Universität des Saarlandes

Fachbereich 9 – Mathematik

Mathematischer Preprint

Fast Deterministic Method of Solving the Boltzmann Equation for Hard Spheres

A. Bobylev, S. Rjasanow

Preprint No. 3

Saarbrücken 1999

Universität des Saarlandes

Fachbereich 9 – Mathematik

Fast Deterministic Method of Solving the Boltzmann Equation for Hard Spheres

A. Bobylev
Keldysh Institute of Applied Mathematics
Russian Academy of Sciences
Miusskaya Sq. 4
125047 Moscow
Russia
bobylev@phys.unit.no

S. Rjasanow
Fachbereich 9 - Mathematik
Universität des Saarlandes
Postfach 151150
66041 Saarbrücken
Germany
rjasanow@num.uni-sb.de

submitted: September 16, 1999

Preprint No. 3

Saarbrücken 1999

Edited by
FB 9 – Mathematik
Im Stadtwald
D–66041 Saarbrücken
Germany

Fax:	+ 49 681 302 4443
e-mail:	preprint@math.uni-sb.de
WWW:	http://www.math.uni-sb.de/

Abstract

A special form of the Boltzmann collision operator for the hard spheres model is introduced. The possibilities of fast numerical computation of the collision operator based on this form and the Fast Fourier Transform are discussed. A new difference scheme for the Boltzmann equation for the hard spheres model is developed. The results of some numerical tests and accuracy comparisons with the Direct Simulation Monte Carlo (DSMC) method are presented.

Contents

1 Introduction

An overwhelming majority of numerical methods for the Boltzmann equation are based on Monte-Carlo-type particle schemes (see [7],[1], [10] for a review). The main advantage of such schemes is their high efficiency, all algorithms are linear with respect to a number of particles. Obvious disadvantages of Monte Carlo schemes are stochastic noise and restricted accuracy. The Monte Carlo methods are almost perfect if we are only interested in lower (hydrodynamic) moments for the stationary problems. However, it is not that easy to obtain more detailed information about non-stationary solutions of the Boltzmann equation by using such methods.
On the other hand, the main disadvantage of deterministic methods is their lower efficiency: the numerical work needed to calculate the collision integral for all grid points in the velocity space is, roughly speaking, proportional to at least N^2, where N is a number of grid points in the velocity space.
How to overcome these difficulties? From the algebraic point of view, the computation of the Boltzmann collision integral reduces on the discrete level to the evaluation of a certain quadratic form having a rather complicated matrix. One of the most effective algebraic tools for this purpose is the algorithm of Fast Fourier Transform (FFT) (see [8],[9]). It requires a uniform discretisation in the velocity space which is convenient for the Boltzmann collision operator. It is natural to try to use FFT for our goals. In particular, it is well known [2] that the Boltzmann collision operator for Maxwell molecules is relatively simple in the Fourier representation. We used this property in our previous paper [4] and constructed the fast deterministic scheme for this special case of the intermolecular forces. Another attempt to use the same simplification for the numerical solution of the Boltzmann equation was made in [11]. However, the most widely used (and physically justified) molecular model is the model of particles as hard spheres. Unfortunately, no serious analytic simplification of the Boltzmann collision operator can be obtained for this model by the Fourier transformation. In spite of this fact, we show in the present paper that FFT can be applied successfully to construct an efficient numerical scheme in this practically important case too.
Our scheme is based on the special representation (similar to the Carleman representation [6]) of the collision integral for hard spheres. We derive the corresponding formulae in Section 2. Then, in Section 3, we construct a method (based on FFT) of calculation of the collision integral. In Section

4 the completely conservative numerical scheme for solving the spatially homogeneous equation is constructed. Some numerical results and their comparison with the results obtained by Bird's DSMC method are presented and discussed in Section 5. We note that our new scheme can be used for solving spatially inhomogeneous problems on the basis of the splitting algorithm.

2 Transformation of the collision operator

We consider the following initial value problem for the spatially homogeneous Boltzmann equation for the hard spheres model

$$\frac{\partial f}{\partial t}(v,t) = Q(f,f),\ t>0,\ v\in\mathbb{R}^3,\ f(v,0)=f_0(v)>0 \tag{1}$$

where

$$Q(f,f) = \int_{\mathbb{R}^3_w}\int_{S^2} |u|\left(f(v',t)f(w',t) - f(v,t)f(w,t)\right) dw\, de. \tag{2}$$

We use the following notations in (2)

- $v, w \in \mathbb{R}^3$ are pre-collision velocities;
- dw is the volume element in $\mathbb{R}^3_w$;
- $e \in S^2 \subset \mathbb{R}^3$ is a unit vector;
- de is the surface element on the unit sphere S^2;
- $u = v - w$ is the relative velocity of collision partners;
- v', w' are post-collision velocities defined by
  $$v' = U + \frac{1}{2}|u|e,\ w' = U - \frac{1}{2}|u|e,\ U = \frac{1}{2}(v+w).$$

The first step in our considerations is to rewrite the Boltzmann collision operator, given in (2) in the usual form, in a form which is more convenient for the application of the FFT algorithm. To this end we first prove the following technical lemma.

Lemma 1 *The following identity holds for any appropriate test function* $\Phi(z) : \mathbb{R}^3 \to \mathbb{R}$

$$\int_{\mathbb{R}^3} \Phi(z)\delta\left((z,u) + \frac{1}{2}|z|^2\right) dz = |u|\int_{S^2}\Phi(|u|e - u)de, \tag{3}$$

where $u \in \mathbb{R}^3$ denotes an arbitrary vector, (z, u) the Euclidian scalar product and $\delta(x)$ is the one-dimensional Dirac delta-function.

Proof:
We begin the proof from the left-hand side of (3) noting that

$$\int_{\mathbb{R}^3} \Phi(z)\delta\left((z,u)+\frac{1}{2}|z|^2\right)dz = 2\int_{\mathbb{R}^3} \Phi(z)\delta\left(|z+u|^2-|u|^2\right)dz.$$

Using the substitution $z = \tilde{z} - u,\ dz = d\tilde{z}$ and immediately removing the tilde sign we obtain in the spherical coordinates

$$z = \rho\, e,\ \ \rho \in [0,\infty),\ \ e \in S^2,\ \ dz = \rho^2 d\rho\, de$$

the assertion of the lemma

$$\int_{\mathbb{R}^3} \Phi(z)\delta\left((z,u)+\frac{1}{2}|z|^2\right)dz = 2\int_{\mathbb{R}^3} \Phi(z-u)\delta\left(|z|^2-|u|^2\right)dz$$

$$= 2\int_0^\infty \rho^2\delta\left(\rho^2-|u|^2\right)\int_{S^2}\Phi(\rho\, e-u)de\, d\rho = |u|\int_{S^2}\Phi(|u|e-u)de.$$

■

The next lemma presents a particular form of the Boltzmann collision operator which is very convenient for numerical computation using the FFT algorithm. For simplicity we omit the dependence of the function $f(v,t)$ on the variable t in the following considerations.

