldsymbol{S}\boldsymbol{S}^\dagger \boldsymbol{f} + \nu \boldsymbol{S}^\dagger \boldsymbol{C}\boldsymbol{S}\boldsymbol{S}^\dagger \boldsymbol{f} \\
&= \partial_t \boldsymbol{f}_R + \left(U^a \mathbf{1} + \upsilon \boldsymbol{A}^a_R\right) \partial_{x_a} \boldsymbol{f}_R - \left(\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t} \boldsymbol{A}^a_R + p \frac{\partial U_b}{\partial x^a} \boldsymbol{A}^a_R \boldsymbol{A}^b_R\right) \partial_p \boldsymbol{f}_R \\
&\quad + \left(\frac{1}{\upsilon} \epsilon_{abc} \frac{\mathrm{D}U^a}{\mathrm{D}t} \boldsymbol{A}^b_R \boldsymbol{\Omega}^c_R + \epsilon_{bcd} \frac{\partial U^b}{\partial x^a} \boldsymbol{A}^a_R \boldsymbol{A}^c_R \boldsymbol{\Omega}^d\right) \boldsymbol{f}_R - \omega_a \boldsymbol{\Omega}^a_R \boldsymbol{f}_R + \nu \boldsymbol{C}_R \boldsymbol{f}_R = 0\,.
\end{aligned} \tag{5.58}$$

Notice in particular that $\boldsymbol{\Omega}^a_R \coloneqq \mathrm{i}\boldsymbol{S}^\dagger \boldsymbol{\Omega}^a \boldsymbol{S}$ includes the factor i.

It is left to show that eq. (5.58) is actually a real system of equations, i.e. that all the appearing matrices and $\boldsymbol{f}_R$ are real. We begin with the latter and show that the components of $\boldsymbol{f}_R$ are

$$\begin{aligned}
(\boldsymbol{f}_R)_{i(l,m,s)} &= \sum_{l',m'} (\boldsymbol{S}^\dagger)_{i(l,m,s)j(l',m')} (\boldsymbol{f})_{j(l',m')} \\
&= \frac{f_l^m + (-1)^m f_l^{-m}}{\sqrt{2(1+\delta_{m0})}} \delta_{s0} + \frac{\mathrm{i}(f_l^m - (-1)^m f_l^{-m})}{\sqrt{2}} \delta_{s1} = f_{lms}\,.
\end{aligned} \tag{5.59}$$

These are the expansion coefficients of the real spherical harmonic expansion, which we derived in eq. (5.53) and which, as we pointed out there, are real. To prove that the matrices are real, we had to compute them explicitly. The explicit expressions for all real representation matrices can be found in Appendix B.1.

At the end of this section, we stress that the real representation matrices can as well be computed with the help of rotation matrices, i.e.

$$\boldsymbol{A}^y_R = e^{-\frac{\pi}{2}\boldsymbol{\Omega}^z_R} \boldsymbol{A}^x_R e^{\frac{\pi}{2}\boldsymbol{\Omega}^z_R} \qquad \boldsymbol{A}^z_R = e^{-\frac{\pi}{2}\boldsymbol{\Omega}^x_R} \boldsymbol{A}^y_R e^{\frac{\pi}{2}\boldsymbol{\Omega}^x_R} \quad \text{and} \tag{5.60}$$

$$\boldsymbol{\Omega}^y_R = e^{-\frac{\pi}{2}\boldsymbol{\Omega}^z_R} \boldsymbol{\Omega}^x_R e^{\frac{\pi}{2}\boldsymbol{\Omega}^z_R} \qquad \boldsymbol{\Omega}^z_R = e^{-\frac{\pi}{2}\boldsymbol{\Omega}^x_R} \boldsymbol{\Omega}^y_R e^{\frac{\pi}{2}\boldsymbol{\Omega}^x_R}\,. \tag{5.61}$$

5.4 EIGENVECTORS AND EIGENVALUES OF THE REPRESENTATION MATRICES

Numerically solving the system of PDEs for the expansion coefficients, as derived in eq. (5.41) or (5.58), using standard techniques, e.g. the finite volume method or the discontinuous Galerkin method, requires knowledge about the eigenvalues and eigenvectors of the matrix representations. We prove that the

representation matrices of the direction operators have the same eigenvalues and that their eigenvectors are rotated versions of each other. Moreover, we show that the eigenvalues are the roots of the associated Legendre polynomials. The implementation of the discontinuous Galerkin method exploiting this knowledge is the subject of the next chapter. Related statements about the eigenvalues of the representation matrices of the angular momentum operators are included for the sake of completeness.

5.4.1 GENERAL STATEMENTS

A first important observation is that the matrix representations of the $\hat{A}^a$ and $\hat{L}^a$ are Hermitian matrices, because the eigenvalues of Hermitian matrices must be real. Assume $\boldsymbol{\xi}$ is an eigenvector of an arbitrary Hermitian matrix $\boldsymbol{A}$ with eigenvalue λ and with norm $\xi \coloneqq \sqrt{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi}} = 1$, then

$$\begin{aligned} \lambda^* &= (\boldsymbol{\xi}^* \cdot \boldsymbol{A}\boldsymbol{\xi})^* \\ &= (\boldsymbol{\xi})_i (\boldsymbol{A})^*_{ij} (\boldsymbol{\xi})^*_j = (\boldsymbol{\xi})^*_j (\boldsymbol{A}^\mathsf{T})^*_{ji} (\boldsymbol{\xi})_i = \boldsymbol{\xi}^* \cdot \boldsymbol{A}^\dagger \boldsymbol{\xi} = \boldsymbol{\xi}^* \cdot \boldsymbol{A}\boldsymbol{\xi} = \lambda \,. \end{aligned} \tag{5.62}$$

We now prove the above observation for the direction operators $\hat{A}^a$. We prove it for $\hat{A}^x$ and only note that the proof for the other two operators works analogously. We use the definition of the representation matrices given in eq. (5.16) to get the transpose conjugate representation matrix of $\hat{A}^x$, namely[5]

$$\begin{aligned} (\boldsymbol{A}^{x\dagger})_{j(l,m)i(l',m')} &= \left(Y_l^m \,\middle|\, \hat{A}^x Y_{l'}^{m'} \right)^* \\ &= \left(\int_{S^2} Y_l^{m*} \cos\theta Y_{l'}^{m'} \,\mathrm{d}\Omega \right)^* \\ &= \int_{S^2} Y_{l'}^{m'*} \cos\theta Y_l^m \,\mathrm{d}\Omega = \left(Y_{l'}^{m'} \,\middle|\, \hat{A}^x Y_l^m \right) = (\boldsymbol{A}^x)_{i(l',m')j(l,m)} \,. \end{aligned} \tag{5.63}$$

Hence $\boldsymbol{A}^x$ is a Hermitian matrix. The same is true for the representation matrices of $\hat{A}^y$ and $\hat{A}^z$. Hence, $\boldsymbol{A}^{a\,\dagger} = \boldsymbol{A}^a$ for $a \in \{x, y, z\}$.

We proceed by proving that the representation matrices of the angular momentum operators are Hermitian. Knowing their action on the spherical harmonics, see eq. (5.11), we conclude that

$$\begin{aligned} (\boldsymbol{\Omega}^{x\dagger})_{j(l,m)i(l',m')} &= \left(Y_l^m \,\middle|\, \hat{L}^x Y_{l'}^{m'} \right)^* \\ &= m'\delta_{mm'} = m\delta_{m'm} = \left(Y_{l'}^{m'} \,\middle|\, \hat{L}^x Y_l^m \right) = (\boldsymbol{\Omega}^x)_{i(l',m')j(l,m)} \,, \end{aligned} \tag{5.64}$$

where we used that we get the same result when the roles of m' and m are

[5] There is an error in the original publication Schween and Reville (2024): The index maps $i(l', m')$ and $j(l, m)$ are swapped.

interchanged. For $\hat{L}^y$, we find

$$\begin{aligned}
&(\boldsymbol{\Omega}^{y\dagger})_{j(l,m)i(l',m')} \\
&\quad = \left(Y_l^m \,\middle|\, \hat{L}^y Y_{l'}^{m'} \right)^* \\
&\quad = \frac{1}{2}\sqrt{(l'+m'+1)(l'-m')}\delta_{ll'}\delta_{m(m'+1)} \\
&\qquad + \frac{1}{2}\sqrt{(l'+m')(l'-m'+1)}\delta_{ll'}\delta_{m(m'-1)} \\
&\quad = \frac{1}{2}\sqrt{(l+m)(l-m+1)}\delta_{l'l}\delta_{m'(m-1)} + \frac{1}{2}\sqrt{(l+m+1)(l-m)}\delta_{l'l}\delta_{m'(m+1)} \\
&\quad = \left(Y_{l'}^{m'} \,\middle|\, \hat{L}^y Y_l^m \right) = (\boldsymbol{\Omega}^y)_{i(l',m')j(l,m)} \,.
\end{aligned} \tag{5.65}$$

We used that $\delta_{m(m'+1)} = \delta_{m'(m-1)}$ and $\delta_{m(m'-1)} = \delta_{m'(m+1)}$ and, in agreement with it, we replaced m' with $m-1$ and $m+1$ respectively. The proof for $\hat{L}^z$ is analogous.

A second statement about the eigenvalues of the representation matrices follows from the fact that the representation matrices are 'rotated' versions of each other. This implies that their eigenvalues are the same. The proof requires the concept of similar matrices: We say the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are *similar*, if they are related by $\boldsymbol{Y} = \boldsymbol{T}^{-1}\boldsymbol{X}\boldsymbol{T}$. Similar matrices have the same eigenvalues, because their *characteristic polynomial* is the same, i.e.

$$\begin{aligned}
p(\lambda) \coloneqq \det(\boldsymbol{X} - \lambda\boldsymbol{I}) &= \det\left(\boldsymbol{T}(\boldsymbol{T}^{-1}\boldsymbol{X}\boldsymbol{T} - \lambda\boldsymbol{I})\boldsymbol{T}^{-1}\right) \\
&= \det(\boldsymbol{T})\det(\boldsymbol{Y} - \lambda\boldsymbol{I})\det(\boldsymbol{T}^{-1}) = \det(\boldsymbol{Y} - \lambda\boldsymbol{I}) \,.
\end{aligned} \tag{5.66}$$

Equation (5.48) states that $\boldsymbol{A}^y$ is similar to $\boldsymbol{A}^z$ and eq. (5.49) says that $\boldsymbol{A}^x$ is similar to $\boldsymbol{A}^y$. The same is true for the representation matrices of the angular momentum operators, see eq. (5.51). Thus, all three matrices share the same eigenvalues. We define the *spectrum* to be the set of all eigenvalues of a matrix, namely $\sigma(\boldsymbol{X}) \coloneqq \{\lambda \in \mathbb{C} \mid \exists \boldsymbol{w} \neq 0 : \boldsymbol{X}\boldsymbol{w} = \lambda\boldsymbol{w}\}$. The fact that all three matrices have the same eigenvalues can now be expressed as

$$\sigma(\boldsymbol{A}^x) = \sigma(\boldsymbol{A}^y) = \sigma(\boldsymbol{A}^z) \,. \tag{5.67}$$

A third statement concerns the eigenvectors of the representation matrices: The eigenvectors are rotated versions of each other, i.e.

$$\begin{aligned}
\boldsymbol{A}^x\boldsymbol{W}^x = \boldsymbol{W}^x\boldsymbol{\Lambda} &\iff e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\boldsymbol{A}^x e^{\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z} e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\boldsymbol{W}^x = e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\boldsymbol{W}^x\boldsymbol{\Lambda} \\
&\iff \boldsymbol{A}^y e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\boldsymbol{W}^x = e^{-\mathrm{i}\frac{\pi}{2}\boldsymbol{\Omega}^z}\boldsymbol{W}^x\boldsymbol{\Lambda} \\
&\iff \boldsymbol{A}^y\boldsymbol{W}^y = \boldsymbol{W}^y\boldsymbol{\Lambda} \,,
\end{aligned} \tag{5.68}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{A}^x$ on its diagonal and $\boldsymbol{W}$ are matrices whose columns are the corresponding eigenvectors. An analogue computation can be performed for $\boldsymbol{A}^z$.

5.4.2 EIGENVALUES OF THE DIRECTION OPERATORS' MATRICES

In the previous section we showed that if we know the eigenvalues of one of the representation matrices of the direction operators, we know them for all. Since $\boldsymbol{A}^x$ is simpler, i.e. it has less non-zero elements, we focus on determining its eigenvalues and, as we will see, its eigenvalues are the roots of the associated Legendre polynomials.

To see a general pattern emerge, we begin with computing the eigenvalues of $\boldsymbol{A}^x$ for $l_{\max} = 1$ and $l_{\max} = 2$. Notice that for the following ordering of the spherical harmonics

$$\begin{aligned} \mathcal{S}^1 &= \operatorname{span}\{Y_0^0, Y_1^0, Y_1^1, Y_1^{-1}\} && \text{for } l_{\max} = 1 \text{ and} \\ \mathcal{S}^2 &= \operatorname{span}\{Y_0^0, Y_1^0, Y_2^0, Y_1^1, Y_2^1, Y_1^{-1}, Y_2^{-1}, Y_2^2, Y_2^{-2}\} && \text{for } l_{\max} = 2, \end{aligned}$$

the matrix $\boldsymbol{A}^x$ is, as we will see now, tridiagonal. We use eq. (5.32) to compute its matrix elements and introduce the shorthand

$$c_l^m \coloneqq [(l+m)(l-m)/(2l+1)(2l-1)]^{1/2}$$

for the coefficients in the referred equation. This yields for $l_{\max} = 1$

$$\boldsymbol{A}^x = \begin{pmatrix} 0 & c_1^0 & 0 & 0 \\ c_1^0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^x_{1,0} & & & \\ & \boldsymbol{A}^x_{1,1} & & \\ & & \boldsymbol{A}^x_{1,-1} & \end{pmatrix}$$

and for $l_{\max} = 2$ we find

$$\boldsymbol{A}^x = \begin{pmatrix} 0 & c_1^0 & 0 & & & & & & \\ c_1^0 & 0 & c_2^0 & & & & & & \\ 0 & c_2^0 & 0 & & & & & & \\ & & & 0 & c_2^1 & & & & \\ & & & c_2^1 & 0 & & & & \\ & & & & & 0 & c_2^{-1} & & \\ & & & & & c_2^{-1} & 0 & & \\ & & & & & & & 0 & \\ & & & & & & & & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^x_{2,0} & & & & \\ & \boldsymbol{A}^x_{2,1} & & & \\ & & \boldsymbol{A}^x_{2,-1} & & \\ & & & \boldsymbol{A}^x_{2,2} & \\ & & & & \boldsymbol{A}^x_{2,-2} \end{pmatrix}$$

where we introduced the tridiagonal block matrices

$$\boldsymbol{A}^x_{l,m} \coloneqq \begin{pmatrix} 0 & c^m_{|m|+1} & & & \\ c^m_{|m|+1} & 0 & c^m_{|m|+2} & & \\ & c^m_{|m|+2} & \ddots & \ddots & \\ & & \ddots & \ddots & c^m_l \\ & & & c^m_l & 0 \end{pmatrix} \in \mathbb{R}^{(l-m+1)\times(l-m+1)} \tag{5.69}$$

with $|m| \leq l$ and the convention that $\boldsymbol{A}^x_{l,\pm l} \coloneqq 0$. We emphasise that the size of the block matrices $\boldsymbol{A}^x_{l,m}$ varies with l and that the pattern shown in the two

examples above is the same for an arbitrary $l_{\max}$. The block diagonal structure of $\boldsymbol{A}^x$ implies that its characteristic polynomial factors into the characteristic polynomials of its blocks, i.e.

$$\begin{aligned} p(\lambda) = \det(\boldsymbol{A}^x - \lambda\mathbf{1}) &= \prod_{m=-l_{\max}}^{l_{\max}} \det(\boldsymbol{A}^x_{l_{\max},m} - \lambda\mathbf{1}) \\ &= \det(\boldsymbol{A}^x_{l_{\max},0} - \lambda\mathbf{1}) \prod_{m=1}^{l_{\max}} \left[\det(\boldsymbol{A}^x_{l_{\max},m} - \lambda\mathbf{1})\right]^2 , \end{aligned} \tag{5.70}$$

where we used that $\boldsymbol{A}_{l_{\max},m} = \boldsymbol{A}_{l_{\max},-m}$, because $c_l^m = c_l^{-m}$. A direct consequence of eq. (5.70) is that the eigenvalues of $\boldsymbol{A}^x$ (i.e. the roots of its characteristic polynomial) are the eigenvalues of the block matrices $\boldsymbol{A}^x_{l_{\max},m}$.

We now prove that their eigenvalues are the roots of the mth derivative of the Legendre polynomial P_{l+1}. Before we begin the proof we define

$$p_l^m(x) \coloneqq \frac{\mathrm{d}^m}{\mathrm{d}x^m} P_l(x) , \tag{5.71}$$

to denote the mth derivative of the Legendre polynomial. Furthermore, we point out that p_l^m is part of the definition of the associated Legendre polynomials, i.e. $P_l^m(x) = (-1)^m(1-x^2)^{m/2} p_l^m(x)$. Thus, if the eigenvalues of $\boldsymbol{A}^x_{l,m}$ are the roots of the mth derivative of the Legendre polynomial P_{l+1}, they are as well the roots of the associated Legendre polynomial P_{l+1}^m modulo ± 1.

In a first step, we compute $\det(\boldsymbol{A}^x_{l,m} - \lambda\mathbf{1})$ developing it after the last row, i.e.

$$\begin{aligned} \pi_{l+1}^m(\lambda) &\coloneqq \det(\boldsymbol{A}^x_{l,m} - \lambda\mathbf{1}) \\ &= -\lambda \det(\boldsymbol{A}^x_{l-1,m} - \lambda\mathbf{1}) - (c_l^m)^2 \det(\boldsymbol{A}^x_{l-2,m} - \lambda\boldsymbol{I}) \\ &= -\lambda\pi_l^m(\lambda) - (c_l^m)^2 \pi_{l-1}^m(\lambda) . \end{aligned} \tag{5.72}$$

This equation can be read as a recurrence relation for the newly introduced polynomials π_l^m. Fixing m to some integer value and setting $\pi_{|m|}^m(\lambda) \coloneqq 1$ the characteristic polynomials for arbitrary l, m can be computed recursively.

In a second step, we compare the recurrence relation (5.72) with the recurrence relation for the associated Legendre polynomials that we presented in eq. (5.17), which implies a recurrence relation for the mth derivative of the Legendre polynomials, i.e.

$$\begin{aligned} P_{l+1}^m(x) &= x\frac{2l+1}{l-m+1} P_l^m(x) - \frac{(l+m)}{(l-m+1)} P_{l-1}^m(x) \\ \implies p_{l+1}^m(x) &= x\frac{2l+1}{l-m+1} p_l^m(x) - \frac{(l+m)}{(l-m+1)} p_{l-1}^m(x) . \end{aligned} \tag{5.73}$$

We see that the recurrence relations in eq. (5.72) and eq. (5.73) are similar. They differ in the factors in front of the polynomials on the right-hand side. This suggests that $p_l^m(x) \propto \pi_l^m(x)$. If the polynomials are proportional to each other,

they have the same roots; multiplying a function with a constant does not change its roots. The factor of proportionality can be derived by noting that the leading order term of $\pi_l^m(x)$ is $(-1)^{l-m}x^{l-m}$, using an explicit formula for the associated Legendre polynomials and dividing its leading order term by the factor which accompanies it. This results in

$$p_l^m(x) = (-1)^l \frac{(2l-1)!!}{(l-m)!}\pi_l^m(x)\,. \tag{5.74}$$

We summarise and conclude that the eigenvalues of the representation matrices of the operators $\hat{A}^a$ for a fixed $l_{\max}$ are the roots of the polynomials $\pi^m_{l_{\max}+1}$ with $0 \leq m \leq l_{\max}$, which are, modulo a constant of proportionality, the mth derivatives of the Legendre Polynomial P_{l+1}. Hence, the eigenvalues of $\boldsymbol{A}^a$ are the roots of the associated Legendre Polynomials $P^0_{l_{\max}+1}, \ldots, P^{l_{\max}}_{l_{\max}+1}$ modulo ± 1. We note that the roots of the polynomials with $m \neq 0$ have algebraic multiplicity two, see eq. (5.70).

The above conclusions have implications for the possible set of eigenvalues: First, the Legendre polynomials are orthogonal polynomials and the roots of orthogonal polynomials are real, simple and located in the interior of the interval of orthogonality, which is $[-1, 1]$ in the case of the Legendre polynomials, see Milton and Stegun (1964, Section 22.16). Secondly, the roots of P_l^m with $m \neq 0$ are the roots of the derivatives of the Legendre polynomials, i.e. the roots of p_l^m. Rolle's theorem states that the roots of the derivative of a continuous and differentiable function, must lie between the roots of this function. This implies that the roots of the polynomials p_l^m are 'moving' towards zero with increasing m, hence that all eigenvalues of $\boldsymbol{A}^x$ are contained in the interval $[-1, 1]$ and that the largest eigenvalue of $\boldsymbol{A}^x$ is the largest root of the Legendre polynomial $P^0_{l_{\max}+1} = P_{l_{\max}+1}$.

We can get a very good estimate for the largest eigenvalue using a formula to compute estimates of the roots of the Legendre polynomial of degree l, which can be found in Tricomi (1950, eq.13). The formula is

$$x_r = \left(1 - \frac{1}{8l^2} + \frac{1}{8l^3}\right)\cos\left(\frac{4r-1}{4l+2}\pi\right) + \mathcal{O}(l^{-4}) \tag{5.75}$$

where $r \in \{1, \ldots, l/2\}$ for even l and $r \in \{1, \ldots, (l-1)/2\}$ for odd l. The estimate for the largest eigenvalue of $\boldsymbol{A}^x$ is $\lambda_{\max} \approx x_1$.

5.4.3 SUMS OF REPRESENTATION MATRICES

In the last part of this section we show how to compute the eigenvalues and eigenvectors of

$$\eta_x \boldsymbol{A}^x + \eta_y \boldsymbol{A}^y + \eta_z \boldsymbol{A}^z \quad \text{where } \boldsymbol{\eta} = (\eta_x, \eta_y, \eta_z)^T \in \mathbb{R}^3 \tag{5.76}$$

using the eigenvalues and eigenvectors of $\boldsymbol{A}^x$.

In a first step we note that the sum of the representation matrices of the direction operators can be traced back to the scalar product $\boldsymbol{\eta} \cdot \boldsymbol{e}_p = \eta_x \hat{A}^x + \eta_y \hat{A}^y + \eta_z \hat{A}^z$, see eq. (5.37). Secondly, we define the unit vector $\boldsymbol{n} \coloneqq \boldsymbol{\eta}/\eta$. Thirdly, we rotate the coordinate system such that the x-axis of the rotated coordinate system is parallel to $\boldsymbol{n}$. We denote the corresponding rotation matrix with $\boldsymbol{R}_{\boldsymbol{n}}$. This changes the scalar product $\boldsymbol{\eta} \cdot \boldsymbol{e}_p$ to

$$\eta \boldsymbol{n} \cdot \left(\boldsymbol{R}_{\boldsymbol{n}}^{\mathsf{T}} \boldsymbol{R}_{\boldsymbol{n}} \boldsymbol{e}_p\right) = \eta \left(\boldsymbol{R}_{\boldsymbol{n}} \boldsymbol{n}\right) \cdot \left(\boldsymbol{R}_{\boldsymbol{n}} \boldsymbol{e}_p\right) = \eta \tilde{\boldsymbol{n}} \cdot \tilde{\boldsymbol{e}}_p = \eta \boldsymbol{e}_x \cdot \tilde{\boldsymbol{e}}_p = \eta \cos \tilde{\theta} = \eta \tilde{A}_x \,, \tag{5.77}$$

where we used that the coordinates $\tilde{\boldsymbol{n}}$ of $\boldsymbol{n}$ in the rotated coordinate system are $\boldsymbol{e}_x$. Eq. (5.77) shows that the sum of the direction operators that is given by the scalar product $\boldsymbol{\eta} \cdot \boldsymbol{e}_p$ reduces to $\eta \tilde{A}_x$ in the rotated coordinate system and so does its matrix representation (5.76) with respect to the spherical harmonics defined in the rotated coordinate system.

As explained in Section 5.2 transforming representation matrices into a rotated coordinate system requires to know how to rotate spherical harmonics. In eq. (5.44) we presented the necessary rotation matrix $\boldsymbol{U}(\alpha \boldsymbol{r}) = e^{-\mathrm{i}\alpha r_a \boldsymbol{\Omega}^a}$, where the unit vector $\boldsymbol{r}$ is the axis of rotation and α is the angle by which the spherical harmonics are rotated.

Because of eq. (5.47) and since we know that the sum of matrices (5.76) must equal $\eta \boldsymbol{A}^x$ in the rotated frame,

$$\eta \boldsymbol{A}^x = \boldsymbol{U}(\alpha \boldsymbol{r})^\dagger (\eta_x \boldsymbol{A}^x + \eta_y \boldsymbol{A}^y + \eta_z \boldsymbol{A}^z) \boldsymbol{U}(\alpha \boldsymbol{r}) \,. \tag{5.78}$$

The axis of rotation can be computed with $\boldsymbol{r} = \boldsymbol{e}_x \times \boldsymbol{n} / \|\boldsymbol{e}_x \times \boldsymbol{n}\|$ and the angle $\alpha = \boldsymbol{e}_x \cdot \boldsymbol{n}$.

