Riv. Mat. Univ. Parma (7) **4**** (2005), 145-250

LORENZO PARESCHI and GIOVANNI RUSSO (*)

An introduction to the numerical analysis of the Boltzmann equation (**)

Contents

(*) L. Pareschi: Department of Mathematics, University of Ferrara, Via Machiavelli 35, I-44100 Ferrara, Italy, e-mail: pareschi@dm.unife.it; G. Russo: Department of Mathematics and Computer Science, University of Catania, Via A. Doria 6, 95125 Catania, Italy, e-mail: russo@dmi.unict.it

(**) Received 13th July 2005 and in revised form 3rd October 2005. AMS classification 65C05, 65L60, 65R20, 76P05, 82C40. This work was partially supported by the European network HYKE, funded by the EC as contract HPRN-CT-2002-00282, by the INDAM project «Dissipative Kinetic Equations and their Fluid Dynamic Equations: Modelling and Numerical Methods» and by the project NUMSTAT 2005, Comitato dei Sostenitori, funded by the University of Ferrara.

Introduction

The main subject of this work concerns the numerical solution of nonlinear kinetic equation of the form

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \frac{1}{\varepsilon} Q(f,f), \quad x \in \mathbb{R}^{d_x}, \quad v \in \mathbb{R}^{d_v}, \quad d_x, d_v = 1,2,3, \tag{1}$$

where $f = f(x,v,t)$ is a nonnegative function, $\varepsilon > 0$ is a parameter related to the mean free path between particles and $Q(f,f)$ is a (nonlinear) operator. The most well-known example is represented by the Boltzmann equation of rarefied gas dynamics [**30**], [**31**]. Besides other classical examples, like the Landau equation of plasma physics [**77**], kinetic equations play an important role in modelling granular gases [**12**], charged particles in semiconductors [**84**], neutron transport [**71**] and quantum gases [**3**], [**83**]. More recently applications of kinetic equations have been considered for car traffic flows [**74**], chemotactical movements [**34**], tumor immune cells competition [**8**], coagulation-fragmentation processes [**44**], population dynamics [**41**], market economies [**36**], supply chains [**2**] and many other. For a recent introduction to the Boltzmann equation and related kinetic equations we refer the reader to [**39**], [**129**].

Although the scope of our insights is wider, in this paper we will focus mainly on the classical Boltzmann equation. This is motivated not only by its relevance for applications but also because it contains all major difficulties present in other kinetic models. In other words it represents the most challenging case for the development of numerical schemes.

Approximate methods of solution for the Boltzmann equation have a long history tracing back to D. Hilbert, S. Chapmann and D. Enskog [**30**] at the beginning of the last century. The mathematical difficulties related to the Boltzmann equation make it extremely difficult, if not impossible, the determination of analytic solutions in most physically relevant situations. Only in recent years, starting in the 70s with the pioneering works by A. Chorin [**35**] and G. Sod [**122**], the problem has been tackled numerically with particular care to accuracy and computational cost. Even nowadays the deterministic numerical solution of the Boltzmann equation still represents a challenge for scientific computing.

Most of the difficulties are due to the multidimensional structure of the collisional integral, since the integration runs on a highly-dimensional unflat manifold. In addition the numerical integration requires great care since the collision integral is at the basis of the macroscopic properties of the equation. Further difficulties are represented by the presence of stiffness, like the case of small mean free path [**51**] or the case of large velocities [**46**].

For such reasons realistic numerical simulations are based on Monte-Carlo techniques. The most famous examples are the Direct Simulation Monte-Carlo (DSMC) methods by Bird [**9**] and by Nanbu [**92**]. These methods guarantee efficiency and preservation of the main physical properties. However, avoiding statistical fluctuations in the results becomes extremely expensive in presence of non-stationary flows or close to continuum regimes.

In some recently developed Monte Carlo methods, a different approach is taken with the goal of constructing simple and efficient numerical methods for the solution of the Boltzmann equation in regions with a large variation in the mean free path [**104**], [**105**], [**112**], [**114**], [**100**]. These algorithms are based on a suitable time discretization of the Boltzmann equation, first introduced in [**51**].

Among deterministic approximations, perhaps the most popular method is represented by the so called Discrete Velocity Models (DVM) of the Boltzmann equation. These methods [**85**], [**121**], [**13**], [**21**], [**97**] are based on a cartesian grid in velocity and on a discrete collision mechanism on the points of the grid that preserves the main physical properties. Unfortunately DVM are not competitive with Monte Carlo methods in terms of computational cost (typically $O(n^{(2d_v+1)/d_v})$, where n is the total number of discretization parameters in velocity) and their accuracy seems to be at most first order in velocity [**95**], [**96**], [**97**]. Fast algorithms for DVM models have been recently proposed in [**90**].

Another important class of numerical methods is based on the use of spectral techniques in the velocity space. The methods were first derived in [**102**], inspired by previous works on the use of Fourier transform techniques (see [**11**] for instance). The numerical method is based on approximating the distribution function by a periodic function in velocity space, and on its representation by Fourier series. The resulting scheme can be evaluated with a computational cost of $O(n^2)$, which is lower than that of DVM.

The method was further developed in [**106**], [**107**] where evolution equations for the Fourier modes were explicitly derived and spectral accuracy of the method was proved. Strictly speaking these methods are not conservative, since they preserve mass, whereas momentum and energy are approximated with spectral accuracy. This trade off between accuracy and conservations seems to be an unavoidable compromise in the development of numerical schemes for the Boltzmann equation.

Recently in [**89**], [**88**], [**45**], using a suitable representation of the collision operator, the computational cost of spectral methods has been reduced from $O(n^2)$ to $O(n \log_2 n)$ without loosing the spectral accuracy thus making the methods competitive with Monte Carlo. These fast algorithms are restricted to a certain class of particle interactions including pseudo-Maxwell molecules (for $d_v = 2$) and hard spheres (for $d_v = 3$).

We recall here that the spectral method has been applied also to non homogeneous situations [**48**], to the Landau equation [**46**], [**109**], where fast algorithms can be readily derived, and to the case of granular gases [**91**], [**47**]. Finally let us mention that A. Bobylev and S. Rjasanow [**14**], [**15**] have also constructed fast algorithms based on a Fourier transform approximation of the distribution function.