Lemma 2 *The Boltzmann collision operator (2) for hard spheres can be written in the following form*

$$Q(f,f) = \int_{S^2}\int_{S^2}\delta\left((e_1,e_2)\right)\left(\Phi(v,e_1)\Phi(v,e_2) - f(v)\Psi(v,e_1,e_2)\right)de_1 de_2 \quad (4)$$

where the functions $\Phi(v,e)$ and $\Psi(v,e_1,e_2)$ are defined as

$$\Phi(v,e) = \int_{-\infty}^{\infty}|\rho| f(v+\rho\, e)d\rho, \quad (5)$$

$$\Psi(v,e_1,e_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\rho_1||\rho_2| f(v+\rho_1 e_1+\rho_2 e_2)d\rho_1 d\rho_2. \quad (6)$$

Proof:
Using the previous lemma for

$$\begin{aligned} \Phi(|u|e-u) &= f(v')f(w')-f(v)f(w) \\ &= f\left(v+\frac{|u|e-u}{2}\right) f\left(w-\frac{|u|e-u}{2}\right) - f(v)f(w) \end{aligned}$$

we obtain for $Q(f,f)$ the following integral

$$\int_{\mathbb{R}^3\times\mathbb{R}^3} \delta\left((z,u)+\frac{1}{2}|z|^2\right)\left(f\left(v+\frac{1}{2}z\right)f\left(w-\frac{1}{2}z\right)-f(v)f(w)\right)dw\,dz. \quad (7)$$

Then using the substitution $z=2\tilde{z}$ and again immediately omitting the tilde sign we get

$$Q(f,f)=4\int_{\mathbb{R}^3\times\mathbb{R}^3}\delta\left((z,u+z)\right)\left(f(v+z)f(w-z)-f(v)f(w)\right)dw\,dz.$$

The next substitution is $w=y+z+v$. Thus, using this substitution and $\delta(x)=\delta(-x)$ we obtain

$$Q(f,f)=4\int_{\mathbb{R}^3\times\mathbb{R}^3}\delta\left((z,y)\right)\left(f(v+z)f(v+y)-f(v)f(v+y+z)\right)dy\,dz.$$

Now we switch to the spherical coordinates

$$y=\rho_1 e_1,\ z=\rho_2 e_2,\ dy\,dz=\rho_1^2\rho_2^2 d\rho_1 d\rho_2 de_1 de_2,$$

and obtain after simple transformations the following form of the collision operator $Q(f,f)$

$$\begin{aligned} 4\int_{S^2}\int_{S^2}\delta\left((e_1,e_2)\right)\Bigg(&\left(\int_0^\infty \rho_1 f(v+\rho_1 e_1)d\rho_1\right)\left(\int_0^\infty \rho_2 f(v+\rho_2 e_2)d\rho_2\right) \\ &-f(v)\int_0^\infty\int_0^\infty \rho_1\rho_2 f(v+\rho_1 e_1+\rho_2 e_2)d\rho_1 d\rho_2\Bigg)de_1 de_2. \end{aligned}$$

The last step of the proof is to remark that

$$4\delta\left((e_1,e_2)\right)=\delta\left((e_1,e_2)\right)+\delta\left((-e_1,e_2)\right)+\delta\left((e_1,-e_2)\right)+\delta\left((-e_1,-e_2)\right),$$

to use the substitutions $e_1=-\tilde{e}_1,\ \rho_1=-\tilde{\rho}_1,\ e_2=-\tilde{e}_2,\ \rho_2=-\tilde{\rho}_2$ at the proper places and to extend the integration in the inner integrals from $[0,\infty)$ to $(-\infty,\infty)$. Thus, the lemma is proved.

∎

Remark 1 *To the best of our knowledge the collision operator $Q(f,f)$ in the form (7) (for general intermolecular potential) was first used in [3] to derive a complete Landau expansion of $Q(f,f)$. A formula similar to (4) (for a more general case) was used in [5] to prove some inequalities for $Q(f,f)$. However, the two short notes [3], [5] do not contain the proofs of (4), (7). Actually, this sort of representation of $Q(f,f)$ is already implicitly present in Carleman's paper [6].*

Note that $Q(f,f)$ has the special structure (4) only for hard spheres. We have presented the short derivation of the formula (4) since it is the basis for our numerical method.
The integral (5) is called the generalised X-ray transform of the function, while (6) is the generalised Radon transform.
Note that the functions $\Phi(v,e)$ and $\Psi(v,e_1,e_2)$ defined in (5),(6) are integrals of the convolution type and can therefore be computed efficiently using the Fourier transform

$$\varphi(\xi) = \mathcal{F}[f](\xi) = \int_{\mathbb{R}^3} f(v) e^{\imath(v,\xi)} dv. \tag{8}$$

The inverse Fourier transform is then

$$f(v) = \mathcal{F}^{-1}[\varphi](v) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^3} \varphi(\xi) e^{-\imath(v,\xi)} d\xi.$$

The function $\Phi(v,e)$ can be written in the following form

$$\begin{aligned}
\Phi(v,e) &= \mathcal{F}^{-1}\left[\mathcal{F}\left[\Phi\right](\xi,e)\right](v,e) \\
&= \mathcal{F}^{-1}\left[\int_{\mathbb{R}^3} \Phi(v,e) e^{\imath(v,\xi)} dv\right](v,e) \\
&= \mathcal{F}^{-1}\left[\int_{-\infty}^{\infty} |\rho| \left(\int_{\mathbb{R}^3} f(v+\rho\, e) e^{\imath(v,\xi)} dv\right) d\rho\right](v,e).
\end{aligned}$$

Using the substitution $v + \rho\, e = \tilde{v}$ we obtain

$$\begin{aligned}
\Phi(v,e) &= \mathcal{F}^{-1}\left[\varphi(\xi) \int_{-\infty}^{\infty} |\rho| e^{-\imath\rho(e,\xi)} d\rho\right](v,e) \\
&= \mathcal{F}^{-1}\left[\varphi(\xi) d((\xi,e))\right](v,e),
\end{aligned} \tag{9}$$

where $d((\xi, e))$ denotes the one-dimensional Fourier transform of the function $|\rho|$ evaluated at (ξ, e). Since this Fourier transform does not exist in the usual sense we are forced to consider it in the sense of distributions. The distribution d can be computed analytically

$$d(\zeta) = -\frac{2}{|\zeta|^2}.$$

By analogy we get for $\Psi(v, e_1, e_2)$

$$\Psi(v, e_1, e_2) = \mathcal{F}^{-1}\left[\varphi(\xi) d((\xi, e_1)) d((\xi, e_2))\right](v, e_1, e_2). \tag{10}$$

Thus, the computation of the Boltzmann collision operator for the given function $f(v)$ involves the following steps:

1. Computation of the Fourier transform $\varphi(\xi)$ of the function $f(v)$ as in (8).

2. The double integral over the unit spheres (4) containing the $\delta-$function is in fact the averaging of the function

$$\Phi(v, e_1)\Phi(v, e_2) - f(v)\Psi(v, e_1, e_2) \tag{11}$$

   over all possible pairs of unit, mutually orthogonal vectors e_1, e_2. Note that

$$\int_{S^2}\int_{S^2} \delta((e_1, e_2)) de_1 de_2 = 8\pi^2.$$

3. For the given pair e_1, e_2 the computation of the expression (11) results in multiplication of the function $\varphi(\xi)$ with $d((\xi, e_1)), d((\xi, e_2)$ and $d((\xi, e_1))d((\xi, e_2))$ as well as three final Fourier transforms as in (9),(10).

Since the function $d(\zeta)$ is singular at zero we are forced to regularise it. Instead of the infinite interval of integration in (9) we consider

$$d_R(x) = \int_{-R}^{R} |\rho| e^{\imath\rho x} d\rho. \tag{12}$$

This function can be computed easily and we obtain

$$d_R(x) = 2\left((xR)\sin(xR) + \cos(xR) - 1\right)/x^2$$

with $d_R(0) = R^2$.