We can now investigate the eigenvalues of the sum of the representation matrices (5.76). They are determined by

$$\begin{aligned} &\det\left(\eta_x \boldsymbol{A}^x + \eta_y \boldsymbol{A}^y + \eta_z \boldsymbol{A}^z - \lambda^s \mathbf{1}\right) \\ &\quad = \det\left(\boldsymbol{U}(\alpha \boldsymbol{r}) \left(\boldsymbol{U}^\dagger(\alpha \boldsymbol{r}) \left(\eta_x \boldsymbol{A}^x + \eta_y \boldsymbol{A}^y + \eta_z \boldsymbol{A}^z\right) \boldsymbol{U}(\alpha \boldsymbol{r}) - \lambda^s \mathbf{1}\right) \boldsymbol{U}^\dagger(\alpha \boldsymbol{r})\right) \\ &\quad = \det\left(\eta \boldsymbol{A}^x - \lambda^s \mathbf{1}\right) = 0 \,, \end{aligned} \tag{5.79}$$

whence we conclude that the eigenvalues of the sum of the representation matrices (5.76) are the eigenvalues of $\boldsymbol{A}^x$ times η, i.e. its ith eigenvalue is

$$\lambda_i^s = \lambda_i (\eta_x^2 + \eta_y^2 + \eta_z^2)^{1/2} \,, \tag{5.80}$$

where λ_i is the ith eigenvalue of $\boldsymbol{A}^x$. We highlight that this relation between the eigenvalues of $\boldsymbol{A}_x$ and the eigenvalues of the sum of the representation matrices is independent of $l_{\max}$ and holds for each eigenvalue λ_i^s.

The geometrical interpretation of eq. (5.80) is that for a given value η_z the eigenvalues λ_i^s lie on one of the sheets of a circular and two-sheeted hyperboloid that is parameterised by η_x and η_y. To illustrate this we plotted one eigenvalue of the sum of the representation matrices (5.76) for two values of η_z, see Fig. 5.1.

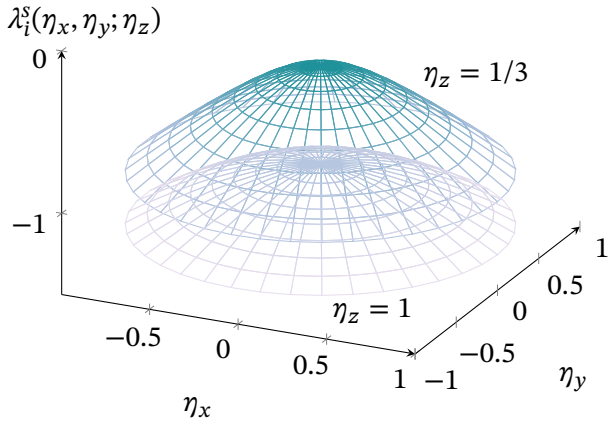

Figure 5.1: Plot of the eigenvalue λ_i^s of $\eta_x \boldsymbol{A}^x + \eta_y \boldsymbol{A}^y + \eta_z \boldsymbol{A}^z$ for varying η_x and η_y and fixed η_z. The plotted λ_i^s corresponds to the eigenvalue $\lambda_i = -0.9062$ of $\boldsymbol{A}^x$ with $l_{\max} = 4$. The depicted circular hyperboloid is described by $(\eta_x^2 + \eta_y^2)/\eta_z^2 - (\lambda_i^s/\lambda_i \eta_z)^2 = -1$.

The eigenvectors of the sum of the representation matrices (5.76) are the rotated eigenvectors of $\boldsymbol{A}^x$, because

$$\begin{aligned} \eta \boldsymbol{A}^x \boldsymbol{W}^x = \eta \boldsymbol{W}^x \boldsymbol{\Lambda} &\iff \eta \boldsymbol{U}(\alpha \boldsymbol{r}) \boldsymbol{A}^x \boldsymbol{U}^\dagger(\alpha \boldsymbol{r}) \boldsymbol{U}(\alpha \boldsymbol{r}) \boldsymbol{W}^x = \eta \boldsymbol{U}(\alpha \boldsymbol{r}) \boldsymbol{W}^x \boldsymbol{\Lambda} \\ &\iff \left(\eta_x \boldsymbol{A}^x + \eta_y \boldsymbol{A}^y + \eta_z \boldsymbol{A}^z\right) \boldsymbol{U}(\alpha \boldsymbol{r}) \boldsymbol{W}^x = \boldsymbol{U}(\alpha \boldsymbol{r}) \boldsymbol{W}^x \boldsymbol{\Lambda}^s \,. \end{aligned}$$

We conclude this section with referring the reader to Garrett and Hauck (2016), who also investigated the eigenstructure of the $\boldsymbol{A}^a$ matrices in the context of radiation transport and found similar results. Our work differs from theirs mainly in identifying the unitary matrices $\boldsymbol{U}$ as rotation matrices. Furthermore we would like to remark, that the eigenvalues of the real matrices $\boldsymbol{A}_R^a$ are the same as the eigenvalues of the $\boldsymbol{A}^a$ matrices, because $\boldsymbol{S}^\dagger \boldsymbol{A}^a \boldsymbol{S} = \boldsymbol{A}_R^a$ is a similarity transformation.

CHAPTER 6

NUMERICAL SOLUTION OF THE SYSTEM OF PARTIAL DIFFERENTIAL EQUATIONS

In the last chapter, we turn our attention to solving the system of partial differential equations for the expansion coefficients of the spherical harmonic expansion. This means that we simulate the transport of charged and energetic particles in tenuous astrophysical plasmas. Notice that as at the time of writing this chapter, the content is submitted to a journal and a preprint is available (Schween et al., 2024).

We remind ourselves that our aim is to solve the semi-relativistic VFP equation (3.60). In order to reduce the dimensionality of the problem and to deal with physical situations in which the diffusion approximation is not applicable, we treat the momentum dependence of the distribution function with a spherical harmonic expansion that we truncate at arbitrary $l_{\max}$. Computing the single particle distribution function f then amounts to solving the system of PDEs for the expansion coefficients. We choose to solve the real system of PDEs that we derived in eq. (5.58), because it does not contain any linear dependent equations as its complex counterpart. For a better overview, we re-print it here, i.e.

$$\begin{aligned}\frac{\partial \boldsymbol{f}}{\partial t} + \left(U^a \mathbf{1} + v\boldsymbol{A}^a\right)\frac{\partial \boldsymbol{f}}{\partial x^a} - \left(\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t}\boldsymbol{A}^a + p\frac{\partial U_b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^b\right)\frac{\partial \boldsymbol{f}}{\partial p} \\ + \left(\frac{1}{v}\epsilon_{abc}\frac{\mathrm{D}U^a}{\mathrm{D}t}\boldsymbol{A}^b\boldsymbol{\Omega}^c + \epsilon_{bcd}\frac{\partial U^b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^c\boldsymbol{\Omega}^d\right)\boldsymbol{f} - \omega_a\boldsymbol{\Omega}^a\boldsymbol{f} + \nu\boldsymbol{C}\boldsymbol{f} = 0\,,\end{aligned} \tag{6.1}$$

where we dropped the subscript R to simplify the notation. Notice that the components of $\boldsymbol{f}$ are the expansion coefficients f_{lms}, see eq. (5.59). We stress that any time-dependent solutions are accurate only to zeroth order in U/c, because when we derived the semi-relativistic VFP equations we dropped corrections of order U/c, cf. the end of Sec. 3.2.4. We discuss approaches to recover higher order accuracy in C.2.

We solve the system of PDEs with the discontinuous Galerkin (dG) method in conjunction with implicit and explicit time stepping methods, because this provides significant flexibility in our choice of spatial and temporal accuracy. The dG method is a finite element (FE) method, ideally suited for advection-reaction equations and, as we will argue in the next section, the system we solve

is exactly this, an advection-reaction equation. We note that typically codes apply either finite difference or finite volume approaches to solve the VFP equation numerically (cf. Bell et al., 2006; Zhang et al., 2024). However, the dG method is also used in the `Gkeyll` code to solve the Vlasov equation in Cartesian spatial and momentum coordinates using (up to) 6D discontinuous basis functions, i.e. without an expansion of f.

In this chapter we detail the dG method for the system of PDEs and we explore the capabilities of of our implementation by means of three physically motivated examples. In particular, we simulate the acceleration of test particles at a parallel shock and compare the results to the analytical predictions derived in Sec. (3.3.1).

Our implementation is available as a free and open-source code called `Sapphire++` (‘**S**imulating **a**strophysical **p**lasmas and **p**articles with **hi**ghly **r**elativistic **e**nergies in `C++`’)[1]. As the acronym `Sapphire++` implies, the code is written in the `C++` programming language. The aim of `Sapphire++` is to combine the advantages of both, i.e. the expansion of the distribution function and the dG method. We note that the implementation is based on the FE library `deal.ii`[2] (Arndt et al., 2021, 2023).

6.1 THE DISCONTINUOUS GALERKIN METHOD

The advection-reaction equation is known to be well suited for an application of the dG method, see Di Pietro and Ern (2012) for a textbook example. In the first part of this section we introduce a set of symbols that allows us to identify the system of PDEs as such an advection-reaction equation. Subsequently, we explain what the dG method is and how to apply it to eq. (6.1) and eventually, we briefly introduce the Θ-method, i.e. one of the time-stepping methods that is implemented in `Sapphire++`

6.1.1 ADVECTION-REACTION EQUATION

The system of equations (6.1) can be brought into the following form

$$\frac{\partial \boldsymbol{f}}{\partial t} + (\boldsymbol{\beta} \cdot \tilde{\nabla})\boldsymbol{f} + \boldsymbol{R}\boldsymbol{f} = 0\,, \tag{6.2}$$

[1] https://sapphirepp.org
[2] https://dealii.org

where we introduce the symbols

$$
\begin{aligned}
(\boldsymbol{\beta})^{\alpha} &\coloneqq \begin{cases} U^{\alpha}\mathbf{1} + \upsilon \boldsymbol{A}^{\alpha} & \text{for } \alpha \in \{1,2,3\} \\ -\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t}\boldsymbol{A}^{a} - p\frac{\partial U_b}{\partial x^a}\boldsymbol{A}^{a}\boldsymbol{A}^{b} & \text{for } \alpha = 4 \end{cases}, \\
(\tilde{\nabla})_{\alpha} &\coloneqq \begin{cases} \partial/\partial x^{\alpha} & \text{for } \alpha \in \{1,2,3\} \\ \partial/\partial p & \text{for } \alpha = 4 \end{cases} \text{ and} \\
\boldsymbol{R} &\coloneqq \frac{1}{\upsilon}\epsilon_{abc}\frac{\mathrm{D}U^{a}}{\mathrm{D}t}\boldsymbol{A}^{b}\boldsymbol{\Omega}^{c} + \epsilon_{bcd}\frac{\partial U^{b}}{\partial x^{a}}\boldsymbol{A}^{a}\boldsymbol{A}^{c}\boldsymbol{\Omega}^{d} - \omega_{a}\boldsymbol{\Omega}^{a} + \nu\boldsymbol{C}\,.
\end{aligned}
\tag{6.3}
$$

Equation (6.2) is an *advection-reaction equation*, $\boldsymbol{\beta}^{\alpha}$ are the *advection matrices* and $\boldsymbol{R}$ is the *reaction matrix*. Moreover, we refer to the vector space $\boldsymbol{\xi} = (\boldsymbol{x}, p)^{\mathsf{T}}$ as *reduced phase space*, and use the index α to refer to the components of vectors in the reduced phase space. The above system of partial differential equations is a *linear hyperbolic system*, because the advection matrices are symmetric, which implies that they are diagonalisable and that their eigenvalues are real, see LeVeque (2002, Sec. 2.9). Note that due to the symmetry of $\boldsymbol{A}^{a}$ and of the products $\boldsymbol{A}^{a}\boldsymbol{A}^{b}$, the advection matrices $\boldsymbol{\beta}^{\alpha}$ are also symmetric.

The physical interpretation of the terms in the advection-reaction equations is, firstly, that a combination of expansion coefficients $\boldsymbol{f}$ is advected in the reduced phase-space variables $\boldsymbol{\xi}$ and, secondly, the expansion coefficients are mixed and decay through the reaction matrix $\boldsymbol{R}$.

A unique solution to the system of PDEs (6.2) requires in addition boundary and initial conditions. In the subsequent section we choose a *zero inflow boundary condition*, because the mathematical results concerning the uniqueness of the solution, to which we refer the reader, hold for this choice. However, in `Sapphire++` other boundary conditions are implemented, namely periodic and continuous, which we use in the examples in Sec. 6.2. We denote with $\boldsymbol{f}^{-}$ the inflowing part of $\boldsymbol{f}$ at a specific point on the boundary ∂D of the domain $D \subset \mathbb{R}^4$. Zero inflow at all times is then formally expressed as $\boldsymbol{f}^{-} = 0$ on $\partial D \times [0, t_F]$ where t_F is the final time. We address the question of how to determine $\boldsymbol{f}^{-}$ later, see eq. (6.25) and the explanations thereafter. As an initial condition we choose a smooth function $\boldsymbol{f}(\boldsymbol{x}, p, t = 0) = \boldsymbol{f}^{0}(\boldsymbol{x}, p)$.

Furthermore, we include an additional source term $\boldsymbol{s}(\boldsymbol{x}, p, t)$, which is at least a square-integrable function. The problem we seek to solve thus becomes

$$
\partial_t \boldsymbol{f} + (\boldsymbol{\beta}\cdot\tilde{\nabla})\boldsymbol{f} + \boldsymbol{R}\boldsymbol{f} = \boldsymbol{s} \qquad \text{in } D \times [0, t_F] \tag{6.4}
$$

$$
\boldsymbol{f}^{-} = 0 \qquad \text{on } \partial D \times [0, t_F] \tag{6.5}
$$

$$
\boldsymbol{f}(\boldsymbol{x}, p, 0) = \boldsymbol{f}^{0} \qquad \text{in } D\,. \tag{6.6}
$$

Under the assumption that $\boldsymbol{\beta}$ and $\boldsymbol{R}$ do not depend on time it can be shown via an *energy estimate*, that if a solution $\boldsymbol{f}$ exists, the solution is unique, see Di Pietro and Ern (2012, p.70 Lemma 3.2 and p.332 Lemma 7.26).[3]

[3]We note that Di Pietro & Ern would interpret such a system as an example of a *Friedrich's system*, cf. Di Pietro and Ern (2012, Section 7.1 and Section 7.5)

As mentioned, the advection-reaction system is a linear hyperbolic system. A well established approach to solve such a system numerically is the finite volume (FV) method, because it includes fluxes which allow it to conserve the relevant physical quantities. We apply the discontinuous Galerkin method instead of the FV method, because the dG method is also based on fluxes and, thus, has the same main advantage but, in contrast to FV methods, it is easy to increase the order of accuracy of the spatial discretisation of the solution to the PDE system (6.2). FV methods rely on polynomial reconstruction methods like the (weighted) essentially non-oscillatory (WENO) method, which requires a stencil of cells to achieve higher order accuracy (Shu, 2009). As we show in the next section, the dG method works right away with higher order polynomials, that are independently defined on each cell, thus avoiding the need for a reconstruction algorithm with information from neighbouring cells. In this sense the dG method is more *local* than FV methods. It is this locality which helps to leverage the implementation of algorithms which adapt the cell sizes or the polynomial degree depending on the error of the numerical solution. This is useful in the context of the acceleration of particles around a shock, which benefits from high accuracy in the vicinity of the shock, see for example the grid design in Fig. 6.5.

6.1.2 DISCRETE REPRESENTATION OF THE SOLUTION AND THE FINITE ELEMENT METHOD

We now apply the dG method to the system of PDEs (6.2) and explain how we can exploit properties of the $\boldsymbol{A}^a$ matrices to accelerate and stabilise the resulting algorithm. We note that the content up to eq. (6.18) heavily draws from Di Pietro and Ern (2012, in particular Chap. 1 – 3 and Chap. 7).

The dG method is a finite element method (FEM) in which the discrete approximation of the solution to the PDE is represented by a linear combination of functions, i.e.

$$\boldsymbol{f}_h(t, \boldsymbol{x}, p) = \zeta_j(t)\boldsymbol{\phi}_j(\boldsymbol{x}, p) \quad \text{with } \boldsymbol{\phi}_k \in V_h\,, \tag{6.7}$$

where the summation over j is implied. V_h is a finite dimensional function space and the $\boldsymbol{\phi}_j$ are its basis functions. The subscript h refers to a typical cell size and expresses that V_h depends among other things on the number of cells in which the domain D will be decomposed. The objective of FEMs is to determine the coefficients ζ_j that are called *degrees of freedom* (DoF).

One of the features that distinguishes the dG method from other FEMs is the choice of the function space V_h: The domain D is subjected to a triangulation. The outcome is a set of cells in the reduced phase-space which we denote with $\mathcal{T}_h$. We use lines in one dimension, rectangles in two dimensions and cuboids in three dimensions[4], because `Sapphire++` builds on the `deal.ii` library that

[4]Hypercuboids can be applied in higher dimensions, though `Sapphire++` is not yet equipped to handle these within the standard `deal.ii` framework. Extensions to higher dimensions are implemented in `hyper.deal`, as described in (Munch et al., 2021).

preferably uses this geometry for the cells. Subsequently, a set of functions is defined on each cell T, for example, polynomials up to a certain degree k. These functions are constructed as *tensor products* of 1D *Lagrange polynomials*. Each factor of the tensor product corresponds to one dimension of the cell. The 1D Lagrange polynomials are

$$\ell_i(x) = \prod_{\substack{0\le j\le k \\ i\neq j}} \frac{x - x_j}{x_i - x_j} \quad \text{with } i \in \{0, \dots, k\}, \tag{6.8}$$

where k is the number of interpolation nodes and, thus, the degree of the Lagrange polynomial ℓ_i. For $k \le 2$, the Lagrange polynomials are constructed with equidistant interpolation nodes x_i, whereas for $k > 2$ *Gauss–Lobatto* points are used[5]. We note that the $k+1$ Gauss–Lobatto points are found by forming the union of the roots of the derivative of the degree-k Legendre polynomial $P_k'(x)$ and the interval endpoints $\{-1, 1\}$ (Hesthaven & Warburton, 2008; Milton & Stegun, 1964, p.47 and eq. 25.4.32 respectively).

Because the expansion coefficients f_{lms} of the spherical harmonic expansion depend on $\boldsymbol{\xi} \in \mathbb{R}^{d+1}$, $d+1$ Lagrange polynomials are used to form the tensor product, which in this context is just the ordinary product of the polynomials. For example, for $d = 2$, $\boldsymbol{\xi} = (x, y, p) \in T \subset \mathbb{R}^{d+1}$ and the set of functions defined on the cell T is

$$\mathbb{Q}^k(T) = \operatorname{span}\{\ell_i(x)\ell_j(y)\ell_m(p)\} \qquad \text{with } i, j, m \in \{0, \dots, k\}. \tag{6.9}$$

We note that $\mathbb{Q}^k(T)$ is a vector space with dimension $\dim(\mathbb{Q}^k(T)) = (k+1)^{d+1}$.

The space $\mathbb{Q}^k(T)$ can only be used to represent *one* of the $n = (l_{\max}+1)^2$ expansion coefficients that we collected in the vector $\boldsymbol{f}$. We remedy this with a copy of $\mathbb{Q}^k(T)$ for each expansion coefficient, i.e. we introduce the space $[\mathbb{Q}^k(T)]^n$ with dimension $\dim([\mathbb{Q}^k(T)]^n) = n(k+1)^{d+1}$. Notice that $[\mathbb{Q}^k(T)]^n$ is a space with vectors $\boldsymbol{\phi} \in \mathbb{R}^n$ whose components are elements of $\mathbb{Q}^k(T)$, i.e. they are linear combinations of the products of the Lagrange polynomials.

Eventually, the space V_h can be defined as the direct sum of $[\mathbb{Q}^k(T)]^n$ over of all cells, i.e.

$$V_h \coloneqq \bigoplus_{T\in\mathcal{T}_h} [\mathbb{Q}^k(T)]^n. \tag{6.10}$$

This means that every element in V_h is a sum of the polynomials defined on each cell. Since there is *no* requirement that this sum has to give a continuous function at the cell interfaces, elements of V_h are expected to be discontinuous at cell faces. Such a space is called a *broken polynomial space* (Di Pietro & Ern, 2012, Sections 1.2.4.2 - 3). Moreover, it is from this that the *discontinuous* Galerkin name arises. As stated in the definition of the discrete solution, presented in

[5] Cf. `deal.ii` manual (FE_Q)

eq. (6.7), the functions $\boldsymbol{\phi}_j(\boldsymbol{x}, p)$ are the basis vectors of V_h. The total number of DoFs is the same as the total number of basis functions, namely

$$N = \mathrm{card}(\mathcal{T}_h)\dim([\mathbb{Q}^k(T)]^n) = \mathrm{card}(\mathcal{T}_h)n(k+1)^{d+1}\,, \tag{6.11}$$

where $\mathrm{card}(\mathcal{T}_h)$ is the number of cells in the triangulation $\mathcal{T}_h$.

As with all FEMs, we seek to construct a linear system to determine the DoF $\zeta_j(t)$. This is achieved by multiplying eq. (6.4) with a basis function $\boldsymbol{\phi}_i \in V_h$ from the left, replacing $\boldsymbol{f}$ with its discrete counterpart $\boldsymbol{f}_h \in V_h$ and integrating the equation over the domain $D = \bigcup_{T\in\mathcal{T}_h} T$. This yields

$$\begin{aligned}\sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\left(\partial_t\boldsymbol{f}_h + (\boldsymbol{\beta}\cdot\tilde{\nabla})\boldsymbol{f}_h + \boldsymbol{R}\boldsymbol{f}_h\right) &= \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\left(\partial_t\boldsymbol{\phi}_j + (\boldsymbol{\beta}\cdot\tilde{\nabla})\boldsymbol{\phi}_j + \boldsymbol{R}\boldsymbol{\phi}_j\right)\zeta_j \\ &= \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\boldsymbol{s}\end{aligned} \tag{6.12}$$

for each $i \in \{1,\dots,N\}$. Notice that the integration variables are implicit to improve the readability, i.e. we did not include $\mathrm{d}^{d+1}\xi$ in the integral expressions.

Keeping in mind that $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ are defined 'cell-wise', eq. (6.12) yields a set of equations for each cell T, which is independent of the set of equations determining the coefficients ζ_j on the neighbouring cells. In the language of FEMs, the degrees of freedom on one cell are decoupled from the degrees of freedom on neighbouring cells. Physically, we expect a flux from one cell to the next, because we are solving an advection equation. Thus, we expect that the degrees of freedom of different cells *do* couple and, hence, the linear system (6.12) will *not* yield a correct approximation of the solution to the system of PDEs given in eq. (6.2).

6.1.3 NUMERICAL FLUX

The usual approach to enforce a coupling of the DoFs on neighbouring cells is to manipulate the linear system (6.12) in such a way that there is a flux from one cell to the next, while taking care that the original equation (6.4) is recovered if we use the exact solution $\boldsymbol{f}$ instead of its approximation $\boldsymbol{f}_h$, i.e. that the manipulated system is *consistent* with the original problem.

To investigate the fluxes between the cells, we integrate the advection term $\boldsymbol{\beta}\cdot\tilde{\nabla}\boldsymbol{f}_h$ by parts and apply the divergence theorem, i.e.

$$\begin{aligned}&\sum_{T\in\mathcal{T}_h}\int_T (\boldsymbol{\phi}_i)_k\boldsymbol{\beta}^\alpha_{kl}\tilde{\nabla}_\alpha(\boldsymbol{f}_h)_l \\ &\quad= \sum_{T\in\mathcal{T}_h}\int_{\partial T} n_\alpha\left((\boldsymbol{\phi}_i)_k\boldsymbol{\beta}^\alpha_{kl}(\boldsymbol{f}_h)_l\right) - \int_T (\boldsymbol{\phi}_i)_k\tilde{\nabla}_\alpha\boldsymbol{\beta}^\alpha_{kl}(\boldsymbol{f}_h)_l - \int_T \tilde{\nabla}_\alpha(\boldsymbol{\phi}_i)_k\boldsymbol{\beta}^\alpha_{kl}(\boldsymbol{f}_h)_l \\ &\quad=: \sum_{T\in\mathcal{T}_h}\int_{\partial T}\boldsymbol{\phi}_i\cdot(\boldsymbol{n}\cdot\boldsymbol{\beta})\boldsymbol{f}_h - \int_T \boldsymbol{\phi}_i\cdot(\tilde{\nabla}\cdot\boldsymbol{\beta})\boldsymbol{f}_h - \int_T(\tilde{\nabla}\boldsymbol{\phi}_i\cdot\boldsymbol{\beta})\boldsymbol{f}_h\,,\end{aligned} \tag{6.13}$$

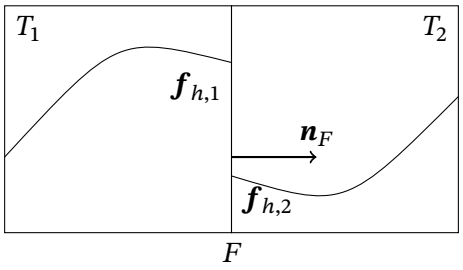

Figure 6.1: Two adjacent cells T_1 and T_2. The discrete representation of the solution $\boldsymbol{f}_h$ is not continuous on the cell interface (Di Pietro & Ern, 2012, cf. Fig. 1.4).

where ∂T denotes the surface of the cell T.

We define the vector $\boldsymbol{J} \coloneqq (\boldsymbol{n} \cdot \boldsymbol{\beta})\boldsymbol{f}$ that has the physical interpretation of a flux in the direction of the normal $\boldsymbol{n}$. This becomes clear when we look at an arbitrary component, say $(\boldsymbol{J})_i = n_\alpha \boldsymbol{\beta}^\alpha_{ij}(\boldsymbol{f})_j =: n_\alpha \boldsymbol{j}^\alpha_i$. The matrices $\boldsymbol{\beta}^\alpha$ 'mix' the components of $\boldsymbol{f}$, which results in a current density $\boldsymbol{j}_i$. This current density is projected onto the normal $\boldsymbol{n}$ and each component of $(\boldsymbol{J})_i$ is the projection of a different current density $\boldsymbol{j}_i$ onto $\boldsymbol{n}$.