Numerical concerns

We can summarize the main different numerical difficulties and requirements specific to the approximation of kinetic equations as follows

- Physical conservation properties, positivity and entropy inequality are very important since they characterize the steady states. Methods that do not maintain such properties need special attention in practical applications.
- The operator $Q(f,f)$ may contain an highly dimensional integral in velocity space. In such cases fast solvers are essential to avoid excessive computational cost. Otherwise full realistic 3D-3D simulations would be impossible even with today faster computers.
- The significant velocity range may vary strongly with space position (steady states may not be compactly supported in velocity space). Thus methods that use a finite velocity range require a great care and may be inadequate in some circumstances.
- Stiffness of the problem for small mean free paths and/or large velocities. Stiff solvers for small mean free path problems may be hard to use when we have to invert a large nonlinear system. Operator splitting may loose accuracy in such cases.
- Schemes must be capable to deal with boundary conditions in complicated geometries and with shocks without introducing excessive numerical dissipation.

In this work we review some of the main results in this field both from a deterministic as well as from a probabilistic viewpoint. In both settings after a presentation of classical methods like discrete velocity modelling and direct simulation Monte Carlo, the emphasis is addressed to spectral methods and time relaxed Monte Carlo methods. Besides the algorithmic aspects and the efficiency of the methods, considerations on stability, accuracy and consistency of the various schemes are reported. Among other methods, we will not discuss in this paper finite differences methods [**94**], [**124**], stochastic weighted particle methods [**118**], [**119**], majorant frequency schemes [**67**] and hybrid methods [**99**], [**100**].

PART I
Deterministic methods

1 - The Boltzmann equation

1.1 - *The model*

The model is characterized by a density function $f(x,v,t)$ describing the time evolution of a monoatomic rarefied gas of particles which move with velocity $v \in \mathbb{R}^3$ in the position $x \in \Omega \subset \mathbb{R}^3$ at time $t > 0$ which satisfies the Boltzmann equation [**30**], [**31**]

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \frac{1}{\varepsilon} Q(f,f). \tag{2}$$

The parameter $\varepsilon > 0$ is called *Knudsen number* and is proportional to the mean free path between collisions. The bilinear collision operator $Q(f,f)$ which describes the binary collisions of the particles acts over the velocity variable only

$$Q(f,f)(v) = \int_{\mathbb{R}^3} \int_{\mathbb{S}^2} B(v, v_*, \omega)[f(v')f(v'_*) - f(v)f(v_*)]\, d\omega\, dv_*. \tag{3}$$

In the above expression, ω is a unit vector of the sphere $\mathbb{S}^2$ and (v', v'_*) represent the collisional velocities associated with (v, v_*). The collisional velocities satisfy microscopic momentum and energy conservation

$$v' + v'_* = v + v_*, \quad |v'|^2 + |v'_*|^2 = |v|^2 + |v_*|^2. \tag{4}$$

The above system of algebraic equations has the following parametrized solution

$$v' = \frac{1}{2}(v + v_* + |v - v_*|\omega), \quad v'_* = \frac{1}{2}(v + v_* - |v - v_*|\omega) \tag{5}$$

where $v - v_*$ is the relative velocity.

The collision kernel $B(v, v_*, \omega)$ is a nonnegative function which characterizes the details of the binary interactions and depends only on $|v - v_*|$ and the scattering angle θ between relative velocities $v - v_*$ and $v' - v'_* = |v - v_*|\omega$

$$cos\theta = \frac{(v - v_*) \cdot \omega}{|v - v_*|}.$$

The kernel has the from

$$B(v, v_*, \omega) = |v - v_*|\sigma(|v - v_*|, \cos\theta), \tag{6}$$

where the function σ is the *scattering cross-section*.

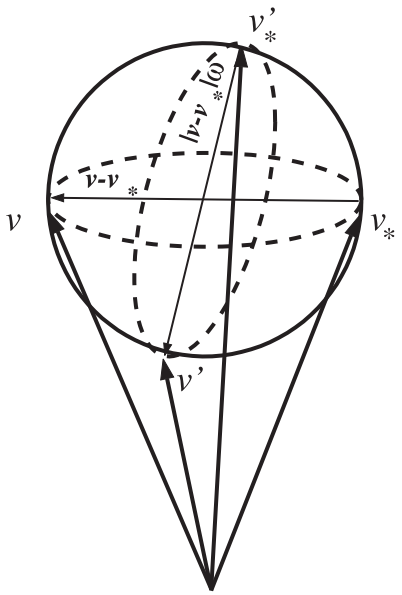

Fig. 1. – The collision sphere

Example 1.

• *In the* hard sphere model *the particles are assumed to be ideally elastic spheres of diameter* $d > 0$ *and thus*

$$\sigma(|v - v_*|, \cos\theta) = \frac{d^2}{4}, \quad B(v, v_*, \omega) = \frac{d^2}{4}|v - v_*|, \tag{7}$$

since the total cross section is $\pi d^2 = 4\pi(d^2/4)$.

• *In the case of inverse k-th power forces between particles, the kernel has the form*

$$\sigma(|v - v_*|, \cos\theta) = b_a(\cos\theta)|v - v_*|^{a-1}, \quad B(v, v_*, \omega) = b_a(\cos\theta)|v - v_*|^a, \tag{8}$$

with $a = (k-5)/(k-1)$. *For* $k > 5$ *we have hard potentials, for* $k < 5$ *we have* soft potentials.

• *The special situation* $k = 5$ *gives the* Maxellian model *with*

$$B(v, v_*, \omega) = b_0(\cos\theta). \tag{9}$$

• *For numerical purposes, a widely used model is the* Variable Hard Sphere (*VHS*) *model, corresponding to* $b_a(\cos\theta) = C_a$, *where* C_a *is a positive constant, and hence*

$$\sigma(|v - v_*|, \cos\theta) = C_a|v - v_*|^{a-1}, \quad B(v, v_*, \omega) = C_a|v - v_*|^a. \tag{10}$$

The collision integral $Q(f,f)$ can be written in different equivalent forms, according to the parametrization used for the collisional velocities. Using the identity

$$\int_{\mathbb{S}^2} (u\cdot n)+\phi(n(u\cdot n))\,dn = \frac{|u|}{4}\int_{\mathbb{S}^2}\phi\left(\frac{u-|u|\omega}{2}\right)d\omega \tag{11}$$

obtained by the transformation $\omega = e - 2(e\cdot n)\,,n$, we get the frequently used form

$$Q(f,f)(v) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} \tilde{B}(v,v_*,\omega)[f(v')\,f(v'_*) - f(v)\,f(v_*)]\,d\omega\,dv_* \tag{12}$$

with

$$v' = v - ((v-v_*)\cdot\omega)\omega, \quad v'_* = v_* + ((v-v_*)\cdot\omega)\omega, \tag{13}$$

and

$$\tilde{B}(v,v_*,\omega) = 2|v-v_*||\cos\theta|\sigma(|v-v_*|,\ 1-2|\cos\theta|). \tag{14}$$

The hard sphere case corresponds to

$$\tilde{B}(v,v_*,\omega) = \frac{d^2}{2}|v-v_*||\cos\theta|, \tag{15}$$

whereas the Maxwellian molecules case gives

$$\tilde{B}(v,v_*,\omega) = 2|\cos\theta|\,b_0(\cos\theta). \tag{16}$$

Remark 1. *For the Maxwellian case the collision kernel $B(v,v_*,\omega)$ is independent of the relative velocity. This case has been widely studied theoretically, in particular exact analytic solutions can be found in the space homogeneous case where $f = f(v,t)$* [**10**].