Remark 2 *The regularisation (12) means that we approximate the collision integral (4) as*

$$Q(f,f) \approx Q_R(f,f), \;\; R \to \infty, \tag{13}$$

where $Q_R(f,f)$ *is given by the same formula (4) in which the integrals (5),(6) are evaluated over the finite interval* $[-R,R]$ *in (5) and over* $[-R,R] \times [-R,R]$ *in (6).*

One can easily check that the approximation does not change the principal properties of the Boltzmann equation (conservation laws, H-theorem, equilibrium solutions and even Galilee invariance).
To choose R we use the following simple criteria. Assume that $f(v) = 0$ for $|v| > R_0$, then the maximal value $|u_{max}|$ of the relative velocity of particles does not exceed $2R_0$. Therefore $Q(f,f) = Q_R(f,f)$ if $R \geq 2R_0$. Roughly speaking, the error of our approximation has the same order as the maximal value of $f(v)$ on the surface of the sphere $|v| = R/2$. Fortunately, the distribution function $f(v)$ usually decreases, such as $\exp(-\alpha|v|^2)$ for large $|v|^2$, therefore we need to estimate α for a specific problem and then choose R such that $\alpha R^2 >> 4$. For the equilibrium value $\alpha = 1/(2T)$ this yields a rough rule $R >> 2\sqrt{2T}$.

3 Computation of the collision integral

The form of the Boltzmann collision operator described above is well suited for the fast numerical computation of the collision operator for the given function $f(v)$. Note that the main numerical work requires the computation of three Fourier transforms for each pair of vectors e_1, e_2. Since the highly efficient algorithm of Fast Fourier Transform can be employed in this situation we will get very high efficiency for the whole procedure.
The averaging over all unit, mutually orthogonal vectors e_1, e_2 can be done in a deterministic or in a stochastic way. If we use a deterministic way we introduce some discrete set of such pairs, compute the expression (11) for each pair and then take the average. The stochastic Monte-Carlo simulation is also very easy to perform. The pairs e_1, e_2 are chosen randomly, and then we take the average and control the stochastic fluctuations. In this paper we concentrate on the deterministic way.

3.1 Generalised X-ray transform

The effective and accurate numerical computation of the generalised X-ray transform (5) is obviously the most crucial step of our algorithm.
The numerical solution begins with the discretisation of the velocity space $\mathbb{R}^3$ using the nodes on the infinite lattice $h_v\mathbb{Z}^3$

$$v_k = V + h_v\,k,\ h_v > 0,\ k \in \mathbb{Z}^3. \tag{14}$$

Here, V denotes the bulk velocity

$$\rho V = \int_{\mathbb{R}^3_v} v f(v,t)dv = \int_{\mathbb{R}^3_v} v f_0(v)dv, \tag{15}$$

$$\rho = \int_{\mathbb{R}^3_v} f(v,t)dv = \int_{\mathbb{R}^3_v} f_0(v)dv. \tag{16}$$

Note that the density ρ, the bulk velocity V as well as the energy density per unit volume W

$$W = \frac{1}{2}\int_{\mathbb{R}^3_v} |v|^2 f(v,t)dv = \frac{1}{2}\int_{\mathbb{R}^3_v} |v|^2 f_0(v)dv \tag{17}$$

remain conserved during the time evolution. h_v denotes some positive discretisation parameter. Then the continuous Fourier transform (8) can be replaced by the following discrete one

$$\tilde{\varphi}(\xi) = h_v^3 \sum_{k\in\mathbb{Z}^3} f(v_k)e^{\imath(v_k,\,\xi)}.$$

In order not to overload the following formulae we omit the tilde sign from $\tilde{\varphi}(\xi)$ having in mind that we are now dealing with some approximation for the function $\varphi(\xi)$ defined in (8).
The function $\varphi(\xi)$ will be evaluated in the discrete set of points

$$\xi_j = h_\xi\,j,\ h_\xi > 0,\ j \in \mathbb{Z}^3$$

as follows

$$\varphi_j = \varphi(\xi_j) = h_v^3 e^{\imath(V,\,\xi_j)} \sum_{k\in\mathbb{Z}^3} f_k e^{\imath h_v h_\xi(k,\,j)}, \tag{18}$$

where the abbreviation f_k is used for $f(v_k)$. The algorithm of FFT requires the following relation between the mesh sizes h_v and h_ξ

$$h_v h_\xi = \frac{2\pi}{n},\ n = 2^{n_l},\ n_l \in \mathbb{N}. \tag{19}$$

The last assumption is due to the especially effective FFT if n is a power of two.
Now it is easy to see that the values

$$\sum_{k \in \mathbb{Z}^3} f_k e^{\imath \frac{2\pi}{n}(k,j)}$$

are $n-$periodic with respect to each single component of the vector j. It is therefore sufficient to compute them for

$$j \ :\ -\frac{n}{2} + 1 \le j_l \le \frac{n}{2},\ l = 1,2,3.$$

The values $e^{\imath \frac{2\pi}{n}(k,j)}$ are also $n-$periodic with respect to the single components of the vector k. Thus, if the parameters h_v and n are chosen so that the values of the function $f(v)$ can be neglected outside of the cube

$$Q = [-L, L]^3,\ 2L = h_v n,\ h_v = \frac{2L}{n} \tag{20}$$

then we restrict the infinite summation in (18) to

$$\begin{aligned} \varphi_j &= \varphi(\xi_j) = h_v^3 e^{\imath(V,\xi_j)} \sum_k f_k e^{\imath \frac{2\pi}{n}(k,j)}, \\ k &: -\frac{n}{2} + 1 \le k_l \le \frac{n}{2},\ l = 1,2,3. \end{aligned} \tag{21}$$

Here we discuss how to choose the parameter L correctly. As in Remark 2 from the previous Section, we assume for the moment that $f(v) = 0$ for $|v| > R_0$. Then we need to choose $R \ge 2R_0$ to get the equality $Q(f,f) = Q_R(f,f)$. Moreover it can be shown that $[Q(f,f)](v) = 0$ if $|v| \ge \sqrt{2}\,R_0$. Using the Fourier Transform method we implicitly introduce a periodic extension of the function $f(v)$ outside of the cube $[-L,L]^3$. Therefore we need to choose the parameter L sufficiently large to avoid any contribution of neighbouring domains (i.e. outside of the cube). Simple geometric consideration shows that in order to obtain correct (non-zero) values of $Q(f,f)$ inside the sphere $|v|^2 = 2R_0^2$ we need to choose

$$L \geq \frac{1}{2}\left(R + R_0(1+\sqrt{2})\right) \geq \frac{3+\sqrt{2}}{2} R_0.$$

In practice it is reasonable to use the following values of R and L (for given R_0): $R = 2R_0$ and $L = 2.5R_0$. Finally we can formulate the following criteria of the choice of R and L without mentioning R_0: choose the basic cube $[-L,L]^3$ and put $R = 0.8L$ in the above formula, then the error of the evaluation of $Q(f,f)$ can be estimated as $\max_{|v|=R/2} f(v)$.
The next step is to rearrange the numbering of the components of the three-dimensional vector $f \in \mathbb{R}^{n^3}$ as follows

$$\tilde{f}_{\tilde{k}} = f_k \tag{22}$$

where the components of the vector $\tilde{k}$ are defined

$$\tilde{k}_l = \begin{cases} k_l & , \quad k_l \geq 0 \\ n + k_l & , \quad k_l < 0 \end{cases} , \ l = 1,2,3. \tag{23}$$

Thus, there is a one-to-one correspondence between k and $\tilde{k}$. Using (22) and the obvious property

$$e^{\imath \frac{2\pi}{n}(k,j)} = e^{\imath \frac{2\pi}{n}(\tilde{k},\tilde{j})}$$

we rewrite (21) as

$$\tilde{\varphi}_{\tilde{j}} \ = \ \varphi(\xi_j) = h_v^3 e^{\imath (V,\xi_j)} \sum_{\tilde{k} \in Q_n} \tilde{f}_{\tilde{k}} e^{\imath \frac{2\pi}{n}(\tilde{k},\tilde{j})}$$

or in the matrix form

$$\begin{aligned} \tilde{\varphi} &= h_v^3 D_V F_3 \tilde{f}, \quad \tilde{\varphi}, \tilde{f} \in \mathbb{C}^N, \ N = n^3, && (24) \\ D_V &= \mathrm{diag}(e^{\imath (V,\xi_j)}, \ \tilde{j} \in Q_n), && (25) \\ F_3 &\in \mathbb{C}^{N \times N}. \end{aligned}$$