In a next step, we focus on the sum over the surface integrals in eq. (6.13) and, in particular, we will look at a single cell interface F as the one depicted in Fig. 6.1. An integral over such a cell interface F consists in a contribution from cell T_1 and another contribution from cell T_2, namely

$$\begin{aligned} &\int_F \boldsymbol{\phi}_{i_1} \cdot (\boldsymbol{n}_{T_1} \cdot \boldsymbol{\beta})\boldsymbol{f}_{h,1} + \boldsymbol{\phi}_{i_2} \cdot (\boldsymbol{n}_{T_2} \cdot \boldsymbol{\beta})\boldsymbol{f}_{h,2} \\ &\qquad = \int_F \boldsymbol{\phi}_{i_1} \cdot (\boldsymbol{n}_F \cdot \boldsymbol{\beta})\boldsymbol{f}_{h,1} - \boldsymbol{\phi}_{i_2} \cdot (\boldsymbol{n}_F \cdot \boldsymbol{\beta})\boldsymbol{f}_{h,2} \end{aligned} \tag{6.14}$$

The introduction of the subscript 1 and 2 reflects that the basis functions of V_h are defined on each cell, i.e. there is a set of basis functions defined on T_1 and another one defined on T_2. Moreover, the outward normal $\boldsymbol{n}_{T_1} = \boldsymbol{n}_F$ and $\boldsymbol{n}_{T_2} = -\boldsymbol{n}_F$, see Fig. 6.1. Note that the flux through F is not unique, because $\boldsymbol{f}_h$ is discontinuous. Physically, we expect that $\boldsymbol{f}$ is continuous and, hence, that the flux is single-valued, which motivates the replacement of the two fluxes appearing in eq. (6.14) with a *numerical flux* $\mathring{\boldsymbol{J}}_F(\boldsymbol{f}_{h,1}, \boldsymbol{f}_{h,2})$, which is a single-valued function of both values of $\boldsymbol{f}_h$. To ensure consistency, we have to require that the numerical flux reduces to the physical flux, if we use the exact solution $\boldsymbol{f}$, i.e. $\mathring{\boldsymbol{J}}_F(\boldsymbol{f}, \boldsymbol{f}) = \boldsymbol{J} = (\boldsymbol{n} \cdot \boldsymbol{\beta})\boldsymbol{f}$. Using the numerical flux changes the integral over the cell interface F in eq. (6.14) to

$$\int_F (\boldsymbol{\phi}_{i_1} - \boldsymbol{\phi}_{i_2}) \cdot \mathring{\boldsymbol{J}}_F(\boldsymbol{f}_{h,1}, \boldsymbol{f}_{h,2}) =: \int_F [\![\boldsymbol{\phi}_i]\!] \cdot \mathring{\boldsymbol{J}}_F(\boldsymbol{f}_{h,1}, \boldsymbol{f}_{h,2}), \tag{6.15}$$

where we defined the symbol $[\![\boldsymbol{\phi}_i]\!]$ to denote the jump of the basis functions at a cell interface.

We now replace the sum over the surface integrals in eq. (6.13) with a sum over cell interface and boundary face integrals, namely

$$\begin{aligned}\sum_{T\in\mathcal{T}_h}\int_{\partial T}\boldsymbol{\phi}_i\cdot(\boldsymbol{n}\cdot\boldsymbol{\beta})\boldsymbol{f}_h \\ \longrightarrow \sum_{F\in\mathcal{F}_h^i}\int_F [\![\boldsymbol{\phi}_i]\!]\cdot\mathring{\boldsymbol{J}}_F(\boldsymbol{f}_{h,1},\boldsymbol{f}_{h,2}) + \sum_{F\in\mathcal{F}_h^b}\int_F \boldsymbol{\phi}_i\cdot\mathring{\boldsymbol{J}}_F^B(\boldsymbol{f}_h)\,.\end{aligned} \tag{6.16}$$

We stress that the above transition includes the usage of the numerical flux, which is a deliberate manipulation of the linear system (6.12). Furthermore, we introduced the sets $\mathcal{F}_h^i$ and $\mathcal{F}_h^b$ whose elements are the cell interfaces and the faces of the boundary cells respectively. $\mathring{\boldsymbol{J}}_F^B(\boldsymbol{f}_h)$ denotes the numerical flux through the latter.

Eventually, we obtain the manipulated linear system for the coefficients ζ_j that couples neighbouring DoFs through two replacements: First, we replace the sum over the surface integrals in eq. (6.13) with the sum over the face integrals in eq. (6.16). Secondly, we replace the advection term in the original linear system (6.12) with the result of the previous replacement. This yields

$$\begin{aligned}\sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\partial_t\boldsymbol{f}_h + \int_T\boldsymbol{\phi}_i\cdot\{\boldsymbol{R}-(\tilde{\nabla}\cdot\boldsymbol{\beta})\}\boldsymbol{f}_h - \int_T(\tilde{\nabla}\boldsymbol{\phi}_i\cdot\boldsymbol{\beta})\boldsymbol{f}_h \\ + \sum_{F\in\mathcal{F}_h^i}\int_F[\![\boldsymbol{\phi}_i]\!]\cdot\mathring{\boldsymbol{J}}_F(\boldsymbol{f}_{h,1},\boldsymbol{f}_{h,2}) + \sum_{F\in\mathcal{F}_h^b}\int_F\boldsymbol{\phi}_i\cdot\mathring{\boldsymbol{J}}_F^B(\boldsymbol{f}_h) = \sum_{T\in\mathcal{T}_h}\int_T\boldsymbol{\phi}_i\cdot\boldsymbol{s}\end{aligned} \tag{6.17}$$

for each $i\in\{1,\dots,N\}$. We note that this is a system of ordinary differential equations (ODEs) which determines the degrees of freedom $\zeta_j(t)$ of $\boldsymbol{f}_h$, see eq. (6.7). We call the system *semi-discretised*, because it is discretised in space but not in time.

Whether the solution to the system of ODEs (6.17) approximates the exact solution $\boldsymbol{f}$ depends on many things, inter alia, on the choice of the numerical flux $\mathring{\boldsymbol{J}}_F$ and on the time stepping method used. For example, an *explicit Runge–Kutta method* (ERK) of order two (or three) only converges over time if we assume that the exact solution $\boldsymbol{f}$ and the source term $\boldsymbol{s}$ are smooth enough. We note that any time stepping method only converges if the time step is chosen in agreement with a suitable *CFL-condition*[6]. We conclude that if we choose an appropriate flux, together with an ERK method, and if the exact solution and the source term are smooth enough, the dG method converges in time and space to the exact solution (Di Pietro & Ern, 2012, Lemma 7.27 and Lemma 7.28 and references therein).

[6]Information on time stepping methods like ERK can, for example, be found in Hairer et al. (1993, Chapter II) and an explanation of the CFL-condition is given in LeVeque, 2002, Section 4.4

In general, there are many different choices for the numerical flux, a typical one for a system of equations is the *(local) Lax–Friedrichs flux* (Cockburn, 1998, p. 204). However, if a problem is advection dominated, it makes sense to use this knowledge to determine a more precise numerical flux. An upwind flux does exactly this, cf. LeVeque (2002, Section 4.8).

The upwind flux is defined as

$$\mathring{\boldsymbol{J}}_F^U(\boldsymbol{f}_{h,1}, \boldsymbol{f}_{h,2}) \coloneqq \boldsymbol{W}\left(\boldsymbol{\Lambda}_+ \boldsymbol{W}^\mathsf{T} \boldsymbol{f}_{h,1} + \boldsymbol{\Lambda}_- \boldsymbol{W}^\mathsf{T} \boldsymbol{f}_{h,2}\right) \tag{6.18}$$

where $\boldsymbol{W}$ and $\boldsymbol{\Lambda}$ are the eigenvectors and eigenvalues of the matrix $(\boldsymbol{n} \cdot \boldsymbol{\beta})$. The eigenvalue matrix is written as a sum of two matrices, i.e. $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_+ + \boldsymbol{\Lambda}_i$ where $\boldsymbol{\Lambda}_+$ is a matrix with positive eigenvalues and zeros on its diagonal and $\boldsymbol{\Lambda}_-$ is a corresponding matrix with the negative eigenvalues and zeros on its diagonal. This definition of the upwind flux can, for example, be found in Hesthaven and Warburton (2008, Section 2.4).

We briefly summarise an insightful physical interpretation of the upwind flux (cf. LeVeque, 2002, p. 47). We study a system of advection equations with initial conditions that are discontinuous at $x = 0$, i.e.

$$\partial_t \boldsymbol{q}(x,t) + \boldsymbol{A}\partial_x \boldsymbol{q}(x,t) = 0 \quad \text{with } \boldsymbol{q}(x,0) \coloneqq \boldsymbol{q}^0 \coloneqq \begin{cases} \boldsymbol{q}_1 & \text{for } x < 0 \\ \boldsymbol{q}_2 & \text{for } x \geq x \end{cases}. \tag{6.19}$$

Such a problem is called a *Riemann problem* and we emphasise that it has to be solved at each cell interface F, because $\boldsymbol{f}_h$ is discontinuous, see Fig. (6.1). The solution to this Riemann problem is found by diagonalising $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$ and multiplying eq. (6.19) with $\boldsymbol{V}^{-1}$ from the left. This yields

$$\partial_t \tilde{\boldsymbol{q}} + \boldsymbol{\Lambda}\partial_x \tilde{\boldsymbol{q}} = 0 \quad \text{with } \tilde{\boldsymbol{q}} \coloneqq \boldsymbol{V}^{-1}\boldsymbol{q}\,, \tag{6.20}$$

which is a set of decoupled scalar advection equations and the functions $\tilde{\boldsymbol{q}}$ are called *characteristic variables*. The decoupled advection equations have the solutions $(\tilde{\boldsymbol{q}})_i(x,t) = (\tilde{\boldsymbol{q}}^0)_i(x-\lambda_i t) = (\boldsymbol{V}^{-1}\boldsymbol{q}^0)_i(x-\lambda_i t)$, which can be interpreted as 'shifting' the values of the transformed initial conditions $\tilde{\boldsymbol{q}}_0$ in time at a speed determined by the eigenvalues λ_i, i.e. the profile $\tilde{\boldsymbol{q}}^0$ given at $t = 0$ is advected. The corresponding flux is $\lambda_i(\tilde{\boldsymbol{q}})_i$ and the sign of the eigenvalue determines its direction.

Bearing in mind that $(\boldsymbol{n} \cdot \boldsymbol{\beta})$ is symmetric, the terms $\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_{h,i}$ in the upwind flux (6.18) can physically be interpreted as the characteristic variables of the Riemann problem at a cell interface F. They are multiplied with the eigenvalues of $(\boldsymbol{n}_F \cdot \boldsymbol{\beta})$, which yields a flux. Positive eigenvalues result in a flux pointing in the direction of the normal $\boldsymbol{n}_F$ and negative eigenvalues produce a flux pointing in the opposite direction. The idea of 'upwinding' is to use $f_{h,1}$, i.e. the value of $\boldsymbol{f}_h$ on the left side of the interface, to compute the flux along $\boldsymbol{n}_F$ and $\boldsymbol{f}_{h,2}$ for the flux in the opposite direction. Eventually, the fluxes are multiplied with $\boldsymbol{W}$ to restore the original variables.

A local Lax–Friedrichs flux only needs the maximum eigenvalue of $\boldsymbol{n}_F \cdot \boldsymbol{\beta}$ whereas an upwind flux requires the diagonalisation of $n \times n$ matrices, where $n = (l_{\max} + 1)^2$, at each interface in every time step. We note that the size of the matrices grows quadratically with the order of the spherical harmonic expansion and, if possible, it is best to avoid the repeated computation of eigenvectors and eigenvalues of so large matrices.

However, if we restrict the triangulation $\mathcal{T}_h$ to (hyper-)rectangles, we are able to avoid the necessity to solve an eigenproblem at each interface. Considering that all normals of the faces of a rectangle can be chosen to be parallel to the coordinate axes, the matrix $(\boldsymbol{n}_F \cdot \boldsymbol{\beta})$ simplifies. For example, if we are interested in the upwind flux through a face whose normal points in the x-direction, i.e. $\boldsymbol{n}_F = \boldsymbol{e}_x$, then

$$(\boldsymbol{n}_F \cdot \boldsymbol{\beta}) = (\boldsymbol{e}_x \cdot \boldsymbol{\beta}) = \boldsymbol{\beta}^x = U^x \mathbf{1} + v \boldsymbol{A}^x \,, \tag{6.21}$$

where we used the definition of $\boldsymbol{\beta}^\alpha$ given in eq. (6.3).

Now, let $\boldsymbol{w}$ be an eigenvector of $\boldsymbol{A}^x$ with eigenvalue λ, then

$$(U^x \mathbf{1} + v \boldsymbol{A}^x)\, \boldsymbol{w} = U^x \boldsymbol{w} + \lambda v \boldsymbol{w} = (U^x + \lambda v) \boldsymbol{w} \,. \tag{6.22}$$

Hence, $\boldsymbol{w}$ is *also* an eigenvector of $\boldsymbol{\beta} \cdot \boldsymbol{e}_x = \boldsymbol{\beta}^x = U^x \mathbf{1} + v \boldsymbol{A}^x$, and the corresponding eigenvalue is $U^x + \lambda v$.

We conclude that we have to determine the eigenvectors and eigenvalues of $\boldsymbol{A}^x$ *once* to get the upwind flux in the x-direction at all interfaces and at all times, because the eigenvectors of $\boldsymbol{\beta}^x$ do *not* change and the eigenvalues can be updated by multiplying them with v and adding U^x, see eq. (6.22).

The same is true for the upwind fluxes in the y- and z-direction. The only difference is that we do *not* have to compute the eigenvalues and eigenvectors of $\boldsymbol{A}^y$ and $\boldsymbol{A}^z$. In Sec. 5.4, we showed that all three matrices have the same eigenvalues and that the eigenvectors of $\boldsymbol{A}^y$ and $\boldsymbol{A}^z$ can be computed by rotating the eigenvectors of $\boldsymbol{A}^x$.

The upwind flux in the momentum direction, i.e. in the p-direction, is different, because

$$(\boldsymbol{n}_p \cdot \boldsymbol{\beta}) = \boldsymbol{\beta}^p = -\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t} \boldsymbol{A}^a - p \frac{\partial U_b}{\partial x^a} \boldsymbol{A}^a \boldsymbol{A}^b \tag{6.23}$$

contains sums and products of the $\boldsymbol{A}^a$ matrices and the above arguments, which we used to avoid finding a solution to the eigenproblem, do not apply. Thus, we end up solving an eigenproblem at each interface whose normal points in the p-direction.

6.1.4 NUMERICAL FLUX AT THE BOUNDARIES OF THE DOMAIN

Having defined the numerical flux to be the upwind flux, we almost have an explicit form of the linear system (6.17), which determines the approximate

solution $\boldsymbol{f}_h$. 'Almost', because we have not yet defined the numerical flux $\mathring{\boldsymbol{J}}_F^B(\boldsymbol{f}_h)$ through the boundary.

We note that the approximate solution $\boldsymbol{f}_h$ must fulfil the zero inflow boundary condition (6.5) and that the choice of $\mathring{\boldsymbol{J}}_F^B$ can enforce it. It is this idea that informs the definition of the boundary flux. Assume it was the upwind flux (6.18) as well and that $\boldsymbol{f}_h$ was single-valued on the boundary, i.e. $\boldsymbol{f}_h = \boldsymbol{f}_{h,1} = \boldsymbol{f}_{h,2}$, then the boundary flux would be

$$\mathring{\boldsymbol{J}}_F^B(\boldsymbol{f}_h) = \boldsymbol{W}\big(\boldsymbol{\Lambda}_+\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_h + \boldsymbol{\Lambda}_-\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_h\big)\,, \tag{6.24}$$

and its second term would be the flux into the domain. We can thus enforce zero inflow by setting

$$\boldsymbol{W}\boldsymbol{\Lambda}_-\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_h = 0 \iff \boldsymbol{\Lambda}_-\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_h = 0\,. \tag{6.25}$$

Notice that $\boldsymbol{\Lambda}_-\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_h$ 'picks out' the inflow part of $\boldsymbol{f}_h$, because the multiplication with $\boldsymbol{W}^\mathsf{T}$ yields the characteristic variables and the multiplication with $\boldsymbol{\Lambda}_-$ eliminates all the characteristic variables which do *not* contribute to the inflow. The reason is that the diagonal elements of $\boldsymbol{\Lambda}_-$ corresponding to the outflowing components are zero. This motivates to define the boundary flux to be

$$\mathring{\boldsymbol{J}}_F^B(\boldsymbol{f}_h) \coloneqq \boldsymbol{W}\boldsymbol{\Lambda}_+\boldsymbol{W}^\mathsf{T}\boldsymbol{f}_h\,. \tag{6.26}$$

An implication of setting the inflow components of $\boldsymbol{f}_h$ to zero on the boundary is that the discrete solution $\boldsymbol{f}_h$ fulfils the zero inflow condition (6.5) (cf. Di Pietro & Ern, 2012, Proposition 2.7). We emphasise that we do *not* prescribe values of $\boldsymbol{f}$ on the boundary to enforce this. It is implicit in the definition of $\mathring{\boldsymbol{J}}_F^B$ and every solution $\boldsymbol{f}_h$ to the linear system (6.17) of ODEs is automatically in agreement with it. In the language of FEMs it is said that the boundary conditions are enforced *weakly*[7], which is typical for dG methods.

At the end of this section, we take up a loose thread and note that we now have the concepts to properly define the meaning of $\boldsymbol{f}^-$ used in the formulation of the zero inflow boundary condition, namely we formalise the 'picking-out' of inflow components by introducing the matrix $\mathbf{1}_-$ that has ones where $\boldsymbol{\Lambda}_-$ has non-zero entries and zeros everywhere else. We use this matrix to define the inflow part of $\boldsymbol{f}$, i.e. $\boldsymbol{f}^- \coloneqq \mathbf{1}_-\boldsymbol{W}^T\boldsymbol{f}$.

6.1.5 TIME STEPPING METHOD

With the definition of the boundary flux we finally have an explicit expression for all the terms in the system of ODEs (6.17) and we can now solve it using any of the standard time stepping methods.

[7]Weakly enforced boundary conditions do *not* hold on every point on the boundary, they only hold *almost everywhere*.

In a first step, we bring the system of ODEs in a particularly simple form to ease the application of a time stepping method, i.e.

$$\boldsymbol{M}\frac{\mathrm{d}\boldsymbol{\zeta}}{\mathrm{d}t} + \boldsymbol{D}(t)\boldsymbol{\zeta} = \boldsymbol{h}(t)\,, \tag{6.27}$$

where the components of the vector $\boldsymbol{\zeta}$ are the degrees of freedom of the approximate solution $\boldsymbol{f}_h$, see eq. (6.7). Furthermore, we introduced the symbols

$$\begin{aligned}
(\boldsymbol{M})_{ij} &\coloneqq \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\boldsymbol{\phi}_j\,,\\
(\boldsymbol{D})_{ij} &\coloneqq \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\left\{\boldsymbol{R}-(\tilde{\nabla}\cdot\boldsymbol{\beta})\right\}\boldsymbol{\phi}_j - \int_T(\tilde{\nabla}\boldsymbol{\phi}_i\cdot\boldsymbol{\beta})\boldsymbol{\phi}_j\\
&\quad + \sum_{F\in\mathcal{F}_h^i}\int_F [\![\boldsymbol{\phi}_i]\!]\cdot\mathring{\boldsymbol{J}}_F(\boldsymbol{\phi}_{j,1},\boldsymbol{\phi}_{j,2}) + \sum_{F\in\mathcal{F}_h^b}\int_F \boldsymbol{\phi}_i\cdot\mathring{\boldsymbol{J}}_F^B(\phi_j) \quad \text{and}\\
(\boldsymbol{h})_i &\coloneqq \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\boldsymbol{s}\,.
\end{aligned} \tag{6.28}$$

We apply the *Θ-method* to time step, i.e.

$$\boldsymbol{M}\frac{\boldsymbol{\zeta}^n-\boldsymbol{\zeta}^{n-1}}{\Delta t} = (1-\Theta)\left(\boldsymbol{h}^{n-1}-\boldsymbol{D}^{n-1}\boldsymbol{\zeta}^{n-1}\right) + \Theta\left(\boldsymbol{h}^n-\boldsymbol{D}^n\boldsymbol{\zeta}^n\right)\,, \tag{6.29}$$

which allows us to switch between an implicit and an explicit time stepping. The superscript n means that the respective quantity is evaluated at time step n, for example, $\boldsymbol{\zeta}^n \coloneqq \boldsymbol{\zeta}(n\Delta t)$. Θ takes values in the interval $[0,1]$ and $\Theta = 0$ results in the *forward (or explicit) Euler method*, $\Theta = 1$ gives the *backward (or implicit) Euler method* and $\Theta = 1/2$ corresponds to the *Crank–Nicolson method*.

The initial conditions $\boldsymbol{\zeta}^0$ for eq. (6.27) are the coefficients of $\boldsymbol{f}_h(0) = \zeta_j(t=0)\boldsymbol{\phi}_j$, i.e. the coefficients of the approximate solution at $t=0$. We use the initial condition $\boldsymbol{f}(\boldsymbol{x},p,0) = \boldsymbol{f}^0$ for the exact solution to compute $\boldsymbol{\zeta}^0$. This is achieved by projecting the initial conditions onto the finite element space, i.e.

$$\sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\boldsymbol{f}_h(0) = \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\boldsymbol{\phi}_j\zeta_j(t=0) = (\boldsymbol{M})_{ij}(\boldsymbol{\zeta}^0)_j = \sum_{T\in\mathcal{T}_h}\int_T \boldsymbol{\phi}_i\cdot\boldsymbol{f}^0\,. \tag{6.30}$$

The solution to this linear system of equations is $\boldsymbol{\zeta}^0$.

We highlight that the Θ-method is an implicit method for $\Theta > 0$, i.e. $\boldsymbol{\zeta}^n$ appears on both sides of eq. (6.29). We rearrange eq. (6.29) for $\boldsymbol{\zeta}^n$ and arrive at

$$(\boldsymbol{M}+\Delta t\Theta\boldsymbol{D}^n)\,\boldsymbol{\zeta}^n = \left(\boldsymbol{M}-\Delta t(1-\Theta)\boldsymbol{D}^{n-1}\right)\boldsymbol{\zeta}^{n-1} + \Delta t\left((1-\Theta)\boldsymbol{h}^{n-1}+\Theta\boldsymbol{h}^n\right)\,.$$

This linear system of equations is solved *iteratively* in every time step.

We implemented the dG method and the Θ-method, as outlined in this section, using the finite element library `deal.ii` (Arndt et al., 2021, 2023) in a free and open-source code called `Sapphire++`. We also implemented an explicit fourth order Runge-Kutta method (ERK4) in `Sapphire++`.

Table 6.1: Units and their reference values in Sapphire++

Unit	Definition	Reference	Value		
t^*	$t\underline{\omega}_g$	$\underline{\omega}_g \coloneqq \underline{q}\underline{B}/\underline{m}$	9.579×10^{-3} s^{-1}		
x^*	$x/\underline{r}_g$	$\underline{r}_g \coloneqq \underline{m}\underline{c}/\underline{q}\underline{B}$	3.130×10^{10} m	10^{-6}	pc
p^*	$p/\underline{p}$	$\underline{p} \coloneqq \underline{m}\underline{c}$	5.014×10^{-19} kg m s^{-1}	938	MeV c^{-1}
v^*	$v/\underline{c}$	$\underline{c}$	2.998×10^{8} m s^{-1}		
q^*	$q/\underline{q}$	$\underline{q} \coloneqq e$	1.602×10^{-19} C		
m^*	$m/\underline{m}$	$\underline{m} \coloneqq m_p$	1.673×10^{-27} kg	938	MeV c^{-2}
B^*	$B/\underline{B}$	$\underline{B}$	1.000×10^{-10} T	1	μG

6.2 TESTS AND SIMULATIONS

In this section we investigate the abilities of Sapphire++ in two examples and eventually we apply it to simulate the acceleration of particles at a parallel shock.

The two examples have been selected to showcase specific features of the code and to highlight its *numerical* accuracy. In particular, the first test case shows that the dG space discretisation together with the various time stepping method convergences as theoretically expected. The second test case investigates consequences of the truncation of the spherical harmonic expansion at finite order $l_{\max}$. The last example, i.e. the simulation of diffusive shock acceleration at a parallel shock, shows that Sapphire++ is applicable to actual astrophysical scenarios.

We use dimensionless units when solving the system of PDEs (6.1). The definitions of the units and their reference values are given in Tab. 6.1. Length and time are defined in terms of a reference gyroradius and a gyrofrequency, motivated by the fact that Sapphire++ is written with physical effects occurring on gyroscales in mind.