A simplified one-dimensional space homogeneous Maxwell model is given by the Kac equation [**72**]. *It reads*

$$\frac{\partial f}{\partial t} = \int_{\mathbb{R}}\int_0^{2\pi} \frac{1}{2\pi}[f(v'_*)\,f(v') - f(v)\,f(v_*)]\,d\theta\,dv_* \tag{17}$$

where the collisional velocities are characterized by rotations in the collisional plane

$$v'_* = v\cos\theta - v_*\sin\theta, \quad v'_* = v\sin\theta + v_*\cos\theta. \tag{18}$$

For this model we have only microscopic conservation of energy $(v')^2 + (v'_*)^2 = v^2 + v_*^2$.

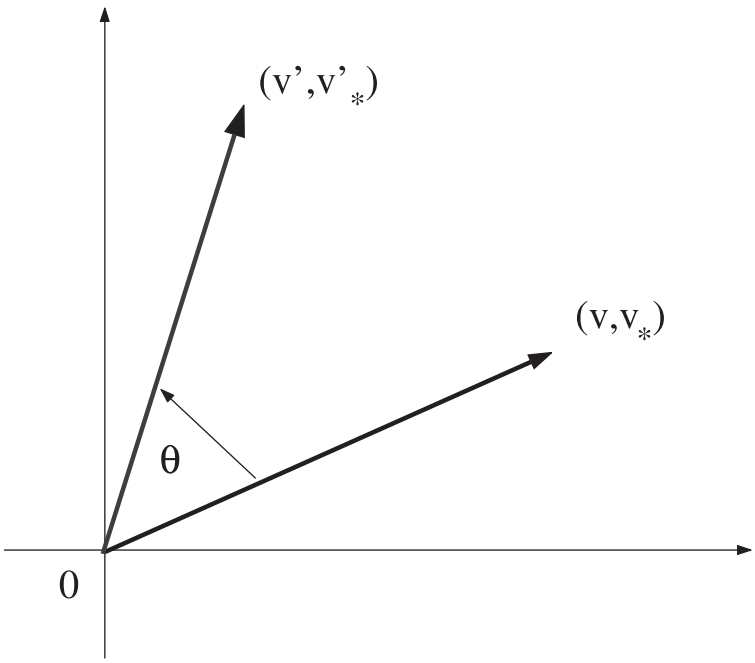

Fig. 2. – Collisions in Kac's model

1.2 - *Physical properties*

During the evolution process, the collision operator preserves mass, momentum and energy, i.e.,

$$\int_{\mathbb{R}^3} Q(f,f)\,\phi(v)\,dv = 0, \quad \phi(v) = 1, v^x, v^y, v^z, |v|^2,$$

and in addition it satisfies Boltzmann's well-known H-theorem

$$\int_{\mathbb{R}^3} Q(f,f)\ln(f(v))\,dv \leq 0.$$

The above properties are a consequence of the following identity that can be easily proved for any test function $\phi(v)$

$$\int_{\mathbb{R}^3} Q(f,f)\,\phi(v)\,dv = -\frac{1}{4}\int_{\mathbb{R}^6}\int_{\mathbb{S}^2} B(v,v_*,\omega)\,[f'f'_* - ff_*][\phi' + \phi'_* - \phi - \phi_*]\,d\omega\,dv_*\,dv$$

where we have omitted the explicit dependence from v, v_*, v', v'_* to simplify the expression.

In order to prove this identity we used the micro-reversibility property $B(v, v_*, \omega) = B(v_*, v, \omega)$ and the fact that the Jacobian of the transformation $(v, v_*) \leftrightarrow (v', v'_*)$ is equal to 1.

A function ϕ such that

$$\phi(v') + \phi(v'_*) - \phi(v) - \phi(v_*) = 0$$

is called a *collision invariant*. It can be shown that a continuous function ϕ is a collision invariant if and only if $\phi \in \operatorname{span}\{1, v, |v|^2\}$ or equivalentely

$$\phi(v) = a + b \cdot v + c|v|^2, \quad a, c \in \mathbb{R}, \quad b \in \mathbb{R}^3.$$

Assuming f strictly positive, for $\phi(v) = \ln(f(v))$ we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^3} Q(f,f) \ln(f)\, dv \\
&= -\frac{1}{4} \int_{\mathbb{R}^6} \int_{\mathbb{S}^2} B(v, v_*, \omega) [f'f'_* - ff_*] [\ln(f') + \ln(f'_*) - \ln(f) - \ln(f_*)]\, d\omega\, dv_*\, dv \\
&= -\frac{1}{4} \int_{\mathbb{R}^6} \int_{\mathbb{S}^2} B(v, v_*, \omega) [f'f'_* - ff_*] \ln\left(\frac{f'f'_*}{ff_*}\right) d\omega\, dv_*\, dv \le 0,
\end{aligned}$$

since the function $z(x, y) = (x - y)\ln(x/y) \ge 0$ and $z(x, y) = 0$ only if $x = y$.

In particular, the equality holds only if $\ln(f)$ is a collision invariant, that is

$$f = \exp(a + b \cdot v + c|v|^2), \quad c < 0.$$

If we define the density, mean velocity and temperature of the gas by

$$\rho = \int_{\mathbb{R}^3} f\, dv, \qquad u = \frac{1}{\rho} \int_{\mathbb{R}^3} vf\, dv, \qquad T = \frac{1}{3R\rho} \int_{\mathbb{R}^3} [v - u]^2 f\, dv,$$

we obtain that the distribution function has the form of a locally *Maxwellian distribution*

$$f(v, t) = M(\rho, u, T)(v, t) = \frac{\rho}{(2\pi RT)^{3/2}} \exp\left(-\frac{|u - v|^2}{2RT}\right).$$

The constant $R = K_B/m$ is called the gas constant, K_B is the Boltzmann constant and m the mass of a particle. Boltzmann's H-theorem implies that any equilibrium distribution function, i.e. any function f for which $Q(f, f) = 0$, has the form of a locally Maxwellian distribution.