Q_n is the following set of indices

$$Q_n = \left\{ j \in \mathbb{Z}^3 : 0 \le j_l \le n-1,\ l = 1,2,3 \right\}.$$

The matrix F_3 involved in (24) is the matrix of the three-dimensional discrete Fourier transform. If n is a power of two as is required in (19) then the computation of φ can be done with $O(N \log(N) = O(n^3 \log(n))$ arithmetical operations using the FFT algorithm. Up to the constant factor the matrix F_3 is unitary

$$F_3^{-1} = \frac{1}{N} F_3^* = \frac{1}{n^3} F_3^*.$$

The numerical evaluation of the generalised X-ray transform (5) in the knots v_k now results in

$$\begin{aligned} \Phi_k &= \frac{1}{(2\pi)^3} h_\xi^3 \sum_{\tilde{j} \in Q_n} e^{-\imath(v_k, \xi_j)} d_R((\xi_j, e)) \tilde{\varphi}_{\tilde{j}} &(26) \\ &= \frac{1}{(2\pi)^3} h_\xi^3 \sum_{\tilde{j} \in Q_n} e^{-\imath \frac{2\pi}{n}(\tilde{k}, \tilde{j})} e^{-\imath(V, \xi_j)} d_R((\xi_j, e)) \tilde{\varphi}_{\tilde{j}} \end{aligned}$$

or in the matrix form

$$\tilde{\Phi} = \frac{1}{(2\pi)^3} h_\xi^3 F^* D_V^{-1} D_e \tilde{\varphi},$$

where the diagonal matrix D_V is defined in (25) and

$$D_e = \operatorname{diag}\left(d_R((\xi_j, e)),\ \tilde{j} \in Q_n \right).$$

The components of the vector $\tilde{\Phi}$ are related to those of Φ in the same way as in (22)

$$\tilde{\Phi}_{\tilde{k}} = \Phi_k.$$

Using (24),(19) and the commutativity of the diagonal matrices we finally obtain

$$\tilde{\Phi} = \frac{1}{n^3} F^* D_e F \tilde{f}. \qquad (27)$$

The accuracy of the formula (27) is defined by three parameters: the cutting parameter R in (12), the size L of the cube Q in (20) and the number of knots in one direction n. There are therefore three different errors. The first error is due to restriction of the infinite integration in (9) to the integration over $[-R, R]$ in (12). The second error is due to restriction of the infinite velocity spaces $\mathbb{R}^3$ to the cube (20). The last error is due to using the midpoint rectangular quadrature instead of the exact integration over the cube (20) and later, for the inverse Fourier transform in (26). Thus we have quadratic accuracy for smooth functions. It is clear that the procedure can easily be improved using other quadrature rules based on uniform discretisation such as the Simpson rule. The only change in this case is an additional diagonal matrix in (27) containing the weights of the quadrature.

3.2 Averaging procedure

The next step is the numerical realisation of the averaging procedure over all pairs of unit, mutually orthogonal vectors e_1, e_2 of the vector $F(e_1, e_2)$. i.e. the computation of the integral

$$\int_{S^2\times S^2} \delta\left((e_1, e_2)\right) F(e_1, e_2) de_1\, de_2. \tag{28}$$

The components of the vector F are

$$F_k(e_1, e_2) = \left(F^* D_1 F f\right)_k \left(F^* D_2 F f\right)_k - f_k \left(F^* D_1 D_2 F f\right)_k\,,\ k \in Q_n. \tag{29}$$

Here we have used the following abbreviations

$$\begin{aligned} D_1 &= \operatorname{diag}\left(d_R((\xi_j, e_1)),\ j \in Q_n\right), \\ D_2 &= \operatorname{diag}\left(d_R((\xi_j, e_2)),\ j \in Q_n\right). \end{aligned}$$

The vector (29) has the following important symmetry properties

$$F(e_1, e_2) = F(\pm e_1, \pm e_2) = F(\pm e_2, \pm e_1). \tag{30}$$

Thus, we are able to reduce the computational work using this symmetry. Then we consider the following parametrisation of the pair of the unit spheres in (28)

$$\begin{aligned} e_1(\phi,\mu,\zeta_1) &= U(\phi,\mu)\,(\cos(\zeta_1),\sin(\zeta_1),0)^T\,, & (31)\\ e_2(\phi,\mu,\zeta_2) &= U(\phi,\mu)\,(-\sin(\zeta_2),\cos(\zeta_2),0)^T\,, & (32)\\ 0\le\phi &< 2\pi,\ -1\le\mu\le 1,\ 0\le\zeta_1<2\pi,\ 0\le\zeta_1<2\pi & (33)\end{aligned}$$

where the three-dimensional rotation matrix $U(\phi,\mu)$ is defined as follows

$$U(\phi,\mu)=\begin{pmatrix} \mu+\sin^2\phi(1-\mu) & -\sin\phi\cos\phi(1-\mu) & \cos\phi\sqrt{1-\mu^2}\\ -\sin\phi\cos\phi(1-\mu) & \mu+\cos^2\phi(1-\mu) & \sin\phi\sqrt{1-\mu^2}\\ -\cos\phi\sqrt{1-\mu^2} & -\sin\phi\sqrt{1-\mu^2} & \mu \end{pmatrix}$$

and a^T is the transpose of a. The idea behind the above parametrisation is the following. The vectors e_1 and e_2 define in the case $e_1\neq\pm e_2$ a plane in $\mathbb{R}^3$. Let

$$e_3=\frac{e_1\times e_2}{|e_1\times e_2|}$$

be the unit normal vector of this plane. The third column of the matrix $U(\phi,\mu)$ is exactly the vector e_3 with the usual parametrisation. The two first columns of the matrix $U(\phi,\mu)$ build an orthogonal basis in this plane. Thus, there are such angles ζ_1 and ζ_2 that the representation (31),(32) valids.
The scalar product (e_1,e_2) now takes the following form

$$(e_1,e_2)=\sin(\zeta_1-\zeta_2)$$

which is convenient for further simplifications. Since the above transformation is orthogonal ($de_1\,de_2=d\phi\,d\mu\,d\zeta_1\,d\zeta_2$) we obtain for the integral (28)

$$\int_0^{2\pi}d\phi\int_{-1}^{1}d\mu\int_0^{2\pi}\int_0^{2\pi}\delta(\sin(\zeta_1-\zeta_2))F(e_1(\phi,\mu,\zeta_1),e_2(\phi,\mu,\zeta_2))d\zeta_1 d\zeta_2.$$

Substituting $\zeta=\zeta_1,\ z=\sin(\zeta_1-\zeta_2)$ with the Jacobian

$$d\zeta_1 d\zeta_2=\frac{d\zeta\,dz}{\sqrt{1-z^2}}$$

in the double integral with respect to ζ_1,ζ_2 leads to

$$\int_0^{2\pi}\int_{-1}^{1}\delta(z)F(e_1(\phi,\mu,\zeta,\zeta),e_2(\phi,\mu,\zeta-\arcsin z))\frac{d\zeta\,dz}{\sqrt{1-z^2}}\,.$$

Removing the $\delta-$function we now obtain the integral (28) in the following form

$$\int_0^{2\pi} d\phi \int_{-1}^{1} d\mu \int_0^{2\pi} F(e_1(\phi,\mu,\zeta),\, e_2(\phi,\mu,\zeta))d\zeta. \tag{34}$$