The Sapphire++ code is designed in a way that the different terms of the VFP equation can be included or excluded in the simulation as required. We apply the following naming scheme:

$$\overbrace{\frac{\partial \boldsymbol{f}}{\partial t}}^{\text{time-evolution term}} + \overbrace{(U^a \mathbf{1} + v\boldsymbol{A}^a)\frac{\partial \boldsymbol{f}}{\partial x^a}}^{\text{spatial advection term}} - \overbrace{\left(\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t}\boldsymbol{A}^a + p\frac{\partial U_b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^b\right)\frac{\partial \boldsymbol{f}}{\partial p}}^{\text{momentum term}} \tag{6.31}$$
$$+ \underbrace{\left(\frac{1}{v}\epsilon_{abc}\frac{\mathrm{D}U^a}{\mathrm{D}t}\boldsymbol{A}^b\boldsymbol{\Omega}^c + \epsilon_{bcd}\frac{\partial U^b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^c\boldsymbol{\Omega}^d\right)\boldsymbol{f}}_{\text{momentum term}} - \underbrace{\omega_a\boldsymbol{\Omega}^a\boldsymbol{f}}_{\text{rotation term}} + \underbrace{\nu\boldsymbol{C}\boldsymbol{f}}_{\text{collsion term}} = \underbrace{\boldsymbol{s}}_{\text{source term}}$$

We emphasise that it is possible to solve the above equation for different configuration space dimensions, namely for $\boldsymbol{x} \in \mathbb{R}^d$ with $d = 1, 2$. If the momentum terms are deactivated, i.e. if monoenergetic particles[8] are simulated d can equal

[8] Up to now, Sapphire++ can only simulate up to three dimensions of the *reduced phase-space*,

3. Additionally, we allow to choose between a linear momentum variable p and a logarithmic momentum variable $\ln p$. The momentum terms are adapted accordingly.

The first two examples retain the time-evolution, the spatial advection term and the rotation term. In the case of the acceleration of particles at a parallel shock all terms are included in the simulation. We note the three test cases are structured in the same way. They start with a description of the equations that are solved numerically and continue with explicitly stating which parameters we used for the presented simulations, i.e. how we setup `Sapphire++`. They end with a short discussion of the aspects we would like to demonstrate with the tests.

6.2.1 CONVERGENCE STUDY

In this example we demonstrate the numerical accuracy of the dG space discretisation and four different time-stepping methods by simulating a simple test-case whose exact solution we present in the next subsection. Moreover, we study how this accuracy changes with different time steps, cell sizes and the polynomial degree of the dG basis functions described in Section (6.1.2).

DESCRIPTION We consider a mono-energetic distribution of particles in a static background plasma ($\boldsymbol{U} = 0$) that is permeated with a magnetic field ($\boldsymbol{B} = B_0 \boldsymbol{e}_z$) with no scattering between the particles and the plasma ($\nu = 0$). This amounts to neglecting the collision, momentum and source terms in eq. (6.31). We allow only spatial derivatives in the x-direction reducing the spatial advection term. In this case the system of equations (6.1) reduces to

$$\frac{\partial \boldsymbol{f}}{\partial t} + v\boldsymbol{A}^x \frac{\partial \boldsymbol{f}}{\partial x} - \omega_z \boldsymbol{\Omega}^z \boldsymbol{f} = 0\,. \tag{6.32}$$

Physically, the system models a distribution function f, homogeneous in y and z, that describes charged particles gyrating about $\boldsymbol{B}$ in the $x-y$ plane.

To arrive at an analytic solution, we consider a toy model and truncate the spherical harmonic expansion at $l_{\max} = 1$. We compute the corresponding matrices $\boldsymbol{A}_x$ and $\boldsymbol{\Omega}_z$ with the formulae presented in eq. (B.4) and (B.7). This gives

$$\frac{\partial \boldsymbol{f}}{\partial t} + \begin{pmatrix} 0 & 0 & v/\sqrt{3} & 0 \\ 0 & 0 & 0 & 0 \\ v/\sqrt{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \frac{\partial \boldsymbol{f}}{\partial x} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\omega_z & 0 \\ 0 & \omega_z & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \boldsymbol{f} = 0\,, \tag{6.33}$$

$\boldsymbol{\xi}$. If the momentum term is deactivated, the reduced phase space is equivalent to configuration space $\boldsymbol{\xi} = \boldsymbol{x}$, i.e. simulations in up to three physical space dimensions are possible.

where $\boldsymbol{f} = (f_{000}\, f_{110}\, f_{100}\, f_{111})^\mathsf{T}$. An immediate consequence is that $f_{111} = \text{const.}$. We, thus, restrict the above system of PDEs to the first three expansion coefficients. For the initial condition of the zeroth order expansion coefficient, we choose a periodic function with period L, namely

$$f_{000}(0,x) = \sum_{n=-\infty}^{\infty} C_n \exp(\mathrm{i}k_n x)$$

where $k_n \coloneqq 2\pi n/L$ is the wave number and $C_n \in \mathbb{C}$. The other three expansion coefficients are identically zero, i.e. $f_{110}(0,x) = f_{100}(0,x) = f_{111}(0,x) = 0$. Moreover, we assume periodic boundary conditions for a box with length L, namely $\boldsymbol{f}(t,0) = \boldsymbol{f}(t,L)$. The periodic boundary conditions in x justify the solution ansatz

$$\boldsymbol{f}(t,x) = \sum_{n=-\infty}^{\infty} \boldsymbol{f}_n(t) \exp(\mathrm{i}k_n x) = \sum_{n=-\infty}^{\infty} \begin{pmatrix} f_{n,1}(t) \\ f_{n,2}(t) \\ f_{n,3}(t) \end{pmatrix} \exp(\mathrm{i}k_n x)\,. \tag{6.34}$$

Plugging this ansatz into the restricted eq. (6.33) and using the orthogonality of the functions $\exp(\mathrm{i}k_n x)$, we arrive at a system of ODEs for the expansion coefficient $\boldsymbol{f}_n$, i.e.

$$\frac{\partial \boldsymbol{f}_n}{\partial t} + \begin{pmatrix} 0 & 0 & \mathrm{i}k_n v/\sqrt{3} \\ 0 & 0 & -\omega_z \\ \mathrm{i}k_n v/\sqrt{3} & \omega_z & 0 \end{pmatrix} \boldsymbol{f}_n = 0\,.$$

with the initial conditions $f_{n,0}(0) = C_n$ and $f_{n,1}(0) = f_{n,2}(0) = 0$. To solve it, we diagonalise it: The eigenvalues of the above matrix are

$$\lambda_0 = 0 \quad \text{and} \quad \lambda_{1,2} = \pm \mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\,, \tag{6.35}$$

where we defined $\zeta_n \coloneqq k_n v/\sqrt{3}$ and its eigenvectors are

$$\boldsymbol{V} = \begin{pmatrix} \mathrm{i}\omega_z/\zeta_n & \mathrm{i}\zeta_n/\sqrt{\omega_z^2+\zeta_n^2} & -\mathrm{i}\zeta_n/\sqrt{\omega_z^2+\zeta_n^2} \\ 1 & -\omega_z/\sqrt{\omega_z^2+\zeta_n^2} & \omega_z/\sqrt{\omega_z^2+\zeta_n^2} \\ 0 & -\mathrm{i} & -\mathrm{i} \end{pmatrix}. \tag{6.36}$$

The solution of the diagonalised system of ODEs thus is

$$\hat{\boldsymbol{f}}_n(t) \coloneqq \boldsymbol{V}^{-1}\boldsymbol{f}_n(t) = \begin{pmatrix} D_1 \\ D_2/2 \exp\left(\mathrm{i}\sqrt{\omega_z^2+\zeta_n^2}t\right) \\ D_3/2 \exp\left(-\mathrm{i}\sqrt{\omega_z^2+\zeta_n^2}t\right) \end{pmatrix}, \tag{6.37}$$

where D_1, D_2 and D_3 are complex constants of integration and, having in mind a later conversion to real quantities, we separated out the factor 1/2. We multiply $\hat{\boldsymbol{f}}_n$ with $\boldsymbol{V}$ to get the original coefficients $\boldsymbol{f}_n$, i.e.

$$
\begin{aligned}
f_{n,0}(t) &= \frac{\mathrm{i}\omega_z}{\zeta_n} D_1 + \frac{\mathrm{i}\zeta_n}{\sqrt{\omega_z^2 + \zeta_n^2}} \left[\frac{D_2}{2} \exp\left(\mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) - \frac{D_3}{2} \exp\left(-\mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) \right] \\
f_{n,1}(t) &= D_1 - \frac{\omega_z}{\sqrt{\omega_z^2 + \zeta_n^2}} \left[\frac{D_2}{2} \exp\left(\mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) - \frac{D_3}{2} \exp\left(-\mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) \right] \\
f_{n,2}(t) &= -\mathrm{i} \left[\frac{D_2}{2} \exp\left(\mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) + \frac{D_3}{2} \exp\left(-\mathrm{i}\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) \right] .
\end{aligned} \tag{6.38}
$$

Combing the above equations with the initial conditions for $\boldsymbol{f}_n$, we get a linear system of equations for D_1, D_2 and D_3 whose solution is

$$
D_1 = -\mathrm{i} \frac{\zeta_n}{\omega_z} \frac{C_n}{1 + (\zeta_n/\omega_z)^2} \quad \text{and} \quad D_{2,3} = \mp \mathrm{i} \frac{\zeta_n}{\sqrt{\omega_z^2 + \zeta_n^2}} C_n \, .
$$

In a last step, we use the obtained solution $\boldsymbol{f}_n$ in our ansatz for $\boldsymbol{f}$, see eq. (6.34), together with the fact that the expansion coefficients f_{000}, f_{110} and f_{100} are real. The latter implies for the initial condition of f_{000} that $C_n = C_{-n}^*$. If we define $C_n \coloneqq A_n/2 - \mathrm{i}B_n/2$ for $n \geq 1$ and $C_0 \coloneqq A_0$, the solution to eq. (6.33) is

$$
\begin{aligned}
f_{000}(t,x) &= \sum_{n=0}^{\infty} \frac{1}{\omega_z^2 + \zeta_n^2} \left[\omega_z^2 + \zeta_n^2 \cos\left(\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) \right] \left[A_n \cos(k_n x) + B_n \sin(k_n x) \right] \\
f_{110}(t,x) &= \sum_{n=1}^{\infty} \frac{\omega_z \zeta_n}{\omega_z^2 + \zeta_n^2} \left[\cos\left(\sqrt{\omega_z^2 + \zeta_n^2}\, t \right) - 1 \right] \left[A_n \sin(k_n x) - B_n \cos(k_n x) \right] \\
f_{100}(t,x) &= \sum_{n=1}^{\infty} \frac{\zeta_n}{\sqrt{\omega_z^2 + \zeta_n^2}} \left[A_n \sin(k_n x) - B_n \cos(k_n x) \right] \\
f_{111}(t,x) &= 0 \, .
\end{aligned} \tag{6.39}
$$

Notice that $\zeta_0 = 0$.

SAPPHIRE++ SETUP We emphasise that the solution presented in eq. (6.39) is the mathematical solution to the system of equations (6.33), and not the physical solution that one could in principle determine, for example via Liouville's theorem. This allows for a direct comparison with the numerical solutions that we compute with `Sapphire++`, including the time-evolution, spatial advection and rotation terms. To match the analytic solution, we choose the dimension of the configuration space to equal one, and we set the expansion order to $l_{\max} = 1$. For the numerical value of the $\boldsymbol{B}$-field we prescribe $\boldsymbol{B}^* = 2\pi \boldsymbol{e}_z$. Here, the asterisk means that the quantities are given in the units described in Tab. 6.1. We fix the

energy of the particles to $\gamma = 2$, implying that $\omega_z^* = q^* B_z^* / (\gamma m^*) = \pi$. To ensure positivity of f_{000}, we consider for the initial condition $A_0 = 2$ and $B_1 = 1$, all other A_n, B_n being set to zero. The size of the box is $L^* = 20$.

Since this is a one dimensional example, the computational grid (or mesh) is a line in all computed cases. The cells have the size $h = \Delta x^* = L^*/N_{\text{cells}}$. Simulations are run until $t_F^* = 10/\omega_z^*$.

RESULTS In this example we are interested in quantitatively comparing the numerical solution to the analytical solution. We restrict ourselves to the isotropic part f_{000} of the distribution function f and, thus, introduce the L_2 norm of the error of f_{000}, i.e.

$$\left\| f_{000,h}^n - f_{000}(t^n) \right\|_{L^2} \coloneqq \left(\int_{L^*} \left| f_{000,h}^n(x^*) - f_{000}(x^*, t^n) \right|^2 \mathrm{d}x^* \right)^{1/2}, \tag{6.40}$$

where $f_{000,h}^n(x^*)$ is the numerical approximation of the solution at time step t^n. This means that we do *not* integrate the error over time. Instead, we introduce the maximum relative error,

$$\text{max rel. } L^2 \text{ error} \coloneqq \max_{t^n} \frac{\left\| f_{000,h}^n - f_{000}(t^n) \right\|_{L^2}}{\left\| f_{000,h}^n \right\|_{L^2}}, \tag{6.41}$$

where the maximum is that of all time steps.

The plot at the top of Fig. 6.2 demonstrates how the error changes when we reduce the time step Δt^* for a fixed spatial resolution Δx^*, comparing the different time-stepping methods implemented in `Sapphire++`. Each data point corresponds to one simulation run. In all these simulations, we use a high spatial resolution of $\Delta x^* = L^*/64 = 0.3125$ with polynomial degree $k = 5$. This ensures that the numerical error of the time-stepping methods is larger than the spatial discretisation error.

The explicit fourth order Runge Kutta (ERK4) and forward Euler (FE) methods are only shown for time steps respecting the following CFL condition (see for example Cockburn, 1998, Sec. 2.3.3 and Di Pietro & Ern, 2012, Sec. 3.1.4):

$$\Delta t^* \leq t_{\text{CFL}} \approx \frac{1}{2k+1} \frac{\Delta x^*}{\beta_{\max}^*}. \tag{6.42}$$

$\beta_{\max}^* = U^* + \lambda_{\max} v^*$ is the maximum velocity of the spatial advection term, with $\lambda_{\max}$ the maximum eigenvalue of the $\boldsymbol{A}^x$ matrix. In this example $\beta_{\max}^* = v^*/\sqrt{3}$.

The error associated to the ERK4 and FE methods scale as Δt^4 and Δt respectively, as expected. For ERK4 the spatial error dominates when $\Delta t^* \approx 10^{-3}$, and the error plateaus. The error of the Crank–Nicolson (CN) method scales as Δt^2, while the implicit backward Euler (BE) method is only first order accurate in Δt

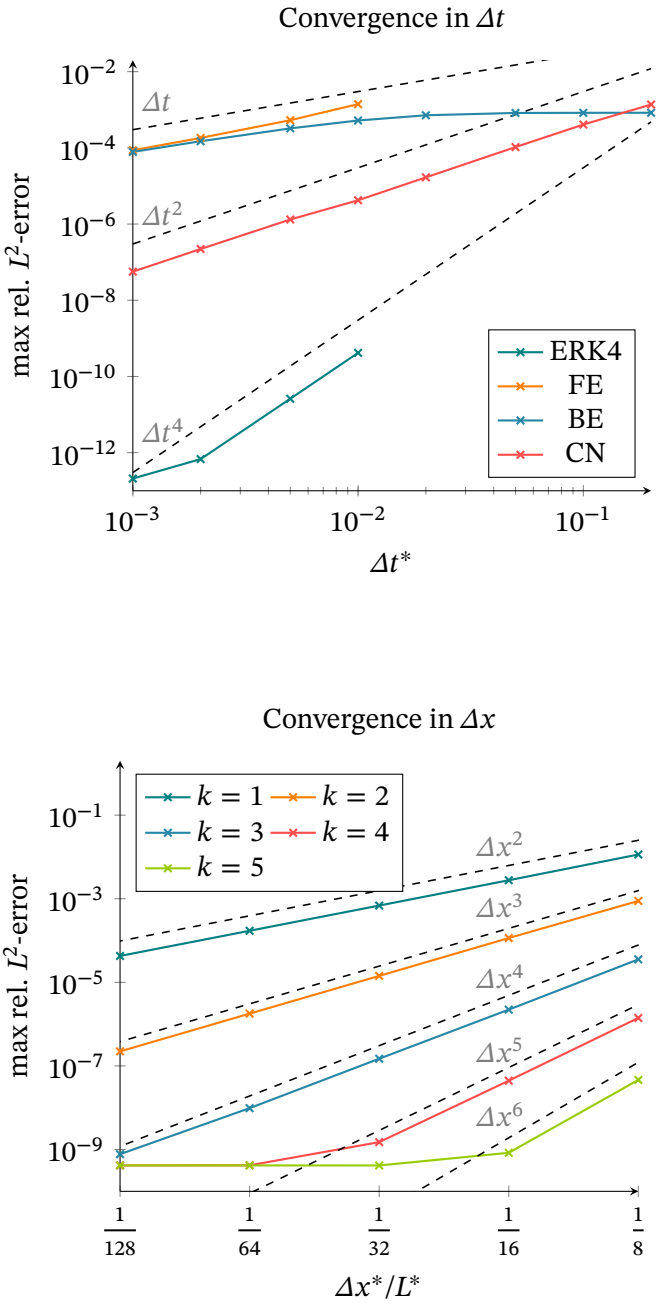

Figure 6.2: Top: Convergence in Δt^* for different time-stepping methods with $\Delta x^* = L^*/64$ and $k = 5$. The CFL condition is violated for $\Delta t^* > t_{\text{CFL}} \approx 10^{-2}$. Bottom: Convergence in Δx^* for different polynomial degrees k with ERK4 and $\Delta t^* = 10^{-2}$.

for small time steps, though the error saturates for large Δt due to the boundedness of the analytical solution: Large errors of the time stepping method cause the numerical solution and the analytical solution to be represented by cosines of different frequencies. Thus the error is bounded by the amplitude of the cosines.

The plot at the bottom of Fig. 6.2 shows the convergence with respect to the spatial resolution Δx^* for different polynomial degrees k. For these simulations we used the ERK4 method with a fixed time step $\Delta t^* = 10^{-2}$, respecting the CFL condition and ensuring that the time stepping error is subdominant (max rel. L^2 error $\sim 10^{-9}$). The error is therefore dominated by the spatial discretisation error up to very high spatial resolution. As expected this error scales as Δx^{k+1} (Cockburn, 1998, Sec. 2.2.4).[9]

6.2.2 ADVECTION IN A CONSTANT MAGNETIC FIELD

As we represent the single particle distribution function f with a series of spherical harmonics, a truncation of this series can result in a discrepancy between the physical solution and the numerical solution computed with `Sapphire++`. In this example we once more consider a test case for which the physical solution is known precisely and demonstrate how truncating the expansion at a finite $l_{\max}$ affects the solution.

DESCRIPTION The test scenario studied in this paragraph is very similar to the one of the previous example, i.e. we compute the distribution of charged and monoenergetic particles moving in a background plasma with a constant magnetic field $\boldsymbol{B} = B_0 \boldsymbol{e}_z$ in which the particles are not scattered ($\nu = 0$). We consider two cases, one in which the background plasma is static ($\boldsymbol{U} = 0$) and another one in which it is moving at a constant velocity $\boldsymbol{U} = U_0 \left(\boldsymbol{e}_x + \boldsymbol{e}_y\right)$. This example is 2D; $\boldsymbol{\xi} = (x, y)^\mathsf{T}$. The corresponding system of equations is

$$\frac{\partial \boldsymbol{f}}{\partial t} + (U_0 \mathbf{1} + v\boldsymbol{A}^x) \frac{\partial \boldsymbol{f}}{\partial x} + (U_0 \mathbf{1} + v\boldsymbol{A}^y) \frac{\partial \boldsymbol{f}}{\partial y} - \omega_z \boldsymbol{\Omega}^z \boldsymbol{f} = 0\,. \tag{6.43}$$

It is possible to obtain the physical solution with the help of Liouville's theorem, i.e. $\mathrm{d}f/\mathrm{d}t = 0$, which holds because we neglect scattering. It implies that the distribution function is constant along the trajectories of the particles, i.e. $f(\boldsymbol{x}(t), \boldsymbol{p}(t)) = f(\boldsymbol{x}_0, \boldsymbol{p}_0)$ where $\boldsymbol{x}_0$ and $\boldsymbol{p}_0$ are the particles' initial positions and momenta. If the phase-space trajectory is $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{\chi}(t)$ and $\boldsymbol{p} = \boldsymbol{p}_0 + \boldsymbol{\pi}(t)$, then $f(t, \boldsymbol{x}, \boldsymbol{p}) = f_0(\boldsymbol{x} - \boldsymbol{\chi}(t), \boldsymbol{p} - \boldsymbol{\pi}(t))$ where f_0 is the initial condition for f. Because in this example the particle trajectories are known, namely they gyrate in the x–y plane about the constant magnetic field, we can compute the physical

[9] If the initial condition would not belong to the Sobolev space H^{k+2} but only to H^{k+1}, the error would scale as $\Delta x^{k+1/2}$.

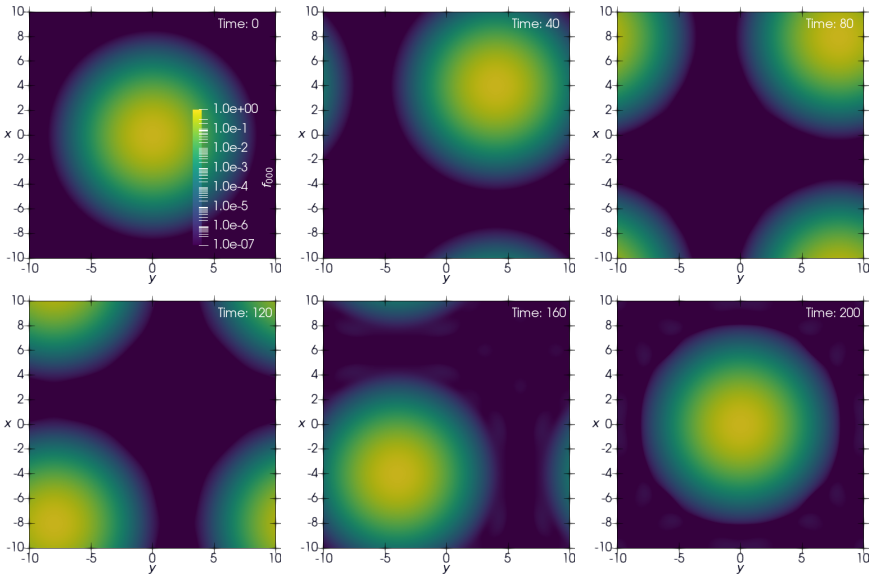

Figure 6.3: Time series showing the advection of the isotropic part of the distribution function f_{000}. We use $l_{\max} = 3$, $\Delta x^* = 20/64$, $k = 3$ and a ERK4 time stepping method with $\Delta t^* = 0.02$.

solution if we specify initial conditions f_0. As we use isotropic initial conditions, f_0 does not depend on $\boldsymbol{p}_0$ and with[10]

$$\boldsymbol{\chi}(t) = \frac{v}{\omega_z} \begin{pmatrix} \sin(\omega_z t + \theta) - \sin\theta \\ -\cos(\omega_z t + \theta) + \cos\theta \end{pmatrix}$$

the physical solution is $f(t, \boldsymbol{x}, \theta) = f_0(\boldsymbol{x} - \boldsymbol{\chi}(t))$.

Using the physical solution to determine the expansion coefficients f_{lms}, which we compute with `Sapphire++`, requires to numerically evaluate the integrals $(Y_{lms} \,|\, f)$ and we prefer to avoid this. We instead exploit the fact that the distribution function returns to its initial values after one gyroperiod $T_g = 2\pi/\omega_g = 2\pi\gamma m/q$. An initially isotropic particle distribution will again be isotropic after multiples of T_g and the isotropic part of the numerically computed distribution function, namely f_{000}, can then be compared to the physical solution. Notice that if the background plasma is not static, but moves with constant, uniform velocity, the actual and the numerically computed distribution function are also translated by a distance $|\boldsymbol{U}|T_g = \sqrt{2}U_0 T_g$ in the direction of $\boldsymbol{U}$.

SAPPHIRE++ SETUP In the computation of the numerical solution, we include the time-evolution term, the spatial advection term and the rotation term. As

[10]Here we use θ instead of φ to denote the gyro phase, because in `Sapphire++` x is the polar direction and in the current example $\boldsymbol{B}$ points out of the x–y plane into the z direction. Moreover, $v_\perp = v$, because we only simulate two dimensions.

the gyration and advection in this example are restricted to the x–y plane, we set the dimension of the configuration space to $d = 2$, i.e. $\boldsymbol{\xi} \in \mathbb{R}^2$. We truncate the expansion either at $l_{\max} = 3$ or $l_{\max} = 5$ to show how the numerical solution converges with increasing spectral resolution.

We choose, in numerical units, the following parameters $U_0^* = 0.1$, $B_0^* = 2\pi$ and the energy of the particle is set by $\gamma = 2$. As an initial condition we choose a Gaussian for the isotropic part of the distribution function:

$$f_{000}(t^* = 0, \boldsymbol{x}^*) = \exp\left(-\frac{x^{*2} + y^{*2}}{2\sigma^{*2}}\right). \tag{6.44}$$

All other expansion coefficients are set to zero. The standard deviation is $\sigma^* = 1.5$, equivalent to approximately five gyro radii.

The computational domain is a periodic box of size $L^* = 20$ that is uniformly refined such that $\Delta x^* = L^*/64$. We use polynomials of degree $k = 3$ in conjunction with a ERK4 time stepping method. The time step size is $\Delta t^* = 0.02$. We picked the spatial and temporal resolution such that the dominating error is produced by the cut-off in $l_{\max}$.

RESULTS In Fig. 6.3 we show the advecting isotropic part of the distribution function at different time steps for a moving background plasma. We note that if we use a periodic box of length $L^* = NU_0^* T_g^*$ with $N \in \mathbb{N}$ in conjunction with the prescribed constant velocity, the distribution function will return to its initial position after N gyroperiods. Note that we chose the parameters such that the particles described by the distribution function will gyrate $N = 100$ times before returning to their initial position at $t^* = 200$, see the lower right plot in Fig. 6.3.