If we define the *H-function*

$$H(f) = \int_{\mathbb{R}^3} f \ln(f)\, dv,$$

we obtain immediately the inequality

$$\frac{dH(f)}{dt} = \int_{\mathbb{R}^3} Q(f,f) \ln(f)\, dv \leq 0.$$

Thus the H-function is monotonically decreasing until f reaches the equilibrium Maxwellian state for which we have

$$H(M) = \rho\left(\ln\left(\frac{\rho}{(2\pi RT)^{3/2}}\right) - \frac{3}{2}\right).$$

1.3 - *Fluid limit*

If we multiply the Boltzmann equation by its collision invariants and integrate the result in velocity space we obtain

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^3} f\phi(v)\, dv + \nabla_x\left(\int_{\mathbb{R}^3} vf\phi(v)\, dv\right) = 0, \quad \phi(v) = 1, v_1, v_2, v_3, |v|^2.$$

These equations describe the balance of mass, momentum and energy. The system of five equations is not closed since it involves higher order moments of the distribution function f.

As $\varepsilon \to 0$, from (2) we have formally $Q(f,f) \to 0$, and thus f approaches the local Maxwellian. In this case the higher order moments of the distribution function can be computed as function of ρ, u, and T and we obtain the closed system of *compressible Euler equations*

$$\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho u) = 0$$

$$\frac{\partial \rho u}{\partial t} + \nabla_x \cdot (\rho u \otimes u + p) = 0$$

$$\frac{\partial E}{\partial t} + \nabla_x \cdot (Eu + pu) = 0$$

$$p = \rho T, \quad E = \frac{3}{2}\rho T + \frac{1}{2}\rho u^2$$

where p is the gas pressure and $\otimes$ denotes the tensor product.

The rigorous passage from the Boltzmann equation to the compressible Euler equations has been investigated by several authors. Among them we mention references [**23**], [**93**]. Higher order fluid models such as the Navier-Stokes model, can be considered using the Chapmann-Enskog and the Hilbert expansion. We refer to [**80**] for a mathematical setting of the problem and to [**56**] for recent theoretical results.

1.4 - *Boundary conditions*

The Boltzmann equation is complemented with the boundary conditions in space for $v \cdot n \geq 0$ and $x \in \partial\Omega$, where n denotes the unit normal, pointing inside the domain Ω. Usually the boundary represents the surface of a solid object (an obstacle or a container). The particles of the gas that hit the surface interact with the atoms of the object and are reflected back into the domain Ω.

Mathematically, such boundary conditions are modelled by an expression of the form [**30**]

$$|v \cdot n|\, f(x,v,t) = \int\limits_{v_* \cdot n < 0} |v_* \cdot n(x)|\, K(v_* \to v, x, t)\, f(x, v_*, t)\, dv_*. \tag{19}$$

This is the so-called *reflective boundary condition* on $\partial\Omega$.

The ingoing flux is defined in terms of the outgoing flux modified by a given boundary kernel K. This boundary kernel is such that positivity and mass conservation at the boundaries are guaranteed,

$$K(v_* \to v, x, t) \geq 0, \qquad \int\limits_{v \cdot n(x) \geq 0} K(v_* \to v, x, t)\, dv = 1.$$

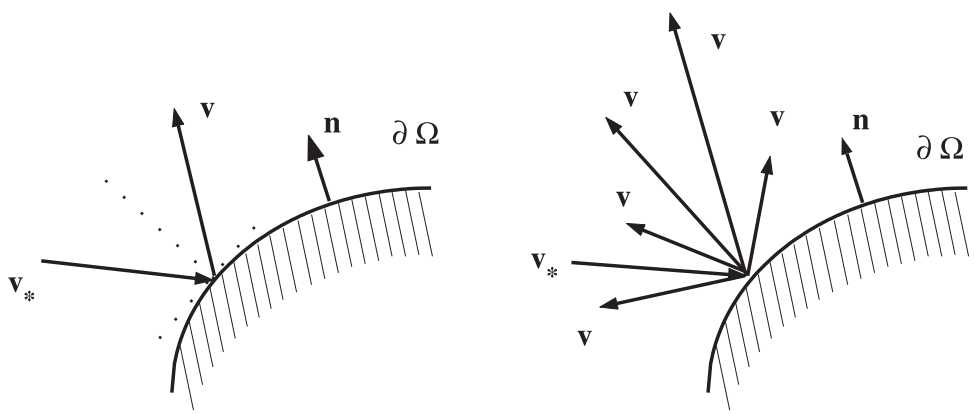

Fig. 3. – Reflection and diffusion at the solid boundary

Commonly used reflecting boundary conditions are the so-called Maxwell's conditions. From a physical point of view, one assumes that a fraction a of molecules is absorbed by the wall and then re-emitted with the velocities corresponding to those in a still gas at the temperature of the wall, while the remaining fraction $(1-a)$ is specularly reflected.

This is equivalent to impose for the ingoing velocities

$$f(x,v,t) = (1-a)\,Rf(x,v,t) + aMf(x,v,t), \tag{20}$$

in which $x \in \partial\Omega$, $v \cdot n(x) \geq 0$. The coefficient a, with $0 \leq a \leq 1$, is called the *accommodation coefficient* and

$$Rf(x,v,t) = f(x, v - 2n(n\cdot v), t), \quad Mf(x,v,t) = \mu(x,t)M_w(v). \tag{21}$$

If we denote by T_w the temperature of the solid boundary, M_w is given by

$$M_w(v) = \exp\left(-\frac{v^2}{2RT_w}\right),$$

and the value of μ is determined by mass conservation at the surface of the wall

$$\mu(x,t) \int\limits_{v\cdot n \geq 0} M_w(v)|v\cdot n|dv = \int\limits_{v\cdot n<0} f(x,v,t)|v\cdot n|dv. \tag{22}$$

For $a = 0$ (specular reflection) the re-emitted molecules have the same flow of mass, temperature and tangential momentum of the incoming molecules, while for $a = 1$ (full accommodation) the re-emitted molecules have completely lost memory of the incoming molecules, except for conservation of the number of molecules.