Using the properties

$$\begin{array}{rcl} e_1(\pi+\phi,-\mu,-\zeta+2\phi) & = & -e_1(\phi,\mu,\zeta), \\ e_2(\pi+\phi,-\mu,-\zeta+2\phi) & = & e_2(\phi,\mu,\zeta), \\ e_1(\phi,\mu,\pi+\zeta) & = & -e_1(\phi,\mu,\zeta), \\ e_2(\phi,\mu,\pi+\zeta) & = & -e_2(\phi,\mu,\zeta), \\ e_1(\phi,\mu,\pi/2+\zeta) & = & e_2(\phi,\mu,\zeta), \\ e_2(\phi,\mu,\pi/2+\zeta) & = & -e_1(\phi,\mu,\zeta) \end{array}$$

and (30) we restrict the integration to the interval $[0,1]$ with respect to μ and to the interval $[0,\pi/2]$ with respect to ζ. Thus, (34) takes the form

$$8\int_0^{2\pi} d\phi \int_0^{1} d\mu \int_0^{\pi/2} F(e_1(\phi,\mu,\zeta),\, e_2(\phi,\mu,\zeta))d\zeta.$$

The discretisation of the parameter domain $[0,2\pi)\times[0,1]\times[0,\pi/2]$ is realised using the nodes

$$\begin{array}{rcl} (\phi_{i_1},\mu_{i_2},\zeta_{i_3}) & = & (h_\phi(i_1-1),-1+h_\mu(i_2-1/2),h_\zeta(i_3-1)), \\ h_\phi & = & \dfrac{2\pi}{n_\phi},\ h_\mu=\dfrac{1}{n_\mu},\ h_\zeta=\dfrac{\pi}{2n_\zeta} \end{array} \tag{35}$$

where $n_\phi, n_\mu, n_\zeta \in \mathbb{N}$ are new discretisation parameters. This choice of discretisation corresponds to the numerical integration using the midpoint rectangular rule and is therefore of the quadratic order of accuracy for smooth functions.
Our numerical tests have shown that the above discretisation does not cover all necessary symmetries which are naturally involved in the analytical form of the integral (28). A completely symmetric integration rule for the unit sphere should involve, together with an integration knot

$$(\alpha,\beta,\gamma)^T \in S^2, \tag{36}$$

all its permutations and changes of the sign:

$$(\pm\alpha, \pm\beta, \pm\gamma)^T,\ (\pm\beta, \pm\alpha, \pm\gamma)^T,\ (\pm\gamma, \pm\beta, \pm\alpha)^T,$$
$$(\pm\alpha, \pm\gamma, \pm\beta)^T,\ (\pm\gamma, \pm\alpha, \pm\beta)^T,\ (\pm\beta, \pm\gamma, \pm\alpha)^T.$$

That means that a non-trivial integration knot (36)

$$\alpha \neq \beta \neq \gamma \neq 0$$

requires 47(!) additional knots in order to keep the symmetry of the problem. Fortunately, most of them are automatically involved in our integration rule having the knots (35) in the parameter domain because of the reduction of the initial domain of integration and if we choose n_ϕ as a multiple of 4. Nevertheless, two additional permutations are necessary to keep the symmetry. We introduce three permutation matrices

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\ P_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Then an approximation for $Q(f,f)(v_k)$ can be computed as

$$q_k = \frac{8\pi^2}{3n_\phi n_\mu n_\zeta} \sum_{j=1}^{3} \sum_{i_1,i_2,i_3=1}^{n_\phi,n_\mu,n_\zeta} \Big((F_3^* D_1 \varphi)_k\, \overline{(F_3^* D_2 \varphi)_k} - f_k \overline{(F_3^* D_1 D_2 \varphi)_k} \Big). \quad (37)$$

Here we have used the following abbreviations

$$\begin{aligned}
\varphi &= \frac{1}{n^3} F_3 \tilde{f}, && (38)\\
D_1 &= \operatorname{diag}\left(d_R((\xi_j, e_{1,j,i_1,i_2,i_3})),\ j \in Q_n\right), && (39)\\
D_2 &= \operatorname{diag}\left(d_R((\xi_j, e_{2,j,i_1,i_2,i_3})),\ j \in Q_n\right), && (40)\\
e_{1,j,i_1,i_2,i_3} &= P_j U(\phi_{i_1}, \mu_{i_2}) \left(\cos(\zeta_{i_3}), \sin(\zeta_{i_3}), 0\right)^T, &&\\
e_{2,j,i_1,i_2,i_3} &= P_j U(\phi_{i_1}, \mu_{i_2}) \left(-\sin(\zeta_{i_3}), \cos(\zeta_{i_3}), 0\right)^T. &&
\end{aligned}$$

Note that the diagonal matrices D_1, D_2 do not depend on the problem and therefore can be computed once in advance and stored in the computer for further use. The arithmetical work for the numerical evaluation of the Boltzmann collision operator at all knots in Q_n is $O(n_\phi n_\mu n_\zeta n^3 \log(n))$.

4 A difference scheme

In this section we discuss an explicit difference scheme for the initial value problem (1) and investigate its conservation properties.

4.1 An explicit scheme

Using the initial condition in (1) we compute the initial vector using the nodes (14)

$$f^0 \in \mathbb{R}^{n^3}, \; f_k^0 = f_0(v_k), \; k : \tilde{k} \in Q_n.$$

Then the time steps are

$$f^{m+1} = f^m + \tau q^m, \; \tau > 0, \; m = 0, 1, \dots , \tag{41}$$

where the vector q^m is defined in (37). Note that now we use $\tilde{f}^m$ in (38) instead of $\tilde{f}$. It is obvious that the time step parameter τ should be chosen small enough to guarantee that all components

$$1 - \tau \frac{8\pi^2}{3n_\phi n_\mu n_\zeta} \sum_{j=1}^{3} \sum_{i_1,i_2,i_3=1}^{n_\phi,n_\mu,n_\zeta} \mathrm{Re}\left(F_3^* D_1 D_2 \varphi^m\right)_k$$

remain positive during the time steps. This is the usual time-step restriction for explicit schemes.

4.2 Conservation properties

One of the most important properties of the Boltzmann equation is the conservation of the density ρ, bulk velocity V and energy density W (15),(16),(17) during the relaxation. The numerical form of these macroscopic quantities is

$$\begin{aligned} \rho_h &= h_v^3 \sum_j f_j^m, \\ \rho_h (V_h)_l &= h_v^3 \sum_j (v_j)_l f_j^m, \; l = 1, 2, 3, \\ (p_h^m)_{ll} &= h_v^3 \sum_j ((v_j)_l)^2 f_j^m, \; l = 1, 2, 3, & (42) \\ W_h &= \frac{1}{2} \left((p_h^m)_{11} + (p_h^m)_{22} + (p_h^m)_{33}\right). & (43) \end{aligned}$$

We first remark that the density ρ_h is conserved because of the following considerations. We use (41) and obtain

$$\rho_h^{m+1} = h_v^3 \sum_k f_k^{m+1} = h_v^3 \sum_k f_k^m + \tau\, h_v^3 \sum_k q_k^m$$

$$= \rho_h^m + \tau \frac{8\pi^2 h_v^3}{3 n_\phi n_\mu n_\zeta} \sum_{j=1}^{3} \sum_{i_1,i_2,i_3=1}^{n_\phi,n_\mu,n_\zeta} \left(F_3^* D_1 \varphi^m, F_3^* D_2 \varphi^m\right) - \left(f^m, F_3^* D_1 D_2 \varphi^m\right).$$

Since

$$F_3 F_3^* = n^3 I, \tag{44}$$

we obtain with (38)

$$\left(F_3^* D_1 \varphi^m, F_3^* D_2 \varphi^m\right) - \left(f^m, F_3^* D_1 D_2 \varphi^m\right) = 0,\ \forall i_1, i_2, i_3, j.$$