In Fig. 6.4 the initial condition is compared with the result at $t^* = 200$, by plotting the residual

$$|f_h(\boldsymbol{x}^*, t^* = 200) - f_h(\boldsymbol{x}^*, t^* = 0)| \,. \tag{6.45}$$

Note that the residual is computed with the discrete representation of the initial condition, i.e. $f_h^0(\boldsymbol{x}^*)$, and not with its physical counter part $f(t^* = 0, \boldsymbol{x}^*)$. This removes the error that is introduced by the initial projection of the physical solution onto our finite element space from our comparison. For the case of the advecting isotropic distribution, we only plot the results for $l_{\max} = 3$, whereas in the static case we compare $l_{\max} = 3$ with $l_{\max} = 5$.

The latter shows that truncating the expansion of the distribution function at $l_{\max} = 3$ results in a solution that deviates slightly from the expected result, i.e. matching the initial condition. The deviation is greatly reduced for larger $l_{\max}$. The characteristic ring pattern is also a consequence of the truncation. Physically, the only frequency in the example is the gyrofrequency. However, a truncation at a finite $l_{\max}$ introduces more frequencies, cf. the factor ζ_n of the analytical solutions of the previous example summarised in eq. (6.39). The difference in frequency leads to the interference pattern that is shown. The fact

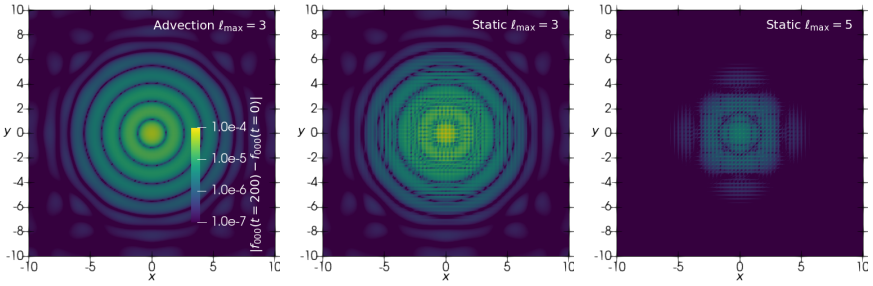

Figure 6.4: Residual for the different cases. Left: advection ($U_0^* = 0.1$) with $l_{\max} = 3$, Middle: static ($U_0^* = 0$) with $l_{\max} = 3$, Right: static ($U_0^* = 0$) with $l_{\max} = 5$. Spatial resolution and time step are the same as in Fig. 6.3.

that we see rings is due to the axial symmetry of the (numerical and analytical) solution.

6.2.3 DIFFUSIVE SHOCK ACCELERATION AT A PARALLEL SHOCK

In this example we use `Sapphire++` to simulate the time dependent acceleration of charged particles at a parallel shock front. We will compare the numerical results with the steady-state solution that we derived in Sec. 3.3.1 and with an approximate time-dependent solution that we took from Drury (1991) and Forman and Drury (1983).

DESCRIPTION As done in Sec. 3.3, we presuppose an infinitely planar shock with both the magnetic field $\boldsymbol{B}$ and the velocity field of the background plasma $\boldsymbol{U}$ aligned with the shock normal $\boldsymbol{n}$. The $\boldsymbol{B}$-field is continuous whereas the velocity jumps at the the shock, see eq. (3.61). The coordinate system is chosen such that the x-axis is parallel to the shock normal, see Fig. 3.1.2. Using these assumptions in combination with the diffusion approximation, cf. eq. (3.63), in the semi-relativistic VFP equation leads to the cosmic-ray transport equation (3.71) for a parallel shock.

In Sec. 3.3.1 we derived the steady-state solution for the isotropic part and the dipole anisotropy using a mixed-coordinate frame leaving the p-dependence unspecified, see eq. (3.77). We, subsequently, transformed the distribution function into the rest frame of the shock wave and determined the p-dependence, namely $A(p)$, for a monoenergetic and isotropic injection of particles at the shock front, see eq. (3.86). Transforming back into the mixed-coordinate frame, which

we use in `Sapphire++`, yields

$$\begin{aligned} f_1(x,p_1) &= f_1^0(x,p_1) + \cos\theta_1 f_{1,1}^1(x,p_1) \\ &= A(p_1)\left(1 - 3\cos\theta_1 \frac{U_1}{v_1}\right) \exp\left(\int_0^x \frac{U_1}{\kappa_1(x,p_1)}\,\mathrm{d}x\right) \qquad \text{for } x < 0 \\ f_2(x,p_2) &= f_2^0(x,p_2) = A(p_2) \qquad \text{for } x \geq 0\,, \end{aligned} \tag{6.46}$$

where

$$A(p_i) = \frac{Q}{U_1 p_{\mathrm{inj}}} \frac{3r}{r-1} \left(\frac{p_i}{p_{\mathrm{inj}}}\right)^{-3r/(r-1)}\,. \tag{6.47}$$

This is the distribution function with which we will compare the numerically computed one.

Up to now, we concentrated on the steady-state solution. We are also interested in investigating the temporal evolution of the distribution function and how it compares to the numerical results. Drury (1991, Sec. 3) presented an *approximate* analytic expression for the time-dependent spectrum at the shock, namely

$$f^0(t, x=0, p_1) = A(p_1)\phi(t) = A(p_1)\int_0^t \zeta(t')\,\mathrm{d}t'\,, \tag{6.48}$$

where, as above, f^0 is the isotropic part of the distribution and $\zeta(t)$ can be interpreted as the acceleration time distribution at $x = 0$ for the acceleration of particles from the injection momentum p_{inj} to momentum p_1, i.e. $\zeta(t)\,\mathrm{d}t$ is the probability that it takes the time t to accelerate a particle to p_1.

For the case of a scattering frequency that is independent of p and the same in the up- and downstream of the shock, i.e.

$$\kappa_1 = \kappa_2 = \lambda v/3 = v^2/3\nu = p^2/(3m^2\gamma^2\nu)\,,$$

the probability density function ζ and its cumulative distribution function ϕ are

$$\begin{aligned} \zeta(t) &= \frac{1}{\sqrt{2\pi c_2}} \left(\frac{t}{c_1}\right)^{-3/2} \exp\left(\frac{-c_1(t-c_1)^2}{2tc_2}\right) \quad \text{and} \\ \phi(t) &= \frac{1}{2}\left[\exp\left(\frac{2c_1^2}{c_2}\right)\operatorname{erfc}\left(\sqrt{\frac{c_1^3}{2tc_2}} + \sqrt{\frac{c_1 t}{2c_2}}\right) + \operatorname{erfc}\left(\sqrt{\frac{c_1^3}{2tc_2}} - \sqrt{\frac{c_1 t}{2c_2}}\right)\right], \end{aligned} \tag{6.49}$$

where c_1 and c_2 are the first two cumulants of the acceleration time distribution. The cumulant c_1 is the mean acceleration time, which is

$$\begin{aligned} c_1 &\coloneqq t_{\mathrm{acc}} \coloneqq \langle t \rangle \\ &= \frac{3}{U_1 - U_2} \int_{p_{\mathrm{inj}}}^{p_1} \left(\frac{\kappa_1}{U_1} + \frac{\kappa_2}{U_2}\right) \frac{\mathrm{d}p}{p} = \frac{r}{2U_1^2\nu} \frac{r+1}{r-1} \ln\left(\frac{1+p_1^2}{1+p_{\mathrm{inj}}^2}\right), \end{aligned} \tag{6.50}$$

see Drury (1983, eq. 3.31). The second cumulant c_2, which is the variance of the acceleration time, is given by (see Drury, 1983, eq. (3.32)):

$$\begin{aligned} c_2 \coloneqq \sigma_{\mathrm{acc}}^2 &= \left(\langle t^2\rangle - \langle t\rangle^2\right) \\ &= \frac{6}{U_1 - U_2}\int_{p_{\mathrm{inj}}}^{p_1}\left(\frac{\kappa_1^2}{U_1^3} + \frac{\kappa_2^2}{U_2^3}\right)\frac{\mathrm{d}p}{p} \\ &= \frac{1}{3\nu^2}\frac{r}{U_1^4}\frac{r^3+1}{r-1}\left[\frac{1}{1+p_1^2} - \frac{1}{1+p_{\mathrm{inj}}^2} + \ln\left(\frac{1+p_1^2}{1+p_{\mathrm{inj}}^2}\right)\right]. \end{aligned} \tag{6.51}$$

We note that the analytic expression for the temporal evolution of the spectrum given in the eqs. (6.48) and (6.49) is exact, if the diffusion coefficients are momentum independent and satisfy $\kappa_1/U_1^2 = \kappa_2/U_2^2$ (Toptyghin, 1980). However, in Forman and Drury (1983, Sec. 4) it was pointed out that $\zeta(t)$ could be used as an approximation to a general acceleration time distribution, i.e. for arbitrary diffusion coefficients, if the mean acceleration time and its variance are computed using the formulas presented in the eqs. (6.50) and (6.51). Additionally, it is required that $\zeta(t)$ is normalised to unity[11] (Drury, 1991, Sec. 3). Since our diffusion coefficient κ does depend weakly on p at low momenta and is the same in the up- and downstream of the shock wave, we expect the time-dependent spectrum in eq. (6.48) to merely approximate the true time dependence.

SAPPHIRE++ SETUP We note that the setup in this example does not match exactly the equations used to derive the analytical solution given in eq. (6.46) and (6.47); for pragmatic reasons, the shock is modelled as a narrow transition of finite thickness, represented by a tanh profile for the velocity $\boldsymbol{U}$, and the point injection of the particles at the shock is approximated with a Gaussian.

In the simulation all terms are included, i.e the time-evolution, the spatial advection, the momentum, rotation and collision terms, though the rotation term is not expected to contribute to the solution. The dimension of the configuration space is set to $d = 1$. Since the momentum terms are included, the reduced phase space is $\boldsymbol{\xi} = (x, p)^{\mathsf{T}}$. We truncate the spherical harmonic expansion at $l_{\mathrm{max}} = 1$, because in the derivation of the analytical solution we used the diffusion approximation and the diffusion approximation is equivalent to such a truncation. The resulting system of PDEs consists of four equations for the expansion coefficients $f_{000}, f_{100}, f_{110}$ and f_{111}. Since we restrict the simulation to one spatial dimension and choose $\boldsymbol{B}$ and $\boldsymbol{U}$ to be aligned with the x-axis, the equations for f_{110} and f_{111} decouple, see eqs. (3.67)–(3.69).[12]. If the coefficients are initially zero and the particles are injected homogeneously in the y–z plane,

[11] We numerically integrated $\zeta(t)$ using the cumulants c_1 and c_2 as given in the text and found that its normalisation is correct within the errors of the used integration method.

[12] Note that $f_{110} \propto f_2^1$ and $f_{111} \propto f_3^1$

they remain so. Thus the only equations containing non-zero terms are

$$\frac{\partial \boldsymbol{f}}{\partial t} + \begin{pmatrix} U & v/\sqrt{3} \\ v/\sqrt{3} & U \end{pmatrix} \frac{\partial \boldsymbol{f}}{\partial x} - \frac{\partial U}{\partial x} \begin{pmatrix} p/3 & \gamma m/\sqrt{3} \\ \gamma m/\sqrt{3} & 3p/5 \end{pmatrix} \frac{\partial \boldsymbol{f}}{\partial p} - \frac{\partial U}{\partial x} \begin{pmatrix} 0 & 2/\sqrt{3}v \\ 0 & 2/5 \end{pmatrix} \boldsymbol{f} + \begin{pmatrix} 0 & 0 \\ 0 & \nu \end{pmatrix} \boldsymbol{f} = \begin{pmatrix} s_{000} \\ 0 \end{pmatrix}, \tag{6.52}$$

where $\boldsymbol{f} = (f_{000}\ f_{100})^{\mathsf{T}}$. We included a source term on the right-hand side to represent the injection of the particles (assumed to be isotropic). We note all quantities in the above equation are dimensionless, see definitions in Tab. 6.1.

The relation between the isotropic part f^0 and anisotropic part f_1^1 of the analytical solution, see eq. (6.46), and the expansion coefficients f_{000} and f_{100} is determined by the fact that

$$\begin{aligned} f(t, x, p, \theta) &= f_{000}(t, x, p) Y_{000} + f_{100}(t, x, p) Y_{100}(\theta) \\ &= f^0(t, x, p) + f_1^1(t, x, p) \cos\theta \,, \end{aligned} \tag{6.53}$$

which implies that

$$f^0 = \frac{1}{\sqrt{4\pi}} f_{000} \quad \text{and} \quad f_1^1 = \sqrt{\frac{3}{4\pi}} f_{100} \,. \tag{6.54}$$

Notice that we dropped the distinction between up- and downstream, i.e. we removed the subscript i. The reason is that the p-variable in `Sapphire++` is defined in the mixed-coordinate frame, and, thus the distinction is implicit. Practically, if $x < 0$ it is p_1 and if $x > 0$ it is p_2.

Since the change in p, i.e. in energy, comes from the derivative in the velocity field $\boldsymbol{U}$, we cannot use the discontinuous velocity profile described in eq. (3.61). Instead we approximate it with

$$U^*(x^*) = \frac{U_1^*}{2r} \left[1 + r + (1 - r) \tanh(x^*/L_s^*)\right] , \tag{6.55}$$

where L_s^* is the shock width[13]. The shock parameters are chosen such that they plausibly model a supernova remnant shock. A typical speed for such a shock is a few thousand kilometres per second, e.g. $U_1^* = 1/60$. Generally, it is assumed that these shocks are strong, i.e. their compression ratio is $r = 4$. The shock width is chosen to be a fraction of the scattering mean free path, i.e. $L_s^* = 1/25$. The velocity profile is plotted in Fig. 6.5.

[13]For a discussion on the effect of a finite shock-thickness, see Achterberg and Schure (2011) and Drury et al. (1982). The power-law index is modified to

$$-\frac{3r}{r-1} - \frac{9}{2(r-1)} \frac{U_1}{v_1} \frac{L_s}{\lambda} \,.$$

For the results shown, the correction is 10^{-3}.

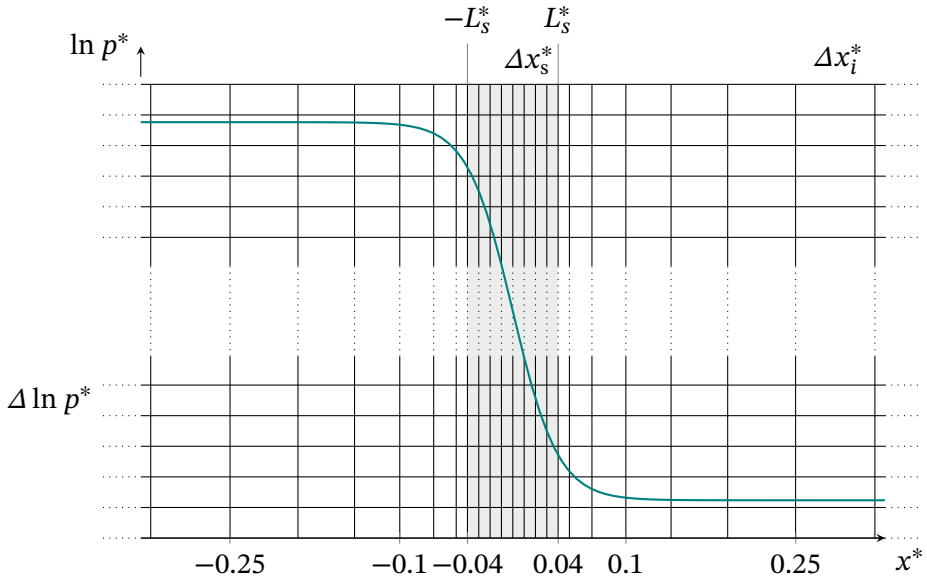

Figure 6.5: A detail from the computational grid of the diffusive shock acceleration simulation. It shows the design of the grid around the shock (highlighted in a light grey) and how it resolves the velocity profile $U(x)$ (drawn in teal).

We set the scattering frequency to $\nu^* = 1$. As expected on physical grounds, the $\boldsymbol{B}$-field does not appear in the system of PDEs, see eq. (6.52). Nonetheless, it is included in the simulation and set to $\boldsymbol{B}^* = B_0^* \boldsymbol{e}_x$, where $B_0^* = 1$. This implies that the gyro frequency is $\omega_g^* = 1/\gamma$ and the ratio $\nu^*/\omega_g^* = \gamma$, which means that a low energy particle ($\gamma \approx 1$) is scattered once per gyration about the magnetic field and high energy particles are scattered about γ times. Typically astrophysical plasmas are assumed to be *magnetised*, i.e. $\nu^*/\omega_g^* \lesssim 1$, and, hence, a constant scattering frequency is not a realistic choice. However, it simplifies the form of the exponential term in the analytical solution for the upstream distribution function, see eq. (6.46), and reduces simulation times.

We use a Gaussian distribution of particles to model the monoenergetic source, i.e.

$$s(x^*, p^*) = \frac{Q^*}{2\pi\sigma_p^*\sigma_x^*} \exp\left(-\frac{(x^* - x_{\text{inj}}^*)^2}{2\sigma_x^{*2}}\right) \exp\left(-\frac{(p^* - p_{\text{inj}}^*)^2}{2\sigma_p^{*2}}\right). \tag{6.56}$$

The spherical harmonic expansion of the source is $s(x^*, p^*) = s_{000}(x^*, p^*)Y_{000}$ with $s_{000} = \sqrt{4\pi}s(x^*, p^*)$. Particles are injected directly at the shock, i.e. $x_{\text{inj}}^* = 0$. The rate is $Q^* = 0.1$ and the injection momentum is $p_{\text{inj}}^* = 2$. The standard deviations of the Gaussian distribution are $\sigma_x^* = \sigma_p^* = 1/8$.

The initial conditions for the expansion coefficients are $\boldsymbol{f}_0(t^* = 0) = 0$, i.e. initially there are no particles in the computational domain. The boundaries in x-direction are treated differently in the up- and downstream region. At the

upstream boundary we use the zero inflow boundary condition as described above and presented in eq. (6.26). At the downstream boundary, we expect that the gradient of the asymptotic solution is zero, i.e. $\partial f/\partial x = 0$. We thus allow the inflow to be determined by the values of the approximate solution $\boldsymbol{f}_h$ on the boundary, i.e.

$$\hat{\boldsymbol{J}}_F^B(\boldsymbol{f}_h) = \boldsymbol{W}\left(\boldsymbol{\Lambda}_+ \boldsymbol{W}^\mathsf{T} \boldsymbol{f}_h + \boldsymbol{\Lambda}_- \boldsymbol{W}^\mathsf{T} \boldsymbol{f}_h\right) . \tag{6.57}$$

We refer to this as the continuous boundary condition. The boundaries in p-direction fulfil the zero inflow boundary condition.

The computational domain is $D = [-280, 280] \times [\ln(0.1), \ln(100)]$. We require the spatial grid to cover multiple diffusion lengths, i.e. $L_d^* \coloneqq \kappa_1^*/U_1^* = \lambda_1^* v_1^*/(3U_1^*) \approx 20 \approx x_{\max}^*/14$. The dimension in the p-direction covers multiple orders of magnitude to show that `Sapphire++` produces an extended power law. We adapt the cell size in x-direction to resolve the shock region accurately, see Fig. 6.5. In the shock region (highlighted in grey) we choose a constant cell size $\Delta x_\mathrm{s}^* = 0.01$ whereas the cell size outside the shock region increases as $\Delta x_i^* = \sinh(i * 0.01)$. The coarse resolution in the outer parts of the domain allows us to simulate large upstream and downstream regions. The cell size in $\ln p$-direction is $\Delta \ln p^* = (\ln(100) - \ln(0.21))/256 \approx 0.027$.

For the time evolution we use the implicit Crank–Nicolson method with the time step $\Delta t^* = 1$. The simulation is run up to a final time of $t_F^* = 5 \times 10^5$ to capture the steady-state solution.[14]

An overview of the simulation parameters is collected in Tab. 6.2.

RESULTS In Fig. 6.6 we compare the steady-state analytic solution with the numerical solution computed with `Sapphire++`. In the panel at the top, the numerically computed spectrum at the shock is compared to the analytic expectation given in eq. (6.47), i.e. a power law with spectral index $\alpha = -3r/(r-1) = -4$ and normalisation $N \coloneqq Q^*/(p_\mathrm{inj}^* U_1^*) 3r/(r-1) = 12$. The spectral index of the numerical solution is $|\alpha_\mathrm{num} - 4| = 1.5 \times 10^{-3}$ and it extends up to $p^* = 100$, which is the boundary of the computational domain in the p-direction. The resolution of the ordinate of the log-log plot is too low to see that the normalisation of the numerical solution $N_\mathrm{num} = 12.14$ is off by $(N - N_\mathrm{num})/N \approx 0.012 = 1.2\%$. We speculate that this discrepancy is due to approximating a point injection of particles with a Gaussian distribution.

In the panel at the bottom, we compare the spatial profile of the numerical solutions with the analytic solution that we summarised in eq. (6.46) and evaluated at $p^* = 10$. The discrepancy between the computed isotropic part and the analytical result at the left boundary of the spatial domain is due to the boundary condition that enforces zero inflow. There is also a small difference in the downstream normalisation, which is not visible due to the log scaling of the

[14] We confirmed the results using an explicit fourth order Runge–Kutta (ERK4) method. But as it requires a much smaller time step, we terminated the simulation at an earlier time.

Table 6.2: Simulation parameters modelling a supernova remnant shock

Parameter	Value	Description
U_1^*	1/60	Velocity of the supernova remnant shock
r	4	Compression ratio of the shock
L_s^*	1/25	Width of the shock's velocity profile
ν^*	1	Scattering frequency
B_0^*	1	Strength of the magnetic field
Q^*	0.1	Injection rate at the shock
p_{inj}^*	2	Injection momentum of the particles
x_{inj}^*	0	Location of the injection
σ_x^*	1/8	Width of the source in x-direction
σ_p^*	1/8	Width of the source in p-direction
D	$[-280, 280] \times [\ln(0.1), \ln(100)]$	Computational domain
Δt^*	1	Time step size
t_F^*	5×10^5	Final time of the simulation

$f^0(x^*)$-axis. This is the same discrepancy as in the normalisation of the particle spectrum discussed in the previous paragraph.

In Fig. (6.7) we plot the temporal evolution of the numerically computed spectrum at the shock's position for a fixed momentum $p^* = 59.9$ and compare it with the *approximate* analytic expression given in the eqs. (6.48) and (6.49). Despite the fact that the setup used in `Sapphire++` only approximates the assumptions leading to the steady-state spectrum and its time-dependent counterpart and the fact that the analytic expression for the temporal evolution is also merely an approximation, the two curves follow each other closely. This indicates that the temporal evolution of the spectrum is captured accurately by `Sapphire++`.

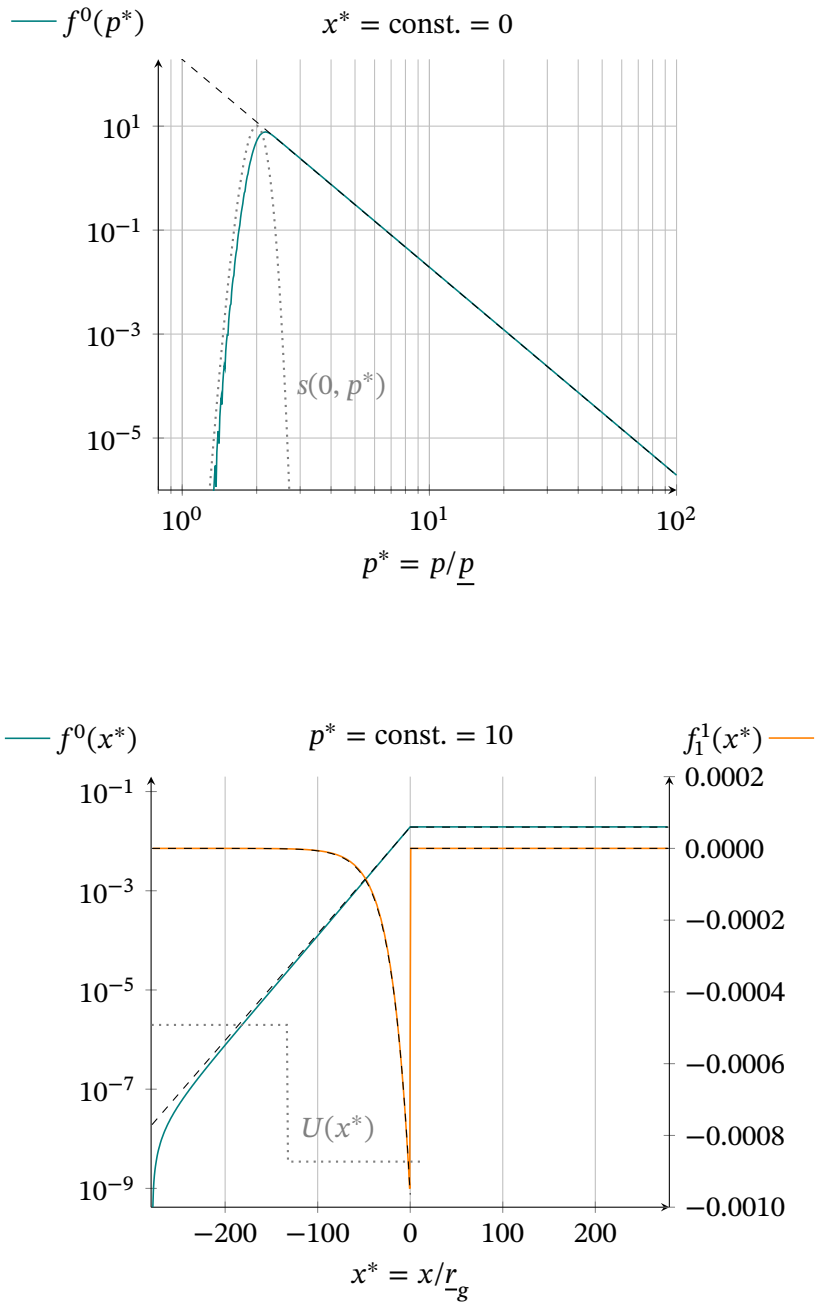

Figure 6.6: Comparison between analytical solution and numerical solution. Top panel: A log-log plot of the particle spectrum at the shock. Bottom panel: A plot of the isotropic and anisotropic part of the distribution function for a constant momentum.
Coloured plots present the numerical results and the dashed plots show the analytical solution.
The dotted plots show the width of the source term $s(0, p^*)$ (top panel) and the velocity profile $U^*(x^*)$ of the shock in its rest frame (bottom panel).