More complex boundary conditions for rarefied gas dynamics (RGD) can be imposed using the boundary conditions of Cercignani and Lampis [**32**]. These can be written as

$$f(x,v,t) = \int P(v,v')\,f(x,v',t)\,dv' \tag{23}$$

where

$$P(v,v') = (2v/a)\,I'(2(1-a)1/2vv'/a)\,\exp\,(v^2 - (1-a)v'^2)/a \tag{24}$$

in which v and v' are the normal components of the outgoing and incoming velocities respectively, and I' is the modified Bessel function. This satisfies the reciprocity (detailed balance) condition

$$vP(-v', -v)M_w(v) = -v'P(v,v')\,M_w(v'). \tag{25}$$

A consequence of the reciprocity condition is that the Maxwellian distribution M_w is preserved by this boundary condition.

In the case of *inflow boundary conditions*, one assumes that the distribution function of the particles entering the domain is known, i.e.

$$f(x,v,t) = g(v,t), \quad x \in \partial\Omega, \quad v \cdot n > 0.$$

A typical example of such condition is used in shock wave calculations, where one assumes that the distribution function at the boundary of the computational domain is a Maxwellian $M(v)$ and that the incoming flux of particles at the boundary is distributed according to the Maxwellian flux $(v \cdot n)\, M(v)$, $v \cdot n > 0$.

1.5 - *Variants*

1.5.1 - BGK models

A simplified model Boltzmann equation is the so-called BGK model introduced by Bhatnagar, Gross and Krook [**6**]. In this model the collision operator is replaced by a relaxation operator of the form

$$Q_{BGK}(f,f)(v) = M(f) - f, \tag{26}$$

where $M(f)$ is a Maxwellian with the same moments of f.

Conservation of mass, momentum and energy as well as Boltzmann H-theorem are readily satisfied. The equilibrium solutions are clearly Maxwellians

$$Q_{BGK}(f,f) = 0 \Leftrightarrow f = M(f).$$

Numerical computations, as well as the analytic theory, for such model are much simpler then for the full Boltzmann equation. This model has the advantage of describing the right fuid limit. But in the Chapman-Enskog expansion, the transport coefficients obtained at the Navier-Stokes level are not satisfactory. In particular, the Prandtl number P_r (the ratio between heat conductivity and viscosity) is equal to 1. For most gases, we have $P_r < 1$. In particular, the hard-sphere model for a monoatomic gas leads to a Prandtl number very close to $2/3$. The correct Prandtl number can be recovered using more sophisticated BGK models, as the velocity dependent collision frequency BGK models and the Ellipsoidal Statistical BGK (ES-BGK) models [**16**], [**62**].

1.5.2 - Landau models

The Landau model [**77**] is a common kinetic model in plasma physics characterized by the following collision operator

$$Q_L(f,f)(v) = \nabla_v \cdot \int_{\mathbb{R}^d} A(v-v_*)\,[\,f(v_*)\nabla_v f(v) - f(v)\nabla_{v_*} f(v_*)]\,dv_*$$

where $A(z) = \Psi(|z|)\Pi(z)$ is a $d \times d$ nonnegative symmetric matrix and $\Pi(z) = (\pi_{ij}(z))$ is the orthogonal projection upon the space orthogonal to z,

$$\pi_{ij}(z) = \left(\delta_{ij} - \frac{z_i z_j}{|z|^2}\right).$$

We have $\Psi(|z|) = \Lambda|z|^{\alpha+2}$ for inverse-power laws, with $\alpha \geq -3$ and $\Lambda > 0$. The case $\alpha = -3$ is the so-called Coulombian case, of primary importance for applications. In such case the Boltzmann collision operator has no meaning, due to the divergence of the integral, even for smooth functions (a cut-off angular approximation is then used and the Landau equation can be derived in the so called grazing collision limit [**129**]).

Since conservation of mass, momentum, and energy, as well as H-theorem for the entropy are satisfied, equilibrium states are Maxwellians.

1.5.3 - Other models

• Enskog model: takes into account the nonlocality of the interactions induced by the diameter of the interacting spheres (accurately describes the behavior of dense gases). The collision operator is delocalized in space (regularization effect).

• Quantum-Boltzmann models: the nonlinear interactions $f'f'_* - ff_*$ is replaced by

$$f'f'(1 \pm f)(1 \pm f_*) - ff_*(1 \pm f')(1 \pm f'_*).$$

The minus sign corresponds corresponds to fermions (such as alectrons), and the plus sign to bosons (such as photons). The collision operator are called Pauli operator and Bose-Einstein operator respectively.

• Semiconductor-Boltzmann models: the linear Boltzmann equation for semiconductor devices has the form

$$Q_S(f, M) = \int \sigma(v, v_*)\{M(v)f(v_*) - M(v_*)f(v)\}\, dv_*,$$

where M is the normalized equilibrium distribution (Maxwellian, Fermi-Dirac) at the temperature θ of the lattice. The function $\sigma(v, v_*)$ describes the interaction of carriers with phonons.

• Granular gas models: particles undergo inelastic collisions. Energy is dissipated by the model and the steady states are Dirac delta function centered in the mean velocity.

More recently kinetic modelling has been applied to new fields as vehicular traffic flows, biomathematics (chemotaxis, inhalation of sprays), finance (modelling income distributions), coagulation-fragmentation processes, supply chains, and so on (see [**2**], [**12**], [**34**], [**36**], [**44**], [**74**], [**83**], [**84**] and the references therein).

1.6 - *The splitting approach*

The most common approach to solve numerically the full Boltzmann equation is based on an operator splitting [**42**].

The solution in one time step Δt may be obtained by the sequence of two steps. First integrate the space homogeneous equation for all $x \in \Omega$,

$$\frac{\partial \tilde{f}}{\partial t} = \frac{1}{\varepsilon} Q(\tilde{f}, \tilde{f}),$$

$$\tilde{f}(x, v, 0) = f_0(x, v),$$

for a time step Δt (*collision step*) to obtain $\tilde{f} = C_{\Delta t}(f_0)$, and then the transport equation using the output of the previous step as initial condition,

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = 0,$$

$$f(x, v, 0) = \tilde{f}(x, v, \Delta t).$$

for a time step Δt (*transport step*) to get $f = T_{\Delta t}(\tilde{f}) = T_{\Delta t}(C_{\Delta t}(f_0))$.

After computing an approximation of the solution at time Δt, the process may be iterated to obtain the numerical solution at later times. Although this splitting scheme (simple splitting) described above is first order accurate in time it is very popular because it has several nice properties.