Thus, the numerical density is conserved.
Using (14) and the numbering of components as in (23) we can represent the above quantities as the following scalar products

$$\begin{aligned}
\rho_h &= h_v^3 (g^{(0)}, \tilde{f}^m), && (45)\\
g^{(0)} &= e \otimes e \otimes e \in \mathbb{R}^N, && (46)\\
\rho_h (V_h)_1 &= h_v^3 (V_h)_1 (g^{(0)}, \tilde{f}^m) + h_v^3 (g^{(1)}, \tilde{f}^m), && (47)\\
g^{(1)} &= a \otimes e \otimes e \in \mathbb{R}^N, && (48)\\
(p_h^m)_{11} &= h_v^3 ((V_h)_1)^2 (g^{(0)}, \tilde{f}^m) + 2 h_v^3 (V_h)_1 (g^{(1)}, \tilde{f}^m) + h_v^3 (g^{(2)},\ \tilde{f}^m), && (49)\\
g^{(2)} &= b \otimes e \otimes e \in \mathbb{R}^N. && (50)
\end{aligned}$$

Here $\otimes$ denotes the Kronecker product. The $n-$dimensional vectors e, a and b are defined as follows

$$\begin{aligned}
e &= (1, 1, \ldots, 1)^T \in \mathbb{R}^n,\\
a &= h_v (0, 1, \ldots, n/2 - 1, 0, -n/2 + 1, \ldots, -1)^T \in \mathbb{R}^n,\\
b &= h_v^2 (0^2, 1^2, \ldots, (n/2 - 1)^2, (n/2)^2, (-n/2 + 1)^2, \ldots, (-1)^2)^T \in \mathbb{R}^n.
\end{aligned}$$

Using (45) in (47) and then (42),(45) in (49) we rewrite these conditions as

$$\begin{aligned}
\rho_h &= h_v^3 (g^{(0)}, \tilde{f}^m), && (51)\\
0 &= h_v^3 (g^{(1)}, \tilde{f}^m), && (52)\\
(p_h^m)_{11} - \rho_h (V_h)_1 &= h_v^3 (g^{(2)}, \tilde{f}^m). && (53)
\end{aligned}$$

Our next aim is to rewrite (51),(52) and (53) in terms of the Fourier transform φ^m of the vector $\tilde{f}^m$ with

$$\varphi^m = \frac{1}{n^3} F_3 \tilde{f}^m.$$

The three-dimensional Fourier transform matrix F_3 fulfils

$$\varphi^m = h_v^3 F_3 f^m, \;\; F_3 = F_1 \otimes F_1 \otimes F_1. \tag{54}$$

Here F_1 denotes the matrix of the one-dimensional Fourier transform having the elements

$$f_{lm} = e^{\imath \frac{2\pi}{n} lm}, \quad l, m = 0, \ldots, n-1.$$

Using (44) we rewrite (38) as

$$\tilde{f}^m = F_3^* \varphi^m$$

and the scalar products (51),(52) and (53) as

$$\begin{aligned} \rho_h &= h_v^3 (g^{(0)}, F_3^* \varphi^m) = h_v^3 (F_3 g^{(0)}, \varphi^m), \\ 0 &= h_v^3 (g^{(1)}, F_3^* \varphi^m) = h_v^3 (F_3 g^{(1)}, \varphi^m), \\ (p_h^m)_{11} - \rho_h (V_h)_1 &= h_v^3 (g^{(2)}, F_3^* \varphi^m) = h_v^3 (F_3 g^{(2)}, \varphi^m). \end{aligned}$$

Thus, we need to compute the Fourier transforms of the vectors (46), (48) and (50). Using

$$f_{lm} = \overline{f_{l,n-m}}, \;\; m = 1, \ldots, n/2$$

and the well-known property

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

of the Kronecker product we get

$$\begin{aligned} F_3 g^{(0)} &= (F_1 \otimes F_1 \otimes F_1)(e \otimes e \otimes e) = F_1 e \otimes F_1 e \otimes F_1 e, \\ F_3 g^{(1)} &= (F_1 \otimes F_1 \otimes F_1)(a \otimes e \otimes e) = F_1 a \otimes F_1 e \otimes F_1 e, \\ F_3 g^{(2)} &= (F_1 \otimes F_1 \otimes F_1)(b \otimes e \otimes e) = F_1 b \otimes F_1 e \otimes F_1 e. \end{aligned}$$

Therefore the problem remains to compute the one-dimensional Fourier transforms of the vectors e, a and b. The Fourier transform of the vector e is trivial

$$(\psi_e)_l = (F_1 e)_l = n\delta_{l,0}, \;\; \psi_e = (n, 0, \dots, 0)^T$$

where $\delta_{l,m}$ denotes the Kronecker symbol. The following technical lemma will be useful to compute the Fourier transform of the vectors a and b.

Lemma 3 *The following relations are valid*

$$\sum_{k=1}^{n/2-1} k \, \sin\left(\frac{2\pi}{n}kj\right) = \begin{cases} 0 & , \;\; j = 0 \\ \frac{n}{4}(-1)^{j+1}\cot\left(\frac{\pi}{n}j\right) & , \;\; j = 1, \dots n-1 \end{cases} , \tag{55}$$

$$\sum_{k=1}^{n/2-1} k^2 \, \cos\left(\frac{2\pi}{n}kj\right) = \begin{cases} \frac{1}{24}(n-2)(n-1)n & , \;\; j = 0 \\ \frac{n}{4}(-1)^{j+1}\left(\frac{n}{2} - \sin^{-2}\left(\frac{\pi}{n}j\right)\right) & , \;\; j = 1, \dots n-1 \end{cases} \tag{56}$$

Proof:
We use the equality

$$\sum_{k=1}^{n/2-1} q^k = \frac{q^{n/2}-1}{q-1} - 1$$

for

$$q = e^{\imath kx},$$

differentiate it once for (55) and twice for (56) and finally evaluate it at

$$x = \frac{2\pi}{n}j.$$

The real parts of these expressions lead to (55) and (56) correspondingly.

■

Now we compute the Fourier transforms of the vectors a and b using the property (55).

$$\beta_j = (F_1 a)_j = \sum_{k=0}^{n-1} a_k f_{j,k} = h_v \sum_{k=1}^{n/2-1} k\left(f_{j,k} - f_{j,n-k}\right)$$

$$
\begin{aligned}
&= 2\imath\, h_v \sum_{k=1}^{n/2-1} k\mathrm{Im}(f_{j,k}) = 2\imath\, h_v \sum_{k=1}^{n/2-1} k\, \sin\left(\frac{2\pi}{n}kj\right) \\
&= \begin{cases} 0 & ,\quad j=0 \\ \imath h_v \frac{n}{2}(-1)^{j+1}\cot\left(\frac{\pi}{n}j\right) & ,\quad j=1,\ldots,n-1 \end{cases}.
\end{aligned}
$$

Note that

$$\beta_1 = \beta_{n/2} = 0,\ \beta_j = -\beta_{n-j},\ j=1,\ldots,n/2-1$$

Thus, for the three-dimensional Fourier transform of the vector $g^{(1)}$ we get the following result

$$(F_3 g^{(1)})_j = F_3 g^{(1)}_{(j_1,j_2,j_3)} = n^2 \beta_{j_1}\delta_{j_2,0}\delta_{j_3,0}$$

The equation (55) now takes the form

$$0 = n^2 h_v^3 \sum_{j_1=1}^{n/2-1} \beta_{j_1}(\varphi_{j_1,0,0} - \varphi_{n-j_1,0,0}) = n^3 h_v^4 \sum_{j_1=1}^{n/2-1} (-1)^{j_1}\cot(\frac{\pi}{n}j_1)\mathrm{Im}\varphi_{(j_1,0,0)}.$$