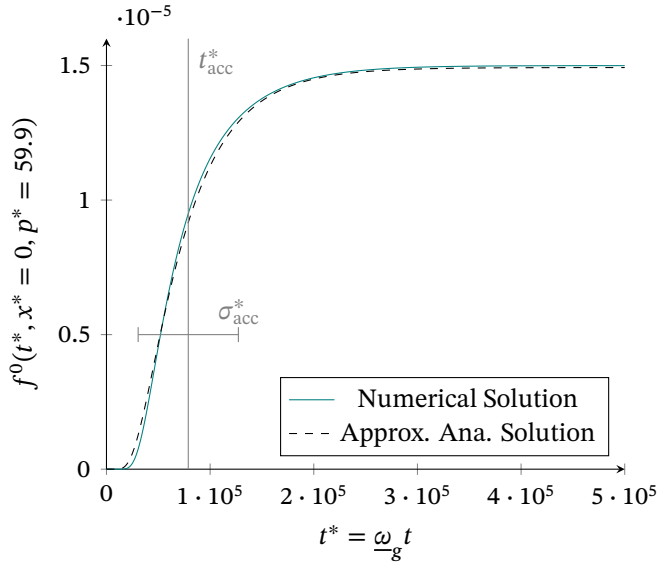

Figure 6.7: The plot shows the temporal evolution of the isotropic part of the particle spectrum f^0 at $x^* = 0$ and $p^* = 59.9$. The dashed curve shows the approximate analytic expression and the teal-coloured curve shows the numerical result.
The mean acceleration time is $t^*_{\text{acc}} = 78909.74$ and the standard deviation is $\sigma^*_{\text{acc}} = 48859.1$.

Chapter 7

Concluding remarks

In the introduction we explicitly stated a research question that guides our efforts, namely 'How does the non-linear interaction between the particles and the plasma, in which they are accelerated, influences the acceleration process and the maximum energy the particles can reach?'. We seek to answer this question quantitatively using a model that couples a kinetic description of the energetic and charged particles with an ideal MHD description of the background plasma. From the perspective of our research question, we haven taken the first step and developed a robust numerical algorithm that solves the VFP equation.

We pass the important milestones in review. The first two chapters of this thesis provided us with the basis of our research. In particular, we defined the one (or single) particle distribution function, see eq. (2.6), and showed that its evolution is determined by the Boltzmann equation (2.31). We pointed out that the collision term can be modelled with a Fokker–Planck approach as presented in eq. (2.62). We, moreover, derived the ideal MHD equations (2.93)–(2.96). To model the transport of charged and energetic particles that interact with a tenuous astrophysical plasma, we combined the kinetic description and the fluid description. An essential part of this combination is an explicit expression for the collision operator. We used the interaction of a charged particle with an Alfvén wave, supposed to be prototypical for the interactions of the particles with the background plasma, to derive the expression given in eq. (3.37). Its derivation assumed that the momentum coordinates were defined in the rest frame of the wave. This led us to the introduction of the mixed-coordinate system. We employed a Lorentz transformation to formulate the (fully) relativistic VFP equation in this coordinate system, see eqs. (3.44) and (3.56). The assumption that the background plasma moves with velocities that are small in comparison to the speed of light, allowed us to drop terms of the order $\mathcal{O}(U/c)$ in the relativistic VFP equation. The result was the semi-relativistic VFP equation (3.60), the starting point of our research.

The key concept of our approach towards a solution of the semi-relativistic VFP equation is the spherical harmonic expansion of the single particle distribution function. It mitigates computational costs, because it separates out the angular dependence in momentum space and, thus, reduces the number of independent variables from six to four. The spherical harmonics are a basis of

the space of the square integrable functions that are defined on the unit sphere. This space is a Hilbert space and it is the theory of Hilbert spaces setting the mathematical framework of our research. In Chapter 4 we derived formulae to convert between the coefficients of the spherical harmonic and, the equivalent, Cartesian tensor expansion. The formulae are based on the insight that both expansions represent the same function, though using a different basis in the space of spherical harmonics. They, thus, take the form of the basis transformation stated in eq. (4.42). To derive the basis transformation matrix we took advantage of the fact that the relation between the coefficients of the (Cartesian) and spherical multipole expansion of the electrostatic (or gravitational) potential is the same as the relation between the expansion coefficients representing the distribution function. This implies that the formulae can also be applied to convert between spherical and Cartesian multipole moments, broadening their scope considerably. A highlight of our derivation is the introduction of a new definition of the (Cartesian) multipole moments that is based on the Kelvin transform, cf. eq. (4.7).

Expressing the single particle distribution function as a series of spherical harmonics comes at the cost of replacing a scalar partial differential equation, namely the VFP equation, with a system of PDEs that determines the unknown expansion coefficients. In Chapter 5 we presented a new method to derive the system of PDEs. The method is based on operators that act in the Hilbert space of spherical harmonics and their representation matrices, see Sec. 5.1. It allowed us to show that the representation matrices, out of which the system of PDEs is built, are related to each other via rotations of the spherical harmonics, see eqs. (5.48), (5.49) and (5.51). This in turn implies that the representation matrices have the same eigenvalues and that their eigenvectors are rotated versions of each other, see eqs. (5.67) and (5.68). In Sec. 5.4.2 we showed that their eigenvalues are the roots of the associated Legendre polynomials. The fact that the spherical harmonics are complex whereas the single particle distribution function is real implies that system of PDEs contains linearly dependent equations. In Sec. 5.3 we demonstrated that a basis transformation from the complex to the real spherical harmonics removes the redundant equations.

In Chapter 6 we solved the system of PDEs numerically with the discontinuous Galerkin method. The perspective of the operator method enabled us to identify the system of PDEs as a system of advection-reaction equations, which are particularly suited for an application of the dG method, see the formulation in eq. (6.4). To the best of our knowledge, it is the first time that the dG method was used in conjunction with a spherical harmonic expansion. In Sec. 6.1 we therefore described in detail how it is applied to the system of advection-reaction equations. An essential part of the dG discretisation is the choice of the numerical flux. We decided to use an upwind flux, because it stabilises the method. This choice forces us to solve a Riemann problem at each cell interface, however, we exploit our knowledge about the eigenvalues and eigenvectors of the representation matrices to avoid this in the spatial directions of the computational

domain, cf. Sec. 6.1.4. In the p-direction we numerically compute the necessary eigenvalues and eigenvectors. We implemented the dG method using the finite element library `deal.ii` and called our implementation `Sapphire++`. We used a toy model to test the spatial and temporal accuracy of `Sapphire++`. It behaves as theoretically expected, see Fig. 6.2. In Sec. 6.2.2 we also investigated the effects of truncating the spherical harmonic expansion on the qualitative behaviour and the accuracy of the numerical solutions. Eventually, we simulated the acceleration of charged particles at a parallel shock and demonstrated that the result agrees with the analytical solutions that we presented in Sec. 3.3.1. This is shown in Fig. 6.6 and 6.7. We conclude that `Sapphire++` is ready to be used in astrophysical applications.

Coming back to our research question, it is clear that numerically solving the VFP equation is only the first part. Our future efforts will centre about the implementation of the ideal MHD equations to model the background plasma and finding ways to couple the two codes. This would allow us to study self-consistently how the feedback of the particles changes their acceleration. Though, we would like to mention that other research directions are possible, namely the inclusion of higher-order relativistic corrections for relativistically moving background plasmas, see e.g. Appendix C.2, or an extension of the dG method to inertial confinement fusion plasmas.

Appendix A

Cartesian tensors and spherical harmonics

A.1 Notes on the definition of the Cartesian multipole moments

A.1.1 The multipole moments are tensors

We would like to use our definition of Cartesian multipole moments, which is given in eq. (4.6), to show that they are tensors. In the physics literature a rank l-th tensor $T^{(l)}$ is defined as a quantity having 3^l components $T^{(l)}_{i_1\cdots i_l}$, which transform under coordinate transformations according to the following scheme (cf. Goldstein et al., 2014, p. 189):

$$\hat{T}^{(l)}_{i_1\cdots i_l} = (\boldsymbol{A})_{i_1}{}^{j_1}\cdots(\boldsymbol{A})_{i_l}{}^{j_l} T^{(l)}_{j_1\cdots j_l}\,.$$

Our definition of the Cartesian multipole moments implies this, i.e.

$$\begin{aligned}\hat{Q}^{(l)}_{i_1\cdots i_l} &= \int \hat{\rho}(\hat{\boldsymbol{x}})\hat{M}^{(l)}_{i_1\cdots i_l}(\hat{\boldsymbol{x}})\,\mathrm{d}^3\hat{x} = \int \hat{\rho}(\hat{\boldsymbol{x}})\hat{\mathcal{K}}\left[\hat{\partial}_{i_1}\cdots\hat{\partial}_{i_l}\frac{1}{|\hat{\boldsymbol{x}}|}\right]\mathrm{d}^3\hat{x}\\ &= (\boldsymbol{A})_{i_1}{}^{j_1}\cdots(\boldsymbol{A})_{i_l}{}^{j_l}\int \rho(\boldsymbol{x})\mathcal{K}\left[\partial_{j_1}\cdots\partial_{j_l}\frac{1}{|\boldsymbol{x}|}\right]\mathrm{d}^3x\\ &= (\boldsymbol{A})_{i_1}{}^{j_1}\cdots(\boldsymbol{A})_{i_l}{}^{j_l} Q^{(l)}_{j_1\cdots j_l}\,.\end{aligned}$$

Note that we dropped the primes in the definition of the Cartesian multipole moments to increase the readability.

A.1.2 The $\hat{P}$ operator

We defined the $\hat{P}$ operator implicitly using the description: "[It] produces the sum over the pairs of indices needed to assure symmetry". We now give a more explicit definition.

$\hat{P}$ acts on a product of components of $\boldsymbol{x}$ and Kronecker deltas, i.e.

$$\hat{P}(\delta_{i_1 i_2}\cdots\delta_{i_{2k-1}i_{2k}} x_{i_{2k+1}}\cdots x_{i_l})\,.$$

Note that there are k Kronecker deltas and $l - 2k$ components of $\boldsymbol{x}$.

The P operator produces a sum of products, which must not change if arbitrary indices are exchanged, i.e. which is *symmetric*. Moreover, note that in each product all indices are distinct (no index appears twice).

For $k = 0$, we define

$$\hat{P}(x_{i_1} \cdots x_{i_l}) \coloneqq x_{i_1} \cdots x_{i_l} .$$

Note that a product is commutative, hence this (trivial) sum is symmetric. For $k = 1$, one Kronecker delta is part of the products, and the sum created by $\hat{P}$ will be symmetric, if it contains all the terms (products) with Kronecker deltas whose two indices correspond to all possible pairs, which can be formed from the l indices. For example, if $l = 3$, then there are $\binom{3}{2} = 3$ pairs of indices and accordingly

$$\hat{P}(x_{i_1}\delta_{i_2 i_3}) = x_{i_1}\delta_{i_2 i_3} + x_{i_2}\delta_{i_1 i_3} + x_{i_3}\delta_{i_1 i_2} .$$

This guarantees the symmetry of the sum, because whenever an index of a Kronecker delta is exchanged with another index (may it be an index of a component of $\boldsymbol{x}$ or of another Kronecker delta), the newly created pair is matched by another term with a Kronecker delta whose pair of indices turns into the Kronecker delta's original combination of indices. Generally, if $k = 1$, the sum produced by $\hat{P}$ contains $\binom{l}{2}$ terms.

k Kronecker deltas need to be equipped with k pairs of indices. Since no index appears twice in a product, the pairs must be distinct, or more technically, disjoint. We can collect all distinct pairs in a set. The $\hat{P}$ operator must produce a sum over all *sets* of distinct pairs of indices to make it symmetric. For example, if $l = 6$ and $k = 2$, the product $\delta_{i_1 i_2}\delta_{i_3 i_4} x_{i_5} x_{i_6}$ will be part of the sum. The set of distinct pairs is $\{(i_1, i_2), (i_3, i_4)\}$. If you exchange an index of one Kronecker delta with an index of the other, then a different set of distinct pairs is created. And to make the sum symmetric, another term whose set of distinct pairs (pertaining to the Kronecker deltas) matches the newly created set must be included in the sum. Because a corresponding exchange of indices in the other term's set of pairs recreates the original one. Moreover, if you exchange an index of a Kronecker delta with an index of a component of $\boldsymbol{x}$, a new set of distinct pairs is created, which also requires its counterpart to keep the sum symmetric. Hence, $\hat{P}$ produces

$$\binom{l}{2k} \frac{\binom{2k}{2}\binom{2k-2}{2} \cdots \binom{2}{2}}{k!} = \binom{l}{2k} (2k-1)!! \tag{A.1}$$

terms. The first factor counts the number of possible combinations of indices, which are then available to form sets of distinct pairs from them. The second factor counts the number of possible sets of distinct pairs. One way to approach the second factor is to think of a tennis tournament. In its first round you have to think of whom plays against whom and this is tantamount to build sets of distinct pairs. If you like to know how many sets there are, you can do the

computation encoded in the second factor: choose two players from all the available players, then choose two players form the left over players and so on. Since the order of the formed pairs is irrelevant, you divide by $k!$ (note that there are k pairs).

Building the products in the sum, produced by the $\hat{P}$ operator, can be done systematically. Assume there are l indices.

If you choose from them $2k$ indices, the possible combinations can be constructed as follows: Begin with i_1, traverse the tree of possible combinations, which is determined by the other $l-1$ indices and the amount of chosen indices, namely $2k$. The pattern of this traversal is illustrated in the following example. In a next step choose i_2 and traverse the tree of possible combinations. But this time there are only $l-2$ indices left to form this tree; proceed until only one possibility is left. For example, $l = 6$ and $k = 2$, then

$i_1i_2i_3i_4$	$i_1i_2i_3i_5$	$i_1i_2i_3i_6$
$i_1i_2i_4i_5$	$i_1i_2i_4i_6$	$i_1i_2i_5i_6$
$i_1i_3i_4i_5$	$i_1i_3i_4i_6$	
$i_1i_3i_5i_6$		
$i_1i_4i_5i_6$		
$i_2i_3i_4i_5$	$i_2i_3i_4i_6$	
$i_2i_3i_5i_6$		
$i_2i_4i_5i_6$		
$i_3i_4i_5i_6$		

are the possible combinations of indices.

In a second step, each combination has to be used to build sets of distinct pairs. Take for example the first combination $i_1i_2i_3i_4$. The possible sets of distinct pairs are

$$\{(i_1,i_2)(i_3,i_4)\},\{(i_1,i_3),(i_2,i_4)\} \text{ and } \{(i_1,i_4)(i_3,i_2)\}\,.$$

These sets were constructed by swapping the index i_2 with i_3 and i_4. If it had not been four but six indices, the sets of distinct pairs would have been constructed by creating the above three sets with four indices and, subsequently, swapping one element of the left over pair with all the indices appearing in the three sets. This process can be continued as needed. The terms in the sum corresponding to the above sets of distinct pairs are

$$\delta_{i_1i_2}\delta_{i_3i_4}x_{i_5}x_{i_6} + \delta_{i_1i_3}\delta_{i_2i_4}x_{i_5}x_{i_6} + \delta_{i_1i_4}\delta_{i_2i_4}x_{i_5}x_{i_6}\,.$$

The $\hat{P}$ operator then sums over all sets of distinct pairs which are created for each of the above combination of indices.

A.1.3 $c_{l,k}$ COEFFICIENTS

The coefficients $c_{l,k}$ appearing in the explicit expression for the multipole functions given in eq. (4.11) can be computed. The multipole functions are harmonic

polynomials, i.e. they are solutions to Laplace's equation. This implies that

$$0 = \Delta M^{(l)}_{i_1 \cdots i_l} = \sum_{k=0}^{\left\lfloor \frac{l}{2} \right\rfloor} c_{l,k} \Delta \left(|\boldsymbol{x}|^{2k} R_{l,2k} \right) \tag{A.2}$$

The product rule tells us that

$$\Delta fg = \Delta f g + 2 \nabla f \cdot \nabla g + f \Delta g \,.$$

Hence, we calculate the following derivatives

$$\begin{aligned} \Delta \, |\boldsymbol{x}|^{2k} &= 2k(2k+1) \, |\boldsymbol{x}|^{2k-2} \\ \nabla \, |\boldsymbol{x}|^{2k} \cdot \nabla R_{l,2k} &= 2k(l-2k) |\boldsymbol{x}|^{2k-2} R_{l,2k} \\ \Delta R_{l,2k} &= (2k+2) R_{l,2k+2} \,. \end{aligned} \tag{A.3}$$

The first derivative is a straightforward computation. We explain the second and the third derivative. We begin with replacing $R_{l,2k}$ in the second derivative with its definition, i.e.

$$\partial_j \, |\boldsymbol{x}|^{2k} \, \partial^j \hat{P} (\delta_{i_1 i_2} \cdots \delta_{i_{2k-1} i_{2k}} x_{i_{2k+1}} \cdots x_{i_l}) \,.$$

And $\partial_j \, |\boldsymbol{x}|^{2k} = 2k \, |\boldsymbol{x}|^{2k-2} \, x^j$. The action of the derivative with respect to x_j on the terms in the sum produced by the operator $\hat{P}$ is most easily seen if an example term is considered, e.g.

$$\begin{aligned} &\delta_{i_1 i_2} \cdots \delta_{i_{2k-1} i_{2k}} \partial_j \left(x_{i_{2k+1}} \cdots x_{i_l} \right) \\ &\quad = \delta_{i_1 i_2} \cdots \delta_{i_{2k-1} i_{2k}} \delta_{j, i_{2k+1}} x_{i_{2k+2}} \cdots x_{i_l} + \cdots + \delta_{i_1 i_2} \cdots \delta_{i_{2k-1} i_{2k}} x_{i_{2k+1}} \cdots x_{i_{l-1}} \delta_{j, i_l} \,. \end{aligned}$$

$l - 2k$ terms appear. In every term one of the components of $\boldsymbol{x}$ is replaced with a Kronecker delta. Note that a contraction of the above sum with x^j yields $l - 2k$ times the original term. Whence,

$$\nabla \, |\boldsymbol{x}|^{2k} \cdot \nabla R_{l,2k} = 2k(l-2k) \, |\boldsymbol{x}|^{2k-2} \, R_{l,2k} \,.$$

To illuminate the reasoning behind the derivation of the third derivative, we again consider an example. Let $l = 6$ and $k = 0$, then (by definition) $\hat{P}$ creates only one term, namely $x_{i_1} \cdots x_{i_6}$. If our expression for the third derivative is correct, applying the Laplace operator to this term yields the sum, which is created by P if $k = 1$, times two. We can compute how many terms we expect this sum to have with the formula given eq. (A.1), namely $\binom{6}{2} = 15$. Computing $\Delta \left(x_{i_1} \cdots x_{i_6} \right)$ gives 30 terms. But, for example,

$$\partial_j \left(x_{i_1} \delta_{j i_2} x_{i_3} \cdots x_{i_6} \right) = \cdots + x_{i_1} \delta_{j i_2} x_{i_3} \cdots \delta_{j i_4} x_{i_5} x_{i_6} + \cdots$$

and

$$\partial_j \left(x_{i_1} \cdots \delta_{j i_4} x_{i_6} \right) = \cdots + x_{i_1} \delta_{j i_2} x_{i_3} \cdots \delta_{j i_4} x_{i_5} x_{i_6} + \cdots$$

yield the same term if summed over j. Hence, we get $15 * 2$ terms and this is what we expect.

If we apply the Laplace operator to $\hat{P}$ when $k = 1$, we expect to obtain the sum which is created by $\hat{P}$ when $k = 2$ times four. For $l = 6$ and $k = 2$, the $\hat{P}$ operator produces a sum with 45 terms. Applying Δ to $\hat{P}$ when $k = 1$, yields $15 * 4 * 3 = 180$ terms. Note, this is exactly four times the amount of terms in R_4. And, as the following example shows, the same term in ΔR_2 appears four times, i.e.

$$\begin{aligned}
\partial_j \left(\delta_{i_1 i_2} \delta_{j i_3} x_{i_4} x_{i_5} x_{i_6}\right) &= \cdots + \delta_{i_1 i_2} \delta_{j i_3} \delta_{j i_4} x_{i_5} x_{i_6} + \cdots \\
\partial_j \left(\delta_{i_1 i_2} x_{i_3} \delta_{j i_4} x_{i_5} x_{i_6}\right) &= \cdots + \delta_{i_1 i_2} \delta_{j i_3} \delta_{j i_4} x_{i_5} x_{i_6} + \cdots \\
\partial_j \left(\delta_{i_3 i_4} \delta_{j i_1} x_{i_2} x_{i_5} x_{i_6}\right) &= \cdots + \delta_{i_3 i_4} \delta_{j i_1} \delta_{j_{i_2}} x_{i_5} x_{i_6} + \cdots \\
\partial_j \left(\delta_{i_3 i_4} x_{i_1} \delta_{j i_2} x_{i_5} x_{i_6}\right) &= \cdots + \delta_{i_3 i_4} \delta_{j i_1} \delta_{j_{i_2}} x_{i_5} x_{i_6} + \cdots
\end{aligned}$$

Thus, we get the expected derivative.

This pattern can be generalised to arbitrary l and k. The number of terms created when the Laplace operator is applied to R_{2k} divided by the number of terms which is contained in the sum R_{2k+2} is

$$\frac{\binom{l}{2k}(2k-1)!!(l-2k)(l-2k-1)}{\binom{l}{2k+2}(2k+1)!!} = 2k+2\,.$$

As in the above examples, ΔR_{2k} includes $2k + 2$ times the same terms: Two of them because of the components of $\boldsymbol{x}$ (cf. with the first example) and two for each Kronecker delta (and there are k of them). Thus,

$$\Delta R_{l,2k} = (2k+2) R_{l,2k+2}\,.$$

The coefficients $c_{l,k}$ can be determined by plugging the three derivatives, given in eq. (A.3), into the equation presented at the beginning of this section, namely eq. (A.2). Collecting all factors in front of $|\boldsymbol{x}|^{2k-2} R_{2k}$, yields the following condition

$$2k c_{l,k-1} + (2k(2k+1) + 4k(l-2k)) c_{l,k} = 0\,.$$

This condition implies the recurrence relation

$$c_{l,k} = -\frac{c_{l,k-1}}{2l-(2k-1)}\,. \tag{A.4}$$

Eventually, setting $c_{l,0} \coloneqq (2l-1)!!$ turns it into a closed-form expression

$$c_{l,k} = (-1)^k \frac{(2l-1)!!}{(2l-1)(2l-3)\cdots(2l-(2k-1))} = (-1)^k (2l-(2k+1))!!\,. \tag{A.5}$$

A.2 COMMUTATOR OF THE LADDER OPERATORS

We would like to compute the commutator $[\mathrm{L}_+, \mathrm{D}_z]$. We use that $[\mathrm{L}_i, x^j] = \mathrm{i}\epsilon_{ijk}x^k$ and $[\mathrm{L}_i, \partial_j] = \mathrm{i}\epsilon_{ijk}\partial_k$ (cf. Landau & Lifshitz, 1977, p. 84, eq. 26.4-5) to compute

$$[\mathrm{L}_+, x^j] = \mathrm{i}\epsilon_{1jk}x^k - \epsilon_{2jk}x^k$$

and

$$[\mathrm{L}_+, \partial_j] = \mathrm{i}\epsilon_{1jk}\partial_k - \epsilon_{2jk}\partial_k\,.$$

Where ϵ_{ijk} is the *Levi-Civita symbol*, which is, for example, defined in Jeevanjee (2011, p. 4, eq. 1.1). Since the commutator is linear, we get

$$[\mathrm{L}_+, \mathrm{D}_z] = 2[\mathrm{L}_+, zx^m\partial_m] - [\mathrm{L}_+, |\boldsymbol{x}|^2\,\partial_z] + [\mathrm{L}_+, z]\,.$$

With the above formulae, we obtain for the last two terms

$$\begin{aligned}[\mathrm{L}_+, z] &= -(x + \mathrm{i}y)\\ [\mathrm{L}_+, |\boldsymbol{x}|^2\,\partial_z] &= |\boldsymbol{x}|^2\,[\mathrm{L}_+, \partial_z] = -\,|\boldsymbol{x}|^2\,(\partial_x + \mathrm{i}\partial_y)\,.\end{aligned}$$

Where we exploited that L_+ does not contain a derivative with respect $|\boldsymbol{x}|$ (cf. eq. (4.33)) and, hence, we could take $|\boldsymbol{x}|^2$ out of the commutator. To compute the first term we note that

$$[\mathrm{L}_+, ABC] = [\mathrm{L}_+, A]BC + A[\mathrm{L}_+, B]C + AB[\mathrm{L}_+, C]\,.$$

Where A, B and C are arbitrary operators. Whence,

$$\begin{aligned}[\mathrm{L}_+, zx^m\partial_m] &= -(x + \mathrm{i}y)x^m\partial_m + z(\mathrm{i}\epsilon_{1mk} - \epsilon_{2mk})x^k\partial_m + z(\mathrm{i}\epsilon_{1mk} - \epsilon_{2mk})x^m\partial_k\\ &= -(x + \mathrm{i}y)x^m\partial_m\,.\end{aligned}$$

In the last step we renamed the indices and used that the Levi-Civita symbol is antisymmetric, i.e. $\epsilon_{ijk} = -\epsilon_{ikj}$. Using the last three results gives

$$[\mathrm{L}_+, \mathrm{D}_z] = -(\mathrm{D}_x + \mathrm{i}\mathrm{D}_y)\,.$$

A.3 DIRECT DERIVATION OF THE INVERSE TRANSFORMATION

We derive a closed-form formula for the inverse basis transformation, namely for $\alpha^m_{l,0,q,r}$ and $\alpha^m_{l,1,q,r}$.