- The collision step acts only on v whereas the transport step acts on x. This makes the implementation of the resulting scheme simpler (it allows the use of any existing code designed to solve the free transport equation) and highly parallelizable.
- It makes simpler to design schemes which preserves the physical properties of the equation (conservations, positivity, H-theorem), since these properties essentially depends on the treatment of the collision step.

It is then clear, that after this splitting almost all the main numerical difficulties are contained in the collision step. The discretization of the resulting equations can be performed in a variety of ways (finite volume, finite difference, Monte Carlo methods and so on). The choice of the discretization mainly depends on the method that is used for the solution of the space homogeneous Boltzmann equation.

We point out that the order of accuracy of this simple splitting does not improve even if we solve with great accuracy both collision and convection steps. The accuracy

in time may be improved by a more sophisticated splitting [**125**]. For example Strang splitting is second order accurate, provided both steps are at least second order. It can be written as

$$f = C_{\Delta t/2}(T_{\Delta t}(C_{\Delta t/2}(f_0))),$$

or equivalently as

$$f = T_{\Delta t/2}(C_{\Delta t}(T_{\Delta t/2}(f_0))).$$

Remark 2. *If a scheme is able to treat the collision step in the simple splitting for vanishingly small values of* ε*, then the splitting scheme become a first order* kinetic scheme *for the underlying fluid dynamic limit. In fact the collision step becomes a projection towards the local Maxwellian* $C_{\Delta t}(f_0) = M(f_0)$ *which is then transported by the transport step* $f = T_{\Delta t}(M(f_0))$.

Unfortunately Strang splitting reduces its accuracy to first order in time in this regime. This problem seems to occur for all high order splitting methods developed in the present literature. This drawback is not present if one uses Implicit-Explicit Runge-Kutta methods of the type proposed in [**25**], [**68**], [**108**] *applied to the whole equation without splitting.*

2 - Discrete velocity models

2.1 - *Derivation*

Historically these methods were among the first deterministic methods for discretizing the Boltzmann equation in velocity space [**20**], [**54**]. The discretization is built starting from physical rather then numerical considerations. We assume the gas particles can attain only a finite set of velocities

$$V_N = \{v_1, v_2, v_3, \dots, v_N\}, \quad v_i \in \mathbb{R}^3.$$

Particles collide by simple elastic collisions. The collision $(v_i, v_j) \leftrightarrow (v_k, v_l)$ is *admissible* if $v_i, v_j, v_k, v_l \in V_N$ and preserves momentum and energy

$$v_i + v_j = v_k + v_l, \qquad |v_i|^2 + |v_j|^2 = |v_k|^2 + |v_l|^2. \tag{27}$$

The set of admissible output pairs (v_k, v_l) corresponding to a given input pair (v_i, v_j) will be denoted by C_{ij} and its cardinality by q_{ij}.

The discrete collision operator is obtained by computing first the *transition probabilities* a_{ij}^{kl} of the collision $(v_i, v_j) \leftrightarrow (v_k, v_l)$ which must satisfy the relations

$$a_{ij}^{kl} \geq 0, \qquad \sum_{k,l=1}^{N} a_{ij}^{kl} = 1, \qquad \forall i,j = 1, \dots, N.$$

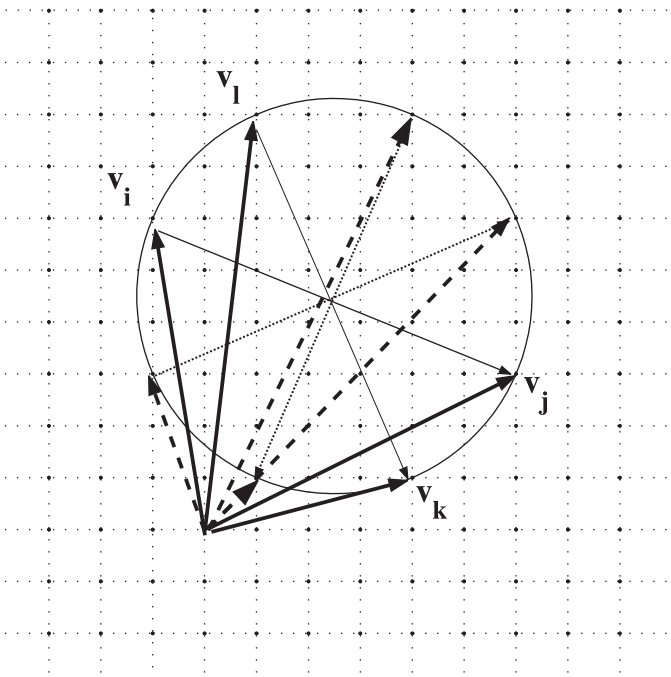

Fig. 4. – Sketch of a planar model based on a cartesian grid. For the collision (v_i, v_j) in the figure we have 3 admissible output collision pairs (v_k, v_l) hence $q_{ij} = 3$. Note that in general few grid points will belong to the collision circle

Example 2. *All output pairs are assumed to be equally probable*

$$a_{ij}^{kl} = \begin{cases} \dfrac{1}{q_{ij}} & \textit{if } (v_i, v_j) \leftrightarrow (v_k, v_l) \textit{ admissible} \\ 0 & \textit{if } (v_i, v_j) \leftrightarrow (v_k, v_l) \textit{ not admissible} \end{cases} .$$

Note that $a_{ij}^{kl} = a_{kl}^{ij}$ and $a_{ij}^{kl} = a_{ji}^{kl} = a_{ji}^{lk} = a_{ij}^{lk}$ (microreversibility). Next we introduce the transition rates $A_{ij}^{kl} = S|v_i - v_j| a_{ij}^{kl}$, where S is the cross sectional area of particles, and write the discrete Boltzmann equation as

$$\frac{\partial f_i}{\partial t} + v_i \cdot \nabla_x f_i = Q_i(f, f), \tag{28}$$

with

$$Q_i(f, f) = \sum_{j,k,l=1}^{N} A_{ij}^{kl} (f_k f_l - f_i f_j), \tag{29}$$

where f_i, f_j, f_k, f_l are the distribution densities of particles with velocities $v_i,\ v_j,\ v_k,\ v_l$.