The Fourier transform of the vector b can be computed similarly

$$
\begin{aligned}
\gamma_j &= (F_1 b)_j = \sum_{k=0}^{n-1} b_k f_{j,k} = h_v^2\left(\frac{n^2}{4}f_{j,n/2} + \sum_{k=1}^{n/2-1} k\,(f_{j,k} + f_{j,n-k})\right) \\
&= h_v^2\left(\frac{n^2}{4}(-1)^j + 2\sum_{k=1}^{n/2-1} k^2\mathrm{Re}(f_{j,k})\right) \\
&= h_v^2\left(\frac{n^2}{4}(-1)^j + 2\sum_{k=1}^{n/2-1} k^2\, \cos\left(\frac{2\pi}{n}kj\right)\right) \\
&= h_v^2\begin{cases} \dfrac{n(n^2+2)}{12} & ,\quad j=0 \\ \frac{n}{2}(-1)^j\sin^{-2}\left(\frac{\pi}{n}j\right) & ,\quad j=1,\ldots,n-1 \end{cases}.
\end{aligned}
$$

Note that

$$\gamma_j = \gamma_{n-j},\ j=1,\ldots,n/2-1$$

Thus, for the three-dimensional Fourier transform of the vector $g^{(2)}$ we get the following result

$$(F_3 g^{(2)})_j = F_3 g^{(2)}_{(j_1,j_2,j_3)} = n^2 \gamma_{j_1} \delta_{j_2,0} \delta_{j_3,0}$$

The equation (55) now takes the form

$$(p_h^m)_{11} - \rho_h (V_h)_1 = n^2 h_v^3 \left(\gamma_0 \varphi^m_{(0,0,0)} + \sum_{j_1=1}^{n/2-1} \gamma_{j_1} (\varphi^m_{(j_1,0,0)} + \varphi^m_{(n-j_1,0,0)}) \right)$$

$$= n^3 h_v^5 \left(\frac{n^2+2}{12} \rho_h + \frac{1}{2} \varphi_{(n/2,0,0)} + \sum_{j_1=1}^{n/2-1} (-1)^{j_1} \sin^{-2}\left(\frac{\pi}{n} j_1\right) \mathrm{Re}\varphi^m_{(j_1,0,0)} \right).$$

Using (57) and (57) we obtain the formulae for the functions $\varphi^m_{e_l}$, $l = 1,2,3$

$$\mathrm{Im}\varphi^m_{e_l} = \tan\frac{\pi}{n} \sum_{m=2}^{n/2-1} (-1)^m \cot\left(\frac{\pi}{n} m\right) \mathrm{Im}\varphi^m_{m\,e_l}, \tag{57}$$

$$\mathrm{Re}\varphi^m_{e_l} = \sin^2\frac{\pi}{n} \left(-\frac{(p_h^m)_{ll} - \rho_h (V_h)_l}{n^3 h_v^5} + \frac{n^2+2}{12}\rho_h + \frac{1}{2}\varphi^m_{n/2\,e_l} \right. \tag{58}$$

$$\left. + \sum_{m=2}^{n/2-1} (-1)^m \sin^{-2}\left(\frac{\pi}{n} m\right) \mathrm{Re}\varphi^m_{m\,e_l} \right), \tag{59}$$

$$\varphi_{-e_l} = \overline{\varphi_{e_l}}, \ l = 1,2,3.$$

Thus, the formulae (57),(58) and (60) allow the components $\varphi^m_{e_l}$, $l = 1,2,3$ to be defined such that all numerical moments of the distribution function are conserved during the computation. The imaginary parts of these components are prescribed (this conserves the bulk velocity) as well as the sum s of the real parts (this conserves the energy, cf. (43)). In order to find the concrete values r_l, $l = 1,2,3$ of the real parts for the given $\tilde{r}_l$, $l = 1,2,3$ we minimise the norm of the difference $\|r - \tilde{r}\|_2^2$ under the condition $(r,e) = s$, $e = (1,1,1)^T$. Simple computation yields the following correction

$$r = \tilde{r} - \frac{1}{3}((\tilde{r},e) - s)e.$$

What is remarkable is the very low computational work required by these formulae: it is of the capital order $O(n)$ because only the knots placed on the axes are involved.

5 Numerical examples

In this section we calculate an example of the relaxation using our difference scheme. The initial distribution $f_0(v)$ is given by

$$f_0(v) = \frac{1}{2(2\pi)^{3/2}} \left(\exp\left(-\frac{|v-V_1|^2}{2} \right) + \exp\left(-\frac{|v-V_2|^2}{2} \right) \right),$$

where

$$V_1 = (2,2,0)^T,\ V_1 = (-2,2,0)^T.$$

The initial and asymptotic values of nontrivial moments of the distribution function $f(v,t)$ and the conserved macroscopic quantities for this example are

$$\begin{aligned}
\rho &= 1,\ V = (0,1,0)^T,\ T = 8/3, \\
p_{11}(0) &= 5,\ p_{11}(\infty) = 8/3, \\
p_{12}(0) &= 2,\ p_{12}(\infty) = 0, \\
p_{22}(0) &= 3,\ p_{22}(\infty) = 11/3, \\
p_{33}(0) &= 1,\ p_{33}(\infty) = 8/3, \\
q_1(0) &= 4,\ q_1(\infty) = 0, \\
q_2(0) &= 13,\ q_2(\infty) = 43/3,
\end{aligned}$$

where

$$\begin{aligned}
p_{i,j}(t) &= \int_{\mathbb{R}^3} v_i v_j f(v,t) dv, \\
q_i(t) &= \int_{\mathbb{R}^3} v_i |v|^2 f(v,t) dv,\ i,j = 1,2,3. \qquad (60)
\end{aligned}$$

In order to obtain the reference solution of this problem we use the standard DSMC method [1] with 10 000 particles and generate $N_{rep} = 10$, 100 and 1000 independent trajectories. The next figures display the confidence bands for some of the above moments for $N_{rep} = 10$ as well as the numerical solution obtained using our scheme for $L = 10$, $R = 7$, $n = 16$, $n_\phi = 16$, $n_{mu} = 4$ and $n_\zeta = 4$.

In **Figure 1** it is clear to see that just 10 independent trajectories of the DSMC already lead to sufficient accuracy in the computation of the time evolution of the second moment. The computational time of the DSMC method is only about 2% of our method.

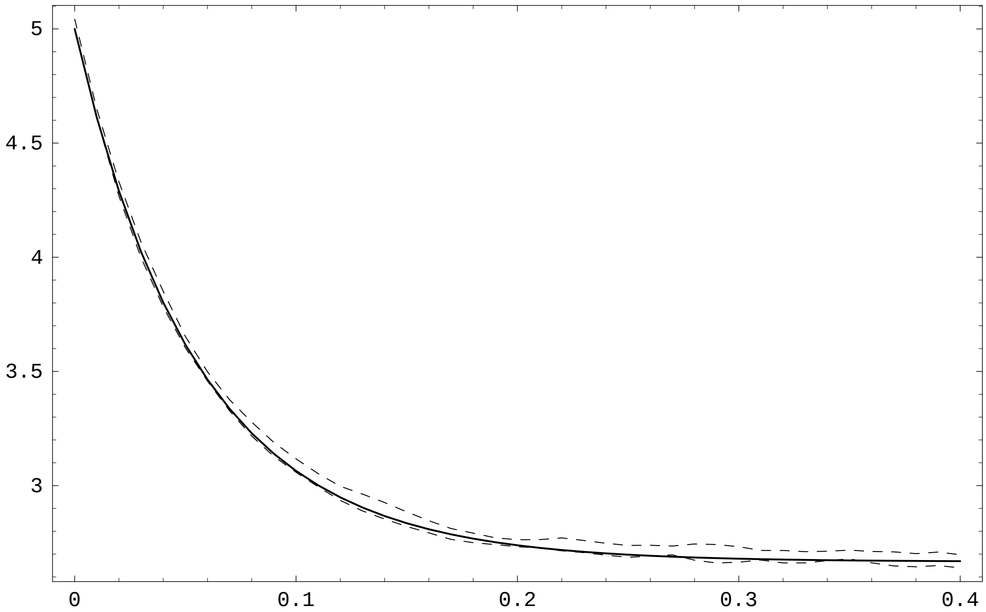

Figure 1 Relaxation of p_{11} and the confidence bands for $N_{rep} = 10$.