To this end, we use that eq. (4.8) and eq. (4.16) describe the same potential ϕ, i.e.

$$\sum_{l=0}^{\infty} \frac{1}{l!r^{2l+1}} Q^{(l)}_{i_1\cdots i_l} x^{i_1}\cdots x^{i_l} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{2l+1}\frac{q^m_l}{r^{2l+1}} r^l Y^m_l(\theta,\varphi)\,.$$

Moreover, at the end of Sec. 4.2.2, we pointed out, that the terms in the above multipole expansion, namely $Q^{(l)}_{i_1\cdots i_l} x^{i_1}\cdots x^{i_l}$, are homogeneous and harmonic polynomials of degree l. The same is true for the solid harmonics $r^l Y_l^m$. This implies that

$$\frac{1}{l!r^{2l+1}} Q^{(l)}_{i_1\cdots i_l} x^{i_1}\cdots x^{i_l} = \sum_{m=-l}^{l} \frac{4\pi}{2l+1} \frac{q_l^m}{r^{2l+1}} r^l Y_l^m(\theta,\varphi)\,. \tag{A.6}$$

The factor $1/r^{2l+1}$ cancels. We can use the notation introduced in eq. (4.12) to express the left-hand side of the above equation as

$$\frac{1}{l!} \sum_{p+q+r=l} \frac{(p+q+r)!}{p!q!r!} Q^{(l)}_{pqr} x^p y^q z^r = \sum_{p+q+r=l} \frac{1}{p!q!r!} Q^{(l)}_{pqr} x^p y^q z^r\,.$$

The numerical factor in front of the Cartesian multipole moment $Q^{(l)}_{pqr}$ reflects that the tensor $Q^{(l)}$ is symmetric and that index combinations corresponding to specific values of p, q and r appear more than once. Hence, eq. (A.6) becomes

$$\sum_{p+q+r=l} \frac{1}{p!q!r!} Q^{(l)}_{pqr} x^p y^q z^r = \sum_{m=-l}^{l} \frac{4\pi}{2l+1} q_l^m r^l Y_l^m(\theta,\varphi)\,. \tag{A.7}$$

The left-hand side of eq. (A.7) is a linear combination of the monomials $x^p y^q z^r$ with $p+q+r=l$. If we are able to express the sum of solid harmonics on the right-hand side of this equation as such a linear combination as well, we can equate coefficients to determine the $Q^{(l)}_{pqr}$. Since the factors in front of the solid harmonics are the spherical multipole moments q_l^m, the components $Q^{(l)}_{pqr}$ of the Cartesian multipole moment must be a sum of them. Eq. (4.40) informs us that we can express the components $Q^{(l)}_{pqr}$ with $p > 1$ as a sum of the components with $p = 0$ or $p = 1$. Thus, we can restrict our attention to the coefficients in front of the monomials with p equal to zero or p equal to one. Furthermore, a look at eq. (4.51), tells us that these coefficients (which must be a sum of the spherical multipole moments) contain the complex conjugate of the $\alpha^m_{l,0,q,r}$ and $\alpha^m_{l,1,q,r}$. Hence, we proceed in two steps: First, we express the solid harmonics as a linear combination of monomials. Secondly, we isolate the $p = 0$ or $p = 1$ part of this sum and, subsequently, equate coefficients.

Step one begins with eq. (4.19), namely

$$r^l Y_l^m(\theta,\varphi) = (-1)^m N_l^m r^{l-m} \frac{\mathrm{d}^m}{\mathrm{d}\cos\theta^m} P_l(\cos\theta)(r\sin\theta\cos\varphi + \mathrm{i}r\sin\theta\sin\varphi)^m\,,$$

and we use the following closed-form expression of the Legendre Polynomials

$$P_l(\cos\theta) = \frac{1}{2^l} \sum_{j=0}^{\left\lfloor \frac{l}{2} \right\rfloor} (-1)^j \binom{l}{j} \binom{2l-2j}{l} \cos^{l-2j}\theta$$

to obtain

$$r^l Y_l^m(\theta,\varphi) = \sum_{j=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} R_{l,j}^m r^{2j} z^{l-m-2j} (x+\mathrm{i}y)^m \,.$$

Where the numerical factor $R_{l,j}^m$ is defined as

$$R_{l,j}^m := \frac{(-1)^{j+m}}{2^l} N_l^m \binom{l}{j} \binom{2l-2j}{l} \frac{(l-2j)!}{(l-2j-m)!} \,.$$

In a next step, we rewrite the expression in parenthesis using the binomial theorem. Furthermore, we apply the multinomial theorem to expand the factor r^{2j}. This yields

$$\begin{aligned} r^l Y_l^m(\theta,\varphi) = \sum_{j=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \sum_{i_1+i_2+i_3=j} \sum_{k=0}^{m} \mathrm{i}^{m-k} R_{l,j}^m \binom{j}{i_1\, i_2\, i_3} \\ \times \binom{m}{k} x^{k+2i_1} y^{m-k+2i_2} z^{l-m-2j+2i_3} \,. \end{aligned} \tag{A.8}$$

We finished step one: we found a way to write the solid harmonics as a linear combination of monomials.

We now isolate the part of the above sum where the exponent of x, namely $p = k + 2i_1$, is either zero or one. Note that if $p = 0$, then $k = 0$ and $i_1 = 0$. And since $i_1 + i_2 + i_3 = j$, the last statement implies that $i_3 = j - i_2$. We conclude that the $p = 0$ part of the sum is

$$\sum_{j=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \sum_{i_2=0}^{j} \mathrm{i}^m R_{l,j}^m \binom{j}{i_2} y^{m+2i_2} z^{l-m-2i_2} = \sum_{j=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \sum_{k=0}^{j} \mathrm{i}^m R_{l,j}^m \binom{j}{k} y^{m+2k} z^{l-m-2k} \,.$$

In the last line we relabelled the index i_2 to k. Since we would like to equate the coefficients in front of the monomials, we should reorder the above sum such that all coefficients in front of monomials with specific values of $q = m + 2k$ and $r = l - m - 2k$ are summed. The result of this reordering is

$$\sum_{j=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \sum_{k=0}^{j} \mathrm{i}^m R_{l,j}^m \binom{j}{k} y^{m+2k} z^{l-m-2k} = \sum_{k=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \left[\sum_{j=k}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \mathrm{i}^m R_{l,j}^m \binom{j}{k} \right] y^{m+2k} z^{l-m-2k} \,. \tag{A.9}$$

When isolating the $p = 1$ part of the sum in eq. (A.8), we get

$$\begin{aligned} &\sum_{j=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \sum_{k=0}^{j} m \mathrm{i}^{m-1} R_{l,j}^m \binom{j}{k} x^1 y^{m-1+2k} z^{l-m-2k} \\ &\quad = \sum_{k=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \left[\sum_{j=k}^{\left\lfloor \frac{l-m}{2} \right\rfloor} m \mathrm{i}^{m-1} R_{l,j}^m \binom{j}{k} \right] x^1 y^{m-1+2k} z^{l-m-2k} \,. \end{aligned} \tag{A.10}$$

Thus, we isolated the $p = 0$ and $p = 1$ part of the expression for a solid harmonic in terms of monomials.

Before we use eq. (A.7) to determine the components $Q^{(l)}_{pqr}$ via equating the coefficients of the two polynomials, we rewrite the right-hand side of this equation as follows

$$\sum_{p+q+r=l} \frac{1}{p!q!r!} Q^{(l)}_{pqr} x^p y^q z^r = \frac{4\pi}{2l+1} \sum_{m=0}^{l} \left(q_l^m r^l Y_l^m + (-1)^m q_l^{-m} r^l Y_l^{m*} \right) .$$

This equation in conjunction with eq. (A.9) implies for the $p = 0$ part of the sums on both sides of it that

$$\begin{aligned}
&\sum_{q=0}^{l} \frac{1}{q!(l-q)!} Q^{(l)}_{0q(l-q)} y^q z^{l-q} \\
&\quad = \frac{4\pi}{2l+1} \sum_{m=0}^{l} \left(\sum_{k=0}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \left[\sum_{j=k}^{\left\lfloor \frac{l-m}{2} \right\rfloor} \mathrm{i}^m R^m_{l,j} \binom{j}{k} \right] y^{m+2k} z^{l-m-2k} \right) \left(q_l^m + q_l^{-m} \right) \\
&\quad = \sum_{q=0}^{l} \left(\sum_{k=0}^{\left\lfloor \frac{q}{2} \right\rfloor} \frac{4\pi}{2l+1} \left[\sum_{j=k}^{\left\lfloor \frac{l-q+2k}{2} \right\rfloor} \mathrm{i}^{q-2k} R^{q-2k}_{l,j} \binom{j}{k} \right] \left(q_l^{q-2k} + q_l^{-(q-2k)} \right) \right) y^q z^{l-q} .
\end{aligned}$$

In the second line, we once more reordered the sum such that all coefficients in front of the monomials with same exponents were grouped together. This implies

$$Q^{(l)}_{0q(l-q)} = \sum_{k=0}^{\left\lfloor \frac{q}{2} \right\rfloor} \alpha^{*q-2k}_{l,0,q,(l-q)} q_l^{q-2k} + \alpha^{*-(q-2k)}_{l,0,q,(l-q)} q_l^{-(q-2k)}$$

for $q \in \{0, 1, \dots, l\}$ and where the *complex conjugate* inverse basis transformation is defined as

$$\alpha^{*q-2k}_{l,0,q,(l-q)} := \frac{4\pi}{2l+1} q!(l-q)! \mathrm{i}^{q-2k} \left[\sum_{j=k}^{\left\lfloor \frac{l-q+2k}{2} \right\rfloor} R^{q-2k}_{l,j} \binom{j}{k} \right] .$$

Note that $\alpha^{*-(q-2k)}_{l,0,q,(l-q)} = \alpha^{*q-2k}_{l,0,q,(l-q)}$.

The above computation can be repeated for the $p = 1$ part of the sums with eq. (A.10). This leads to

$$Q^{(l)}_{1q(l-q-1)} = \sum_{k=0}^{\left\lfloor \frac{q}{2} \right\rfloor} \alpha^{*q+1-2k}_{l,1p,q,(l-q)} q_l^{q+1-2k} + \alpha^{*-(q+1-2k)}_{l,1,q,(l-q)} q_l^{-(q+1-2k)}$$

for $q \in \{0, \dots, l-1\}$ with

$$\alpha^{*q+1-2k}_{l,1,q,(l-q)} := (q+1-2k) \frac{4\pi}{2l+1} q!(l-q-1)! \mathrm{i}^{q-2k} \left[\sum_{j=k}^{\left\lfloor \frac{l-q-1+2k}{2} \right\rfloor} R^{q+1-2k}_{l,j} \binom{j}{k} \right] .$$

Moreover, $\alpha^{*-(q+1-2k)}_{l,1,q,(l-q)} = -\alpha^{*q+1-2k}_{l,1,q,(l-q)}$.

The presented formulae for the αs allow a computation of the inverse basis transformation matrix $\boldsymbol{A}$ without inverting the basis transformation matrix $\boldsymbol{B}$. The arguments to derive the formulae are an adaptation of findings in Johnston, 1960 to our insight that only the components of $Q^{(l)}_{pqr}$ with $p \leq 1$ are needed.

Appendix B

PDEs for the expansion coefficients

B.1 Real representation matrices

At the end of Section (5.3.2), we referred the reader to the Appendix for a proof that the representation matrices corresponding to the real spherical harmonics are real.

We prove this statement by explicitly computing all the appearing matrices, i.e. we evaluate the matrix-matrix products in $\boldsymbol{S}^\dagger \boldsymbol{O}\boldsymbol{S}$, where $\boldsymbol{O}$ is a representation matrix with respect to the complex spherical harmonics of an arbitrary operator. We note that the matrix elements of the inverse basis transformation are

$$\begin{aligned}(\boldsymbol{S}^\dagger)_{i(l',m',s')j(l,m)} \\ = \delta_{ll'}\left(\frac{\delta_{mm'}+(-1)^{m'}\delta_{-m'm}}{\sqrt{2(1+\delta_{m'0})}}\delta_{s'0}+\frac{\mathrm{i}(\delta_{mm'}-(-1)^{m'}\delta_{-m'm})}{\sqrt{2}}\delta_{s'1}\right).\end{aligned} \tag{B.1}$$

We now demonstrate how to compute the matrix representation of an arbitrary operator $\hat{O}$ corresponding to the real spherical harmonics. We assume that we know its matrix representation with the respect to the complex spherical harmonics. Evaluating the necessary matrix-matrix products yields

$$\begin{aligned}&(\boldsymbol{O}_R)_{i(l',m',s')j(l,m,s)} \\ &\quad=\sum_{n,k}\sum_{n',k'}(\boldsymbol{S}^\dagger)_{i(l',m',s')h(n,k)}\boldsymbol{O}_{h(n,k)r(n',k')}(\boldsymbol{S})_{r(n',k')j(l,m,s)} \\ &\quad=\frac{1}{2}\left(\frac{\delta_{s'0}\delta_{s0}}{\sqrt{(1+\delta_{m'0})(1+\delta_{m0})}}-\mathrm{i}\frac{\delta_{s'0}\delta_{s1}}{\sqrt{(1+\delta_{m'0})}}+\mathrm{i}\frac{\delta_{s'1}\delta_{s0}}{\sqrt{(1+\delta_{m0})}}+\delta_{s'1}\delta_{s1}\right) \\ &\qquad\times(\boldsymbol{O}_{s's})_{i(l',m',s')j(l,m,s)}\,.\end{aligned} \tag{B.2}$$

The matrices $\boldsymbol{O}_{s's}$ are defined as

$$\begin{aligned}&(\boldsymbol{O}_{s's})_{i(l',m',s')j(l,m,s)} \\ &\quad\coloneqq \boldsymbol{O}_{i(l',m')j(l,m)}+(-1)^{m'+s'}\boldsymbol{O}_{i(l',-m')j(l,m)} \\ &\qquad+(-1)^{m+s}\boldsymbol{O}_{i(l',m')j(l,-m)}+(-1)^{m'+m+s'+s}\boldsymbol{O}_{i(l',-m')j(l,-m)}\,.\end{aligned} \tag{B.3}$$

Having derived an expression for the real matrix representation for an arbitrary operator, we can apply it to the angular momentum operators, the direction

operators, the collision operator and the rotation operators. We note that $\mathbf{O}_R$ contains complex terms and, thus, we have to check if the resulting representation matrices corresponding to the above list of operators are real.

When computing the matrix representations with respect to the real spherical harmonics, it is important to note that $m', m \geq 0$, hence $\delta_{-m'(m+1)} = \delta_{m'(-m-1)} = 0$ and $\delta_{-m'(m-1)} = \delta_{m'(-m+1)} = \delta_{m'0}\delta_{m1} + \delta_{m'1}\delta_{m0}$. Moreover, $m' = 0$ and $s' = 1$ (or $m = 0$ and $s = 1$) is excluded[4], which we use to remove terms from the resulting expressions. Generally, we remove terms with combinations of Kronecker deltas whose evaluation leads to values of l, m and s which are not allowed.

We begin with the representation matrices of the angular momentum operator. We remind ourselves that we have to include a factor i, because $\boldsymbol{\Omega}_R^a = \mathrm{i}\boldsymbol{S}^\dagger\boldsymbol{\Omega}^a\boldsymbol{S}$. Using eq. (B.2) yields

$$(\boldsymbol{\Omega}_R^x)_{i(l',m',s')j(l,m,s)} = \boldsymbol{\Omega}^x_{i(l',m')j(l,m)}\left(\frac{\delta_{s'0}\delta_{s1}}{\sqrt{1+\delta_{m'0}}} - \frac{\delta_{s'1}\delta_{s0}}{\sqrt{1+\delta_{m0}}}\right) \tag{B.4}$$

and

$$\begin{aligned}(\boldsymbol{\Omega}_R^y)_{i(l',m',s')j(l,m,s)} & \qquad (B.5)\\ &= \frac{\delta_{s'0}\delta_{s1}}{\sqrt{1+\delta_{m'0}}}\left(\boldsymbol{\Omega}^y_{i(l',m')j(l,m)} + \frac{\delta_{l'l}}{2}\sqrt{l(l+1)}\delta_{m'0}\delta_{m1}\right)\\ &\quad - \frac{\delta_{s'1}\delta_{s0}}{\sqrt{1+\delta_{m0}}}\left(\boldsymbol{\Omega}^y_{i(l',m')j(l,m)} + \frac{\delta_{l'l}}{2}\sqrt{l(l+1)}\delta_{m'1}\delta_{m0}\right)\\ (\boldsymbol{\Omega}_R^z)_{i(l',m',s')j(l,m,s)} & \qquad (B.6)\\ &= \frac{\delta_{s'0}\delta_{s0}}{\sqrt{(1+\delta_{m'0})(1+\delta_{m0})}}\\ &\quad \times\left(\mathrm{i}\boldsymbol{\Omega}^z_{i(l',m')j(l,m)} - \frac{\delta_{l'l}}{2}\sqrt{l(l+1)}(\delta_{m'0}\delta_{m1} - \delta_{m'1}\delta_{m0})\right)\\ &\quad + \delta_{s'1}\delta_{s1}\mathrm{i}\boldsymbol{\Omega}^z_{i(l',m')j(l,m)}\,.\end{aligned}$$

We proceed with the matrix representations of the direction operators $\hat{A}^a$ operators. We find that

$$(\boldsymbol{A}_R^x)_{i(l',m',s')j(l,m,s)} = \delta_{s's}\boldsymbol{A}^x_{i(l',m')j(l,m)} \tag{B.7}$$

and

$$
\begin{aligned}
&(\boldsymbol{A}_R^y)_{i(l',m',s')j(l,m,s)} \qquad (B.8)\\
&\quad = \frac{\delta_{s'0}\delta_{s0}}{\sqrt{(1+\delta_{m'0})(1+\delta_{m0})}}\\
&\qquad \times \left((\boldsymbol{A}^y)_{i(l',m')j(l,m)} + \frac{1}{2}\left(a_{l+1}^{-m+1}\delta_{l'(l+1)} - a_l^m \delta_{l'(l-1)}\right)(\delta_{m'0}\delta_{m1} - \delta_{m'1}\delta_{m0})\right)\\
&\qquad + \delta_{s'1}\delta_{s1}(\boldsymbol{A}^y)_{i(l',m')j(l,m)}\\
&(\boldsymbol{A}_R^z)_{i(l',m',s')j(l,m,s)} \qquad (B.9)\\
&\quad = \frac{\delta_{s'0}\delta_{s1}}{\sqrt{1+\delta_{m'0}}}\left(-\mathrm{i}(\boldsymbol{A}^z)_{i(l',m')j(l,m)} + \frac{1}{2}\left(a_{l+1}^{-m+1}\delta_{l'(l+1)} - a_l^m\delta_{l'(l-1)}\right)\delta_{m'0}\delta_{m1}\right)\\
&\qquad - \frac{\delta_{s'1}\delta_{s0}}{\sqrt{1+\delta_{m0}}}\left(-\mathrm{i}(\boldsymbol{A}^z)_{i(l',m')j(l,m)} + \frac{1}{2}\left(a_{l+1}^{-m+1}\delta_{l'(l+1)} - a_l^m\delta_{l'(l-1)}\right)\delta_{m'1}\delta_{m0}\right)
\end{aligned}
$$

We point out that $\boldsymbol{A}^x = \boldsymbol{A}_R^x$.

The same is true for the real representation matrix of the collision operator, namely

$$(\boldsymbol{C}_R)_{i(l'm's')j(l,m,s)} = \frac{l(l+1)}{2}\delta_{l'l}\delta_{m'm}\delta_{s's}\,. \qquad (B.10)$$

It is left to compute the real representation matrices of the used rotation operators, i.e. $e^{\mathrm{i}\frac{\pi}{2}\hat{L}_x}$ and $e^{\mathrm{i}\frac{\pi}{2}\hat{L}_z}$. They are

$$
\begin{aligned}
&\left(e^{\frac{\pi}{2}\boldsymbol{\Omega}_R^x}\right)_{i(l',m',s')j(l,m,s)} \qquad (B.11)\\
&\quad = \delta_{l'l}\delta_{m'm}\\
&\qquad \times \left[\frac{\delta_{s'0}\delta_{s0}}{\sqrt{(1+\delta_{m'0})(1+\delta_{m0})}}\left(\cos\left(\frac{\pi}{2}m\right) + 1\delta_{m'0}\delta_{m0}\right) + \frac{\delta_{s'1}\delta_{s0}}{\sqrt{1+\delta_{m'0}}}\sin\left(\frac{\pi}{2}m\right)\right.\\
&\qquad\qquad \left. - \frac{\delta_{s'0}\delta_{s1}}{\sqrt{1+\delta_{m0}}}\sin\left(\frac{\pi}{2}m\right) + \delta_{s'1}\delta_{s1}\cos\left(\frac{\pi}{2}m\right)\right]\\
&\left(e^{\frac{\pi}{2}\boldsymbol{\Omega}_R^z}\right)_{i(l',m',s')j(l,m,s)} \qquad (B.12)\\
&\quad = \frac{\delta_{s'0}\delta_{s0}}{\sqrt{(1+\delta_{m'0})(1+\delta_{m0})}}\\
&\qquad \times \left((\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,m)} + (-1)^m\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,-m)})\right)\\
&\qquad + \delta_{s'1}\delta_{s1}\left((\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,m)} - (-1)^m\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,-m)})\right)\\
&\qquad\qquad (B.13)
\end{aligned}
$$

For the last equation we assumed that $e^{\frac{\pi}{2}\boldsymbol{\Omega}_R^z}$ is real, which is of course against the idea to prove that it is real. Since $\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)$ is real as well, we concluded

that the two terms in middle of eq. (B.2) must be zero for $\boldsymbol{U}_R(-\pi/2\boldsymbol{e}_z)$. This conclusion led us to conjecture that

$$
\begin{aligned}
(-1)^{m'}\boldsymbol{U}(\pi/2\boldsymbol{e}_z)_{i(l',-m')j(l,m)} &= (-1)^{m}\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,-m)} \\
\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,m)} &= (-1)^{m'+m}\boldsymbol{U}(-\pi/2\boldsymbol{e}_z)_{i(l',m')j(l,-m)}\,.
\end{aligned}
$$

This conjectures give us the above result and numerical experiments convinced us that these conjectures are true.

APPENDIX C

NUMERICAL SOLUTION OF THE SYSTEM OF PDES

C.1 DEFINITION OF THE REAL SPHERICAL HARMONICS

`Sapphire++` solves the system of equations (6.1) and, thus, it computes the expansion coefficients f_{lms}. For a physical interpretation of the results, a reconstruction of the distribution function f may be useful. This requires explicit expressions of the spherical harmonics. Eqs. (5.53) and (5.55) state that the real spherical harmonics are

$$Y_{lms}(\theta,\varphi) = N_{lm} P_l^m(\cos\theta)\left(\delta_{s0}\cos m\varphi + \delta_{s1}\sin m\varphi\right) . \tag{C.1}$$

Where N_{lm} is a normalisation, which is

$$N_{lm} = \sqrt{\frac{2l+1}{2\pi(1+\delta_{m0})}\frac{(l-m)!}{(l+m)!}} , \tag{C.2}$$

and the functions P_l^m are the associated Legendre Polynomials. Their definition is given in (Milton & Stegun, 1964, eq. 8.6.6) and it is

$$P_l^m(\cos\theta) \coloneqq (-1)^m \sin^m\theta \frac{\mathrm{d}^m}{\mathrm{d}(\cos\theta)^m} P_l(\cos\theta) . \tag{C.3}$$

Note that the *Condon–Shortley* phase $(-1)^m$ is included in the definition of the associated Legendre polynomials and *not* in the definition of the spherical harmonics. P_l is the *Legendre polynomial* of degree l. A definition of the Legendre polynomial P_l is given through

$$P_l(\cos\theta) \coloneqq \frac{1}{2^l l!}\frac{\mathrm{d}^l}{\mathrm{d}(\cos\theta)^l}(\cos^2\theta - 1)^l , \tag{C.4}$$

which can, for example, be found in (Milton & Stegun, 1964, eq. 8.6.18).

The real spherical harmonics are related to the complex spherical harmonics via

$$Y_{lms}(\theta,\varphi) = \frac{1}{\sqrt{2(1+\delta_{m0})}}(-i)^s\left(Y_l^m(\theta,\varphi) + (-1)^s Y_l^{m*}(\theta,\varphi)\right) , \tag{C.5}$$

Table C.1: List of real spherical harmonics $Y_{lms}(\theta,\varphi)$ for $l \leq 2$.