2.2 - *Properties*

It is easy to check that the discrete Boltzmann equation satisfies for any test function $\phi_i = \phi(v_i)$

$$\sum_{i=1}^{N} Q_i(f,f)\phi_i = -\frac{1}{4}\sum_{i,j,k,l=1}^{N} A_{ij}^{kl}(f_k f_l - f_i f_j)(\phi_k + \phi_l - \phi_i - \phi_j), \tag{30}$$

and thus the *discrete collision invariants* satisfy

$$\phi_k + \phi_l - \phi_i - \phi_j = 0. \tag{31}$$

Clearly we have $\phi(v_i) = 1, v_i^x, v_i^y, v_i^z, |v_i|^2$ as collision invariants. Note that due to the finite number of velocities we can have models with additional spurious collision invariants or models where the collision invariants are not independents. Models with the correct space of collision invariants are called *normal*. It is remarkable that the normality of a model is determined only by the distribution of the non-zero coefficients in the matrix A_{ij}^{kl}, and does not depend on their values.

The main macroscopic quantities are defined as

$$\rho = \sum_{i=1}^{N} f_i, \qquad u = \frac{1}{\rho}\sum_{i=1}^{N} f_i v_i, \qquad T = \frac{1}{3R\rho}\sum_{i=1}^{N} f_i (v_i - u)^2. \tag{32}$$

In addition taking $\phi_i = \ln(f_i)$ we obtain

$$\sum_{i=1}^{N} Q_i(f,f)\ln(f_i) = -\frac{1}{4}\sum_{i,j,k,l=1}^{N} A_{ij}^{kl}(f_k f_l - f_i f_j)\ln\left(\frac{f_k f_l}{f_i f_j}\right) \leq 0, \tag{33}$$

and hence the discrete analogue of Boltzmann's H-theorem.

By the same arguments as in the continuous case we obtain that the equilibrium states are characterized by the equation

$$f_i = \exp(a + b\cdot v_i + c|v_i|^2), \quad c < 0.$$

However in the discrete case it is not possible to write explicitly a, b, c as functions of the macroscopic quantities ρ, u, T except in some particular cases.

In general one has to solve for a, b, c the system of nonlinear equations characterized by

$$\rho = \sum_{i=1}^{N} \exp(a + b\cdot v_i + c|v_i|^2) \tag{34}$$

$$u = \frac{1}{\rho}\sum_{i=1}^{N} \exp(a + b\cdot v_i + c|v_i|^2) v_i, \tag{35}$$

$$T = \frac{1}{3R\rho}\sum_{i=1}^{N} \exp(a + b\cdot v_i + c|v_i|^2)(v_i - u)^2. \tag{36}$$

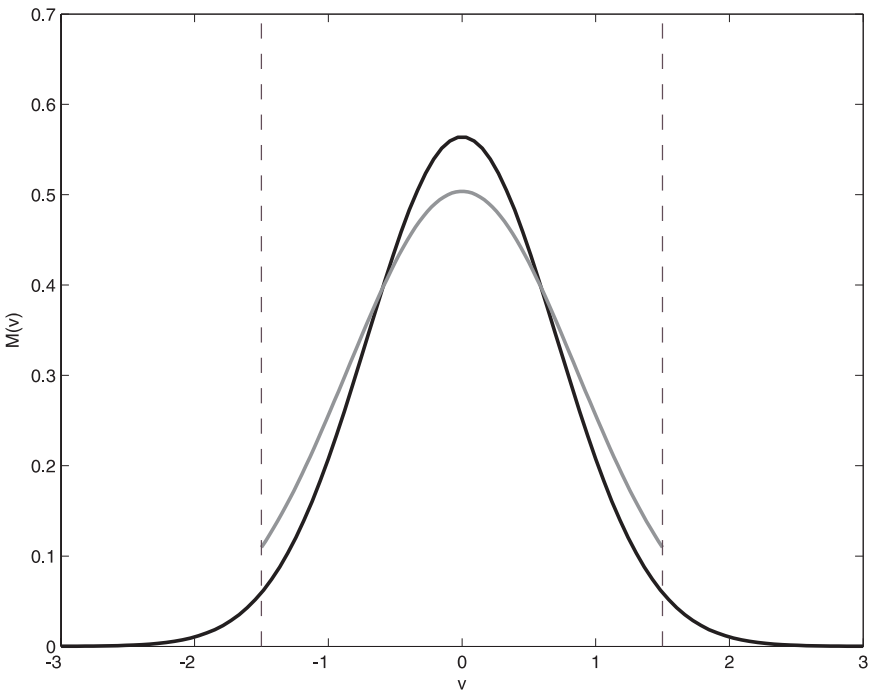

Fig. 5. – Example of discrete Maxwellian on a truncated velocity domain (dashed lines) and corresponding Maxwellian with same mass, momentum and energy

This requires a suitable numerical method if the discrete local Maxwellian equilibrium is needed explicitely.

Similarly we can formally write the corresponding conservation equations. Note however that, since discrete Maxwellian equilibrium states differ from Maxwellians of the Boltzmann equation, the corresponding fluid equations may be different from the classical ones (see Figure 5).

The same difficulties (necessity to solve a nonlinear system and velocity truncation effects) are present also in discrete BGK models of the Boltzmann equation.

2.3 - *Computational considerations*

Although the discrete Boltzmann equation is a discrete model of the Boltzmann equation that has the nice property of preserving the essential physical features (conservations, H-theorem, equilibrium states) from a computational point of view it presents some severe drawbacks.

- The computational cost is very high, typically $O(N^\eta)$, with $\eta > 2$.
- The accuracy of the method is poor, the error behaves as $O(1/n^\mu)$ with $\mu < 1$ and where n is the number of grid points in each direction (for a cartesian grid).

TABLE 1. – *Relative L_∞ error and computational cost (in seconds) for $n = 8$, 10, 12, 14, 16, 18, 20, 22*

n	8	10	12	14	16	18	20	22
E_∞	0.044	0.035	0.033	0.028	0.024	0.021	0.019	0.017
T_{sec}	0.38	1.96	7.25	23.01	66.43	135.47	286.94	638.13

In Tables 1 and 2 we report the results obtained in [**97**] where a careful numerical study of the convergence rate of DVM has been performed in the case of 3D Maxwell molecules. All computations were performed on Sun Dual UltraSparc 1700/167 MHz. Their method in dimension d is derived in a discrete-velocity space V subset of the regular cubic grid

$$\{v_i = hi | i = (i_1, \dots, i_d) \in \mathbb{Z}\},$$

where $h = 1/n$.