However, the situation changes if we are interested in computing the time evolution of the third moment, as shown in **Figure 2**. The width of the confidence bands indicates that the accuracy of the computation using 10 independent trajectories is rather low in this case. There is only the possibility to increase the accuracy using more independent trajectories. The increase of this value to 100 and finally to 1000 leads to better results which are presented in **Figure 3**.

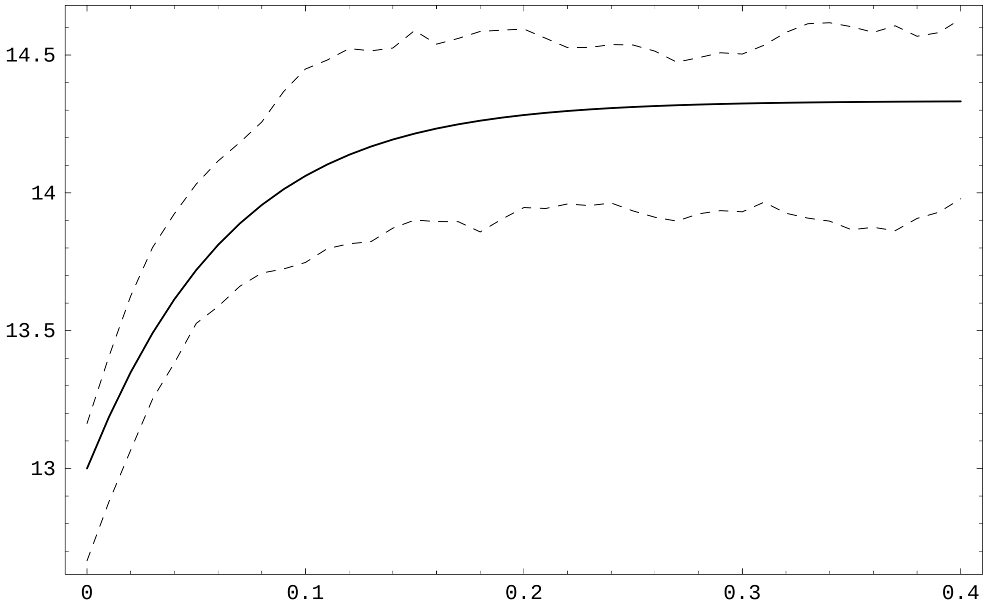

Figure 2 Relaxation of q_2 and the confidence bands for $N_{rep} = 10$.

This figure shows the empirical means for $10, 100$ and 1000 trajectories of DSMC (thin lines) as well as the curve obtained by our method (thick line). There is very close agreement in the results for $N_{rep} = 1000$. However, the computational time of the DSMC method is now twice that of our method.

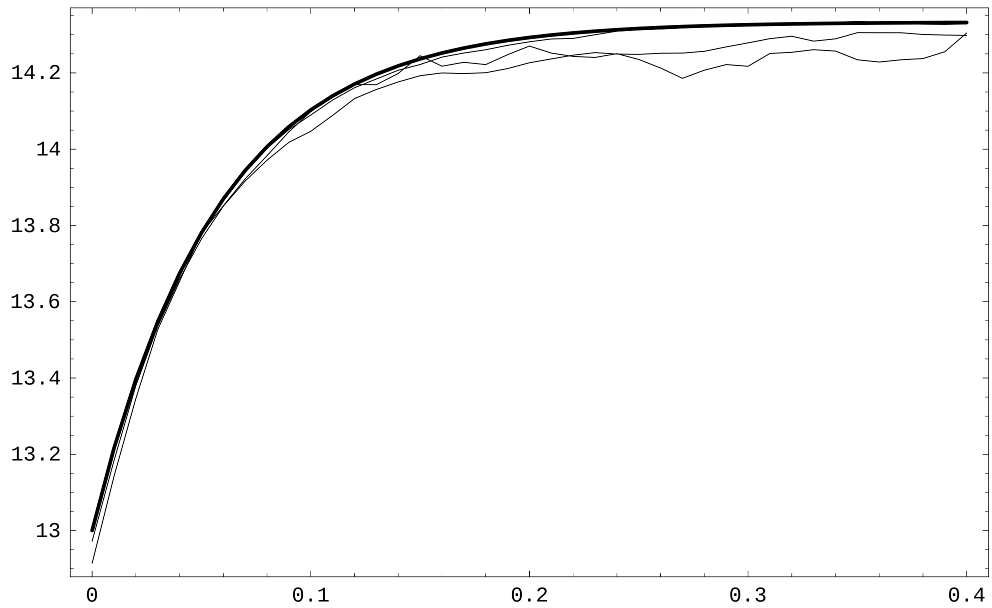

Figure 3 Curves for $N_{rep} = 10, 100, 1000$

It is also important to consider the memory requirements of the method. We need to store $2n^3$ components for the vectors f^m and q^m in (41). It is also useful to store all diagonal matrices (39),(40) ($2n^3\, n_\phi\, n_\mu\, n_\zeta$ components) in order to accelerate the computations. Note that these matrices can be used for all spatial cells if we are solving a spatially inhomogeneous problem. Some additional but rather small storage is required by FFT. Since the parameter n is at least 16 the memory requirements of the method presented are quite high. Especially for the spatially inhomogeneous problems this can lead to serious problems if the variation of the macroscopic quantities in the physical space becomes large. It is clear that the DSMC method is almost free of these difficulties.

6 Acknowledgement

The authors would like to thank the German Academic Exchange Service (DAAD) for providing a visiting professorship at the University of Saarland to the first author during the academic year 1997-1998.

References

[1] G. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows.* Clarendon Press, Oxford, 1994.

[2] A. Bobylev. The Fourier transform method in the theory of the Boltzmann equation for Maxwell molecules. *Doklady Akad. Nauk SSSR*, 225 : 1041–1044, 1975.

[3] A. Bobylev. Expansion of the Boltzmann collision integral in Landau series. *Sov. Phys. Dokl.*, 20 : 740–742, 1976.

[4] A. Bobylev and S. Rjasanow. Difference scheme for the Boltzmann equation based on the Fast Fourier Transform. *Eur. J. Mech. B/Fluids*, 16(2) : 293–306, 1997.

[5] A. Bobylev and V. Vedenjapin. The maximum principle for discrete models of the Boltzmann equation, and the connection between the integrals of direct and inverse collisions of the Boltzmann equation. *Doklady Akad. Nauk SSSR*, 233 : 519–522, 1977.

[6] T. Carleman. *Problèmes Mathematiques dans la Théorie Cinétique des Gas.* Almqvist & Wiksell, Uppsala, 1957.

[7] C. Cercignani, R. Illner, and M. Pulvirenti. *The Mathematical Theory of Dilute Gases.* Springer, New York, 1994.

[8] J. Cooley and J. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19 : 297–301, 1965.

[9] C. V. Loan. *Computational frameworks for the Fast Fourier Transform.* SIAM, Philadelphia, 1992.

[10] H. Neunzert, F. Groppengiesser, and J. Struckmeier. Computational methods for the Boltzmann equation. In R. Spigler, editor, *Applied and Industrial Mathematics*, pages 111–140. Kluwer Acad. Publ., Dordrecht, 1991.

[11] L. Pareschi and B. Perthame. A Fourier spectral method for homogeneous Boltzmann equations. *Transport Theory Statist. Phys.*, 25 : 369–382, 1996.