Y_{lms}	Explicit expression
$Y_{000}(\theta,\varphi)$	$\sqrt{\frac{1}{4\pi}}$
$Y_{100}(\theta,\varphi)$	$\sqrt{\frac{3}{4\pi}}\cos\theta$
$Y_{110}(\theta,\varphi)$	$-\sqrt{\frac{3}{4\pi}}\sin\theta\cos\varphi$
$Y_{111}(\theta,\varphi)$	$-\sqrt{\frac{3}{4\pi}}\sin\theta\sin\varphi$
$Y_{200}(\theta,\varphi)$	$\frac{1}{4}\sqrt{\frac{5}{\pi}}\left(3\cos^2\theta-1\right)$
$Y_{210}(\theta,\varphi)$	$\frac{-1}{2}\sqrt{\frac{15}{\pi}}\sin\theta\cos\theta\cos\varphi$
$Y_{211}(\theta,\varphi)$	$\frac{-1}{2}\sqrt{\frac{15}{\pi}}\sin\theta\cos\theta\sin\varphi$
$Y_{220}(\theta,\varphi)$	$\frac{1}{4}\sqrt{\frac{15}{\pi}}\sin^2\theta\cos\varphi$
$Y_{221}(\theta,\varphi)$	$\frac{1}{4}\sqrt{\frac{15}{\pi}}\sin^2\theta\sin\varphi$

cf. eq. (5.55). More explicitly,

$$Y_{lms}(\theta,\varphi) = Y_l^m(\theta,\varphi) \qquad \text{for } s=0, m=0, \tag{C.6}$$

$$Y_{lms}(\theta,\varphi) = \frac{1}{\sqrt{2}}\left(Y_l^m(\theta,\varphi) + (-1)^m Y_l^{-m}(\theta,\varphi)\right) \qquad \text{for } s=0, m>0 \tag{C.7}$$

$$Y_{lms}(\theta,\varphi) = \frac{1}{\sqrt{2}i}\left(Y_l^m(\theta,\varphi) - (-1)^m Y_l^{-m}(\theta,\varphi)\right) \qquad \text{for } s=1, m>0\,, \tag{C.8}$$

Notice that for $s=1$, m is by definition greater than zero. We list the first few real spherical harmonics in Tab. C.1.

C.2 HIGHER ORDER CORRECTIONS

As mentioned, dropping the relativistic corrections in front of the time derivative is accurate to order $\mathcal{O}(U/c)$, also see the discussion in Sec. 3.2.4. In this appendix we want to demonstrate two different ways, to retain higher order corrections in U/c.

Instead of dropping all terms of order $\mathcal{O}(U/c)$ in our expansion of the Lorentz transformation, see eq. (3.57), we keep the first order terms. This results in the

following the VFP equation in mixed coordinates

$$\left(1+\frac{\boldsymbol{U}\cdot\boldsymbol{v}'}{c^2}\right)\frac{\partial f}{\partial t}+(\boldsymbol{U}+\boldsymbol{v}')\cdot\nabla_x f - \left(\gamma' m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}+(\boldsymbol{p}'\cdot\nabla_x)\boldsymbol{U}\right)\cdot\nabla_{p'} f + q\boldsymbol{v}'\cdot\left(\boldsymbol{B}'\times\nabla_{p'} f\right)=\frac{\nu'}{2}\Delta_{\vartheta',\varphi'} f\,. \quad \text{(C.9)}$$

One approach to retain the relativistic correction in front of the time derivative is to proceed as before and to apply the operator based method, as presented in Ch. 5, to arrive at the following system of PDEs

$$\left(\mathbf{1}+\frac{\upsilon}{c^2}U_a\boldsymbol{A}^a\right)\frac{\partial \boldsymbol{f}}{\partial t}+\left(U^a\mathbf{1}+\upsilon\boldsymbol{A}^a\right)\frac{\partial \boldsymbol{f}}{\partial x^a}-\left(\gamma m\frac{\mathrm{D}U_a}{\mathrm{D}t}\boldsymbol{A}^a+p\frac{\partial U_b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^b\right)\frac{\partial \boldsymbol{f}}{\partial p} + \left(\frac{1}{\upsilon}\epsilon_{abc}\frac{\mathrm{D}U^a}{\mathrm{D}t}\boldsymbol{A}^b\boldsymbol{\Omega}^c+\epsilon_{bcd}\frac{\partial U^b}{\partial x^a}\boldsymbol{A}^a\boldsymbol{A}^c\boldsymbol{\Omega}^d\right)\boldsymbol{f}-\omega_a\boldsymbol{\Omega}^a\boldsymbol{f}+\nu\boldsymbol{C}\boldsymbol{f}=0\,. \quad \text{(C.10)}$$

Note that we dropped the primes. Its implicitly assumed that all quantities related to momentum variables are defined in the rest frame of the background plasma. Using the dG method and the explicit Euler method to discretise the equations yields

$$\tilde{\boldsymbol{M}}\frac{\boldsymbol{\zeta}^n-\boldsymbol{\zeta}^{n-1}}{\Delta t}=\boldsymbol{h}^{n-1}-\boldsymbol{D}^{n-1}\boldsymbol{\zeta}^{n-1}\,, \quad \text{(C.11)}$$

cf. eq. (6.29). Notice that we introduced the modified mass matrix, namely

$$(\tilde{\boldsymbol{M}})_{ij}\coloneqq\sum_{T\in\mathcal{T}_h}\int_T\boldsymbol{\phi}_i\left(\mathbf{1}+\frac{\upsilon}{c^2}U_a\boldsymbol{A}^a\right)\boldsymbol{\phi}_j\,, \quad \text{(C.12)}$$

that includes the relativistic time correction. Solving the system of PDEs involves solving a linear system of equations given by the modified mass matrix. This is computationally more expensive than solving the system of equations corresponding to the ordinary mass matrix, because the modified mass matrix is less sparse. This statement holds true for explicit time stepping. For implicit time steps, the linear system of equations is more complex and we do not expect that incorporating the modified mass matrix affects the solution time much.

In the second approach, we multiply eq. (C.9) with $1-\boldsymbol{U}\cdot\boldsymbol{v}'/c^2$ and neglect terms of order $\mathcal{O}\left((U/c)^2\right)$. This gives

$$\frac{\partial f}{\partial t}+\left(\boldsymbol{U}+\boldsymbol{v}'-\boldsymbol{v}'\frac{\boldsymbol{U}\cdot\boldsymbol{v}'}{c^2}\right)\cdot\nabla_x f-\left(\gamma' m\frac{\mathrm{D}\boldsymbol{U}}{\mathrm{D}t}+(\boldsymbol{p}'\cdot\nabla_x)\boldsymbol{U}\right)\cdot\nabla_{p'} f + q\left(1-\frac{\boldsymbol{U}\cdot\boldsymbol{v}'}{c^2}\right)\boldsymbol{v}'\cdot\left(\boldsymbol{B}'\times\nabla_{p'} f\right)=\frac{\nu'}{2}\left(1-\frac{\boldsymbol{U}\cdot\boldsymbol{v}'}{c^2}\right)\Delta_{\vartheta',\varphi'} f\,. \quad \text{(C.13)}$$

Applying the operator based method, we arrive at the following system,

$$\frac{\partial \boldsymbol{f}}{\partial t} + \left(U^a \mathbf{1} + \upsilon \boldsymbol{A}^a - \frac{\upsilon^2}{c^2} U_b \boldsymbol{A}^a \boldsymbol{A}^b \right) \frac{\partial \boldsymbol{f}}{\partial x^a} - \left(\gamma m \frac{\mathrm{D}U_a}{\mathrm{D}t} \boldsymbol{A}^a + p \frac{\partial U_b}{\partial x^a} \boldsymbol{A}^a \boldsymbol{A}^b \right) \frac{\partial \boldsymbol{f}}{\partial p} + \left(\frac{1}{\upsilon} \epsilon_{abc} \frac{\mathrm{D}U^a}{\mathrm{D}t} \boldsymbol{A}^b \boldsymbol{\Omega}^c + \epsilon_{bcd} \frac{\partial U_b}{\partial x^a} \boldsymbol{A}^a \boldsymbol{A}^c \boldsymbol{\Omega}^d \right) \boldsymbol{f} - \left(\mathbf{1} - \frac{\upsilon}{c^2} U_b \boldsymbol{A}^b \right) \omega_a \boldsymbol{\Omega}^a \boldsymbol{f} + \nu \left(\mathbf{1} - \frac{\upsilon}{c^2} U_a \boldsymbol{A}^a \right) \boldsymbol{C} \boldsymbol{f} = 0 \,. \quad \text{(C.14)}$$

Computing the upwind flux for the term $\propto U_b \boldsymbol{A}^a \boldsymbol{A}^b \partial_{x^a}$ requires the eigenvalues and eigenvectors of the combined matrix $U_b \boldsymbol{A}^a \boldsymbol{A}^b$. So far, we are not aware of an analytical solution of the corresponding eigenproblem. A numerical solution (similar to the term $\propto \frac{\partial U_b}{\partial x^a} \boldsymbol{A}^a \boldsymbol{A}^b \partial_p$) is again computationally expensive.

Finally, we note that another route is to *not* use a mixed-coordinate system. This implies the necessity to derive an explicit form of the scattering operator in the laboratory frame.

Bibliography

Achterberg, A., & Norman, C. A. (2018). Relativistic theory of particles in a scattering flow - I. Basic equations, diffusion, and drift. *Monthly Notices of the Royal Astronomical Society*, *479*(2), 1747–1770. https://doi.org/10.1093/mnras/sty1449

Achterberg, A., & Schure, K. M. (2011). A more accurate numerical scheme for diffusive shock acceleration. *Monthly Notices of the Royal Astronomical Society*, *411*(4), 2628–2636. https://doi.org/10.1111/j.1365-2966.2010.17868.x

Allis, W. P. (1956, January). Motions of ions and electrons. In S. Flügge (Ed.), *Electron–emission gas discharges I* (pp. 383–444, Vol. 4). Springer. https://doi.org/10.1007/978-3-642-45844-6_5

Arndt, D., Bangerth, W., Bergbauer, M., Feder, M., Fehling, M., Heinz, J., Heister, T., Heltai, L., Kronbichler, M., Maier, M., Munch, P., Pelteret, J.-P., Turcksin, B., Wells, D., & Zampini, S. (2023). The `deal.II` library, version 9.5. *Journal of Numerical Mathematics*, *31*(3), 231–246. https://doi.org/10.1515/jnma-2023-0089

Arndt, D., Bangerth, W., Davydov, D., Heister, T., Heltai, L., Kronbichler, M., Maier, M., Pelteret, J.-P., Turcksin, B., & Wells, D. (2021). The `deal.II` finite element library: Design, features, and insights. *Computers & Mathematics with Applications*, *81*, 407–422. https://doi.org/10.1016/j.camwa.2020.02.022

Axler, S., Bourdon, P., & Wade, R. (2001). *Harmonic function theory* (2nd ed., Vol. 137). Springer-Verlag New York. https://doi.org/10.1007/978-1-4757-8137-3

Baade, W., & Zwicky, F. (1934). Remarks on super-novae and cosmic rays. *Physical Review*, *46*, 76–77. https://doi.org/10.1103/PhysRev.46.76.2

Bell, A. R. (1978). The acceleration of cosmic rays in shock fronts - I. *Monthly Notices of the Royal Astronomical Society*, *182*(2), 147–156. https://doi.org/10.1093/mnras/182.2.147

Bell, A. R., Robinson, A. P. L., Sherlock, M., Kingham, R. J., & Rozmus, W. (2006). Fast electron transport in laser-produced plasmas and the KALOS code for solution of the Vlasov–Fokker–Planck equation. *Plasma Physics and Controlled Fusion*, *48*(3), R37–R57. https://doi.org/10.1088/0741-3335/48/3/R01

Bellan, P. M. (2006). *Fundamentals of plasma physics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807183

Beresnyak, A. (2023). *NRL plasma formulary*. U.S. Naval Research Laboratory. https://www.nrl.navy.mil/News-Media/Publications/NRL-Plasma-Formulary/

Billingsley, P. (1995). *Probability and measure* (3rd ed.). John Wiley & Sons.

Blandford, R., & Eichler, D. (1987). Particle acceleration at astrophysical shocks: A theory of cosmic ray origin. *Physics Reports*, *154*(1), 1–75. https://doi.org/10.1016/0370-1573(87)90134-7

Bothe, W., & Kolhörster, W. (1929). Das Wesen der Höhenstrahlung. *Zeitschrift für Physik*, *56*(11), 751–777. https://doi.org/10.1007/BF01340137

Chandrasekhar, S. (1943). Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, *15*, 1–89. https://doi.org/10.1103/RevModPhys.15.1

Cipriani, J., & Silvi, B. (1982). Cartesian expressions for electric multipole moment operators. *Molecular Physics*, *45*(2), 259–272. https://doi.org/10.1080/00268978200100211

Cockburn, B. (1998). An introduction to the discontinuous Galerkin method for convection-dominated problems. In A. Quarteroni (Ed.), *Advanced numerical approximation of nonlinear hyperbolic equations: Lectures given at the 2nd session of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Cetraro, Italy, June 23–28, 1997* (pp. 150–268). Springer. https://doi.org/10.1007/BFb0096353

Courant, R., & Hilbert, D. (1989, April 19). *Methods of mathematical physics* (Vol. 1). WILEY-VCH Verlag GmbH & Co. KGaA. https://doi.org/doi.org/10.1002/9783527617210

Debye, P., & Hückel, E. (1923). Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen. *Physikalische Zeitschrift*, *24*(9), 185–206. https://archive.org/details/1923-debye-huckel-theory-zur-theorie-der-elektrolyte-1/1923-debye-huckel-theory-1923-german-scan-1-high-exposure/page/n7/mode/2up

Decker, R. B. (1988). Computer modeling of test particle acceleration at oblique shocks. *Space Science Reviews*, *48*(3), 195–262. https://doi.org/10.1007/BF00226009

Dendy, R. (1990). *Plasma dynamics* (3rd ed.). Clarendon Press. https://global.oup.com/academic/product/plasma-dynamics-9780198520412?q=dendy&lang=en&cc=us

Di Pietro, D. A., & Ern, A. (2012). *Mathematical aspects of discontinuous Galerkin methods* (Vol. 69). Springer. https://doi.org/10.1007/978-3-642-22980-0

Drury, L. O. (1991). Time-dependent diffusive acceleration of test particles at shocks. *Monthly Notices of the Royal Astronomical Society*, *251*(2), 340–350. https://doi.org/10.1093/mnras/251.2.340

Drury, L. O. (1983). An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas. *Reports on Progress in Physics*, *46*(8), 973–1027. https://doi.org/10.1088/0034-4885/46/8/002

Drury, L. O., Axford, W. I., & Summers, D. (1982). Particle acceleration in modified shocks. *Monthly Notices of the Royal Astronomical Society*, *198*, 833–841. https://doi.org/10.1093/mnras/198.3.833

Efimov, S. (1979). Transition operator between multipole states and their tensor structure [Translated from Teoreticheskaya i Matematicheskaya Fizika, Vol. 30, No. 2, pp. 219–233, May, 1979.]. *Theoretical and Mathematical Physics*, *39*(2), 425–434. https://doi.org/10.1007/BF01014921

Epperlein, E. M., & Haines, M. G. (1986). Plasma transport coefficients in a magnetic field by direct numerical solution of the Fokker–Planck equation. *Physics of Fluids*, *29*(4), 1029–1041. https://doi.org/10.1063/1.865901

Evoli, C. (2020, December). The cosmic-ray energy spectrum. https://doi.org/10.5281/zenodo.4396125

Fakhri, H. (2016). Spherical harmonics $Y_l^m(\theta, \phi)$: Positive and negative integer representations of su(1, 1) for $l - m$ and $l + m$. *Advances in High Energy Physics*, *2016*, 7. https://doi.org/10.1155/2016/3732657

Fermi, E. (1949). On the origin of the cosmic radiation. *Physical Review*, *75*(8), 1169–1174. https://doi.org/10.1103/PhysRev.75.1169

Forman, M. A., & Drury, L. O. (1983). Time-dependent shock acceleration: Approximations and exact solutions. *Proceedings from the 18th International Cosmic Ray Conference*, *2*, 267–270.

Gaisser, T. K., Engel, R., & Resconi, E. (2016). *Cosmic rays and particle physics* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139192194

Garrett, C. K., & Hauck, C. D. (2016). On the eigenstructure of spherical harmonic equations for radiative transport. *Computers & Mathematics with Applications*, *72*(2), 264–270. https://doi.org/10.1016/j.camwa.2015.05.030

Goldstein, H., Poole, C. S., & Safko, J. L. (2014). *Classical mechanics* (3rd ed.) [new international edition]. Pearson.

Hairer, E., Wanner, G., & Nørsett, S. P. (1993, August 5). *Solving ordinary differential equations I: Nonstiff problems* (2nd ed., Vol. 8). Springer. https://doi.org/10.1007/978-3-540-78862-1

Hesthaven, J. S., & Warburton, T. (2008). *Nodal discontinuous Galerkin methods: Algorithms, analysis, and applications* (1st ed., Vol. 54). Springer. https://doi.org/10.1007/978-0-387-72067-8

Hillas, A. M. (1972). *Cosmic rays* (1st ed.). Pergamon Press. https://doi.org/doi.org/10.1016/C2013-0-02469-3

Hillas, A. M. (1984). The origin of ultra-high-energy cosmic rays. *Annual Review of Astronomy and Astrophysics*, *22*, 425–444. https://doi.org/10.1146/annurev.aa.22.090184.002233

Jackson, J. D. (1998). *Classical electrodynamics* (3rd ed.). John Wiley & Sons, Ltd. https://www.wiley.com/en-us/Classical+Electrodynamics%2C+3rd+Edition-p-9780471309321

Jeevanjee, N. (2011). *An introduction to tensors and group theory for physicists* (2nd ed.). Birkhäuser. https://doi.org/10.1007/978-0-8176-4715-5

Johnston, T. W. (1960). Cartesian tensor scalar product and spherical harmonic expansions in Boltzmann's equation. *Physical Review*, *120*, 1103–1111. https://doi.org/10.1103/PhysRev.120.1103

Kirk, J. G. (1994). Particle acceleration. In A. O. Benz & T. J.-L. Courvoisier (Eds.), *Plasma astrophysics* (1st ed., pp. 225–314, Vol. 24). Springer. https://doi.org/10.1007/3-540-31627-2_3

Klimontovich, Y. L. (1967). *The statistical theory of non-equilibrium processes in a plasma* (D. T. Haar, Ed.; H. S. H. Massey & O. M. Blunn, Trans.; Vol. 9). Pergamon Press. https://doi.org/10.1016/C2013-0-06978-2

Krall, N. A., & Trivelpiece, A. W. (1973). *Principles of plasma physics*. McGraw-Hill.

Kulsrud, R. M. (2005). *Plasma physics for astrophysics*. Princeton University Press. https://doi.org/10.1515/9780691213354

Landau, L., & Lifshitz, E. (1977). *Quantum mechanics: Non-relativistic theory* (3rd ed.). Pergamon Press.

LeVeque, R. J. (2002, August). *Finite volume methods for hyperbolic problems*. Cambridge University Press. https://doi.org/10.1017/cbo9780511791253

Longair, M. S. (1994). *High energy astrophysics: Stars, the Galaxy and the interstellar medium* (2nd ed., Vol. 2). Cambridge University Press. https://doi.org/10.1017/CBO9781139170505

Ludwig, A. C. (1991). The generalized multipole technique. *Computer Physics Communications*, *68*(1–3), 306–314. https://doi.org/10.1016/0010-4655(91)90205-Y

Mayer-Kuckuk, T. (2002). *Kernphysik*. Vieweg+Teubner Verlag. https://doi.org/10.1007/978-3-322-84876-5

Meyer-Vernet, N. (1993). Aspects of Debye shielding. *American Journal of Physics*, *61*(3), 249–257. https://doi.org/10.1119/1.17300

Milton, A., & Stegun, I. A. (1964). *Handbook of mathematical functions with formulas graphs and mathematical tables* (10th print. 1972, with corrections). U.S. Dept. of Commerce National Bureau of Standards.

Montgomery, D., & Tidman, D. A. (1964). *Plasma kinetic theory*. McGraw-Hill.

Müller, C. (1966). *Spherical harmonics* (1st ed., Vol. 17). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/BFb0094775

Munch, P., Kormann, K., & Kronbichler, M. (2021). Hyper.deal: An efficient, matrix-free finite-element library for high-dimensional partial differential equations. *ACM Trans. Math. Softw.*, *47*(4). https://doi.org/10.1145/3469720

Nolting, W. (2013). *Grundkurs Theoretische Physik 3 - Elektrodynamik* (10th ed.). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37905-5

Ottmer, E. R. (1959). Elster, Julius [Online version]. In *Neue Deutsche Biographie* (pp. 468–469, Vol. 4). Retrieved October 19, 2023, from https://www.deutsche-biographie.de/pnd118530054.html#ndbcontent

Pasquale, B. (2013). The origin of galactic cosmic rays. *The Astronomy and Astrophysics Review*, *21*, 21–70. https://doi.org/10.1007/s00159-013-0070-7

Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge University Press. http://www.cambridge.org/9780521880688

Reville, B., & Bell, A. R. (2013). Universal behaviour of shock precursors in the presence efficient cosmic ray acceleration. *Monthly Notices of the Royal Astronomical Society*, *430*(4), 2873–2884. https://doi.org/10.1093/mnras/stt100

Risken, H. (1996, September 17). *The Fokker–Planck equation: Methods of solution and applications* (2nd ed., Vol. 18). Springer-Verlag Berlin Heidelberg. https://doi.org/doi.org/10.1007/978-3-642-61544-3

Schram, P. P. J. M. (1991). *Kinetic theory of gases and plasmas*. Springer Science+Business Media Dordrecht. https://doi.org/10.1007/978-94-011-3612-9

Schween, N. W., & Reville, B. (2022a). Multipole-conv: A multipole moment converter. https://github.com/nils-schween/multipole-conv

Schween, N. W., & Reville, B. (2022b). Converting between the Cartesian tensor and spherical harmonic expansion of solutions to the Boltzmann equation. *Journal of Plasma Physics*, *88*(5), Article 905880510. https://doi.org/10.1017/S002237782200099X

Schween, N. W., & Reville, B. (2024). Using spherical harmonics to solve the Boltzmann equation: An operator-based approach. *Monthly Notices of the Royal Astronomical Society*. https://doi.org/10.1093/mnras/stae596

Schween, N. W., Schulze, F., & Reville, B. (2024). Sapphire++: A particle transport code combining a spherical harmonic expansion and the discontinuous Galerkin method [Preprint]. https://doi.org/10.2139/ssrn.4808843

Shalchi, A. (2012). Gyrophase diffusion of charged particles in random magnetic fields: Gyrophase diffusion. *Monthly Notices of the Royal Astronomical Society*, *426*(2), 880–891. https://doi.org/10.1111/j.1365-2966.2012.21690.x

Shkarofsky, I., Johnston, T., & Bachynski, M. (1966). *The particle kinetics of plasmas*. Addison-Wesley Publishing Company.

Shu, C.-W. (2009). High order weighted essentially nonoscillatory schemes for convection dominated problems. *SIAM Review*, *51*(1), 82–126. https://doi.org/10.1137/070679065

Stein, E. M., Shakarchi, R., & Bothe, W. (2003, April 27). *Complex analysis*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691113852/complex-analysis

Stein, E. M., Weiss, G., & Shi, E. (1972). *Introduction to Fourier analysis on Euclidean spaces* (Vol. 32). Princeton University Press. https://doi.org/10.1515/9781400883899

Thomas, A. G. R., Tzoufras, M., Robinson, A. P. L., Kingham, R. J., Ridgers, C. P., Sherlock, M., & Bell, A. R. (2012). A review of Vlasov-Fokker-Planck numerical modeling of inertial confinement fusion plasma. *Journal of Computational Physics*, *231*(3), 1051–1079. https://doi.org/10.1016/j.jcp.2011.09.028

Thorne, K. S. (1980). Multipole expansions of gravitational radiation. *Reviews of Modern Physics*, *52*(2), 299–340. https://doi.org/10.1103/RevModPhys.52.299

Thorne, K. S., & Blandford, R. D. (2017, September 5). *Modern classical physics: Optics, fluids, plasmas, elasticity, relativity, and statistical physics*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691159027/modern-classical-physics

Toptyghin, I. N. (1980). Acceleration of particles by shocks in a cosmic plasma. *Space Science Reviews*, *26*(2), 157–213. https : / / doi . org / 10 . 1007 / bf00167370

Tricomi, F. G. (1950). Sugli zeri dei polinomi sferici ed ultrasferici. *Annali di Matematica Pura ed Applicata*, *31*(1), 93–97. https://doi.org/10.1007/bf02428258

Tzoufras, M., Bell, A., Norreys, P., & Tsung, F. (2011). A Vlasov–Fokker–Planck code for high energy density physics. *Journal of Computational Physics*, *230*(17), 6475–6494. https://doi.org/10.1016/j.jcp.2011.04.034

Tzoufras, M., Tableman, A., Tsung, F. S., Mori, W. B., & Bell, A. R. (2013). A multi-dimensional Vlasov–Fokker–Planck code for arbitrarily anisotropic high-energy-density plasmas. *Physics of Plasmas*, *20*(5), Article 056303. https://doi.org/10.1063/1.4801750

Varshalovich, D. A., Moskalev, A. N., & Khersonskii, V. K. (1988, October). *Quantum theory of angular momentum*. World Scientific. https://doi.org/10.1142/0270

Weiser, M. (1964). Geitel, Hans [Online version]. In *Neue Deutsche Biographie* (p. 164, Vol. 6). Retrieved October 19, 2023, from https://www.deutsche-biographie.de/pnd118538284.html#ndbcontent

Williams, L. L., & Jokipii, J. R. (1991). Viscosity and inertia in cosmic-ray transport: Effects of an average magnetic field. *The Astrophysical Journal*, *371*, 639. https://doi.org/10.1086/169930

Zhang, S. T., Li, X. M., Liu, D. J., Li, X. X., Cheng, R. J., Lv, S., Huang, Z. M., Qiao, B., Liu, Z., Cao, L. H., Zheng, C. Y., & He, X. T. (2024). Vlasov–Fokker–Planck–Maxwell simulations for plasmas in inertial confinement fusion. *Computer Physics Communications*, *294*, Article 108932, 108932. https://doi.org/10.1016/j.cpc.2023.108932