TABLE 2. – *Convergence rates μ and cost exponents η for $n = 8$, 10, 12, 14, 16, 18, 20, 22*

n	8-10	10-12	12-14	14-16	16-18	18-20	20-22
μ	1.0255	0.3227	1.0659	1.1544	1.1337	0.9499	1.1670
η	2.4506	2.3915	2.4974	2.6466	2.0167	2.3745	2.7953

The model is defined by

$$A_{ij}^{kl} = h^{2d-1} \frac{2|k-i| B_{k-i,l-i}}{g(k-i)} \chi_{ij}^{kl}, \quad k \neq i, \tag{37}$$

where $g(k-i)$ denotes the greatest common divisor of the components of the vector $k-i$ and χ_{ij}^{kl} is an indicator function of the set of points satisfying (27). In [**97**] the authors showed that the discrete-velocity model defined by (37) satisfies the discrete mass, momentum, and energy conservation relations and the entropy property. The model has no other collision invariants than linear combinations of 1, v_i and $|v_i|^2$.

The term $B_{k-i,l-i}$ depends on the collision kernel. For VHS molecules we have

$$B_{k-i,l-i} = 2C_a \frac{(v_{k-i}^2 + v_{l-i}^2)^{(a-1)/2}}{|v_{k-i}|}.$$

The method behaves essentially as a first order scheme $\mu \approx 1$ with a computational cost higher then quadratic $\eta \approx 2.45$. Theoretically the computational cost of DVM can be estimated using the Farey series and gives the value $\eta = (2d+1)/d$ [**83**]. In this case since $d = 3$ we have $\eta = 7/3$ which is slightly smaller then the above numerical estimate.

The general approach for proving convergence of discrete collision terms is to view them as multidimensional quadrature formulas for the integral. In fact, numerically the discrete Boltzmann equation corresponds to taking a quadrature formula for the Boltzmann equations where

$$V_N = \{v_1, v_2, v_3, \dots, v_N\} \quad \text{quadrature nodes}$$
$$a_{ij}^{kl} \quad \text{quadrature weights.}$$

Thus for a grid of regular hypercubes with equal weights, using product rectangular rules, we expect a first order convergence to the Boltzmann equation for hard spheres. Conservations assumptions on the weights impose $A_{ij}^{kl} = 0$ for non admissible interactions which can result in a deterioration of the accuracy of the method. We remark that the convergence of DVM to the full Boltzmann equation is a rather delicate task, since it involves the study of the distribution of integer points on spheres/planes in $\mathbb{R}^3$.

To this aim we recall the recent consistency result obtained in [**97**]

Theorem 1. *Let us define the kernel*

$$B^c(x,y) = \frac{4}{|x|\sqrt{x^2+y^2}}\sigma\left(\sqrt{x^2+y^2}, 1 - 2\frac{|x|}{\sqrt{x^2+y^2}}\right), \quad x = v' - v, \quad y = v'_* - v.$$

Assume that the above kernel satisfies $|x|^\beta B^c(x,y) \in C^m(P)$ *for some* $0 \le \beta < 3$ *and* $m \ge 1$. *Here* $P = \{(x,y) \in \mathbb{R}^3 \times \mathbb{R}^3 \,|\, x \cdot y = 0\}$ *with the metric* $(|x|^2 + |y|^2)^{1/2}$. *Take* f *and* $g \in C^m(\mathbb{R}^3)$, *with compact support. Then, for sufficiently small* h,

$$|Q(f,g)(v_i) - Q_i(f,g)| \le Ch^\mu \tag{38}$$

where

$$\mu = \min\left\{\frac{m}{m+3}, \beta, m, \frac{m(m+3-\beta)}{(m+3)(m+3-\beta)+\beta-m}\right\}$$

and the constant C *does not depend on* h *and* i.

It is interesting to note that in the case of hard spheres ($B^c(x,y) = C/|x|$) and C^∞-integrands the theorem states that $\mu = 1 - \varepsilon$ for any $\varepsilon > 0$.

As a consequence of the excessive computational requirements, applications of DVM have been mainly limited to space homogeneous problems or to simple models with few velocities in order to obtain qualitative results for space nonhomogeneous situations.

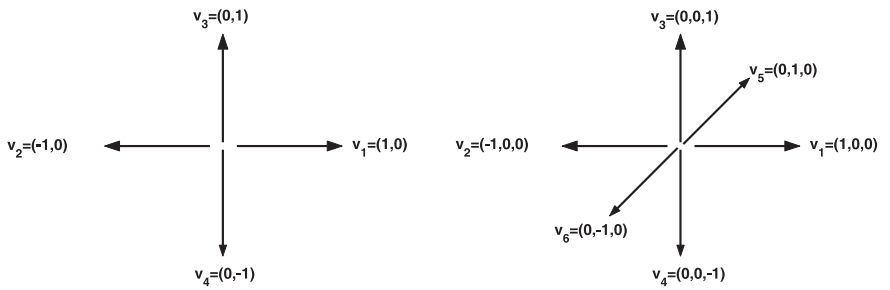

Fig. 6. – Velocities in Broadwell models

2.4 - *Broadwell models*

In this paragraph we consider a simple example of discrete velocity model. As we will see despite its simplicity the model can be used for a qualitative study of the formation of shock waves. More precisely we consider a set of one dimensional Broadwell models [**20**] defined by

$$\frac{\partial f_1}{\partial t} + \frac{\partial f_1}{\partial x} = \frac{1}{\varepsilon}(f_3^2 - f_1 f_2) \tag{39}$$

$$\frac{\partial f_2}{\partial t} - \frac{\partial f_2}{\partial x} = \frac{1}{\varepsilon}(f_3^2 - f_1 f_2) \tag{40}$$

$$\frac{\partial f_3}{\partial t} = \frac{1}{a\varepsilon}(f_1 f_2 - f_3^2) \tag{41}$$

being now f_1, f_2, f_3 the density of particles with velocity $1, -1, 0$ respectively. The parameter $a \geq 1$ is proportional to the number of particle densities moving with zero velocity. In particular for $a = 1$ we have the reduced four velocity Broadwell model, whereas for $a = 2$ it corresponds to the reduced six velocity Broadwell model.

The space of collision invariant has dimension 2 corresponding to conservation of mass and mean velocity

$$\rho = f_1 + f_2 + 2af_3, \quad \rho u = f_1 - f_2. \tag{42}$$

The Maxwellian state is characterized by two constants a and b such that

$$M_1 = a\exp\{b\}, \quad M_2 = a\exp\{-b\}, \quad M_3 = a.$$

In particular it is possible to get the analytic expression of a and b as a function of ρ and u

$$a = \frac{\rho(1-e)}{2a}, \quad b = \log\left[\frac{(1-e)}{a(e-u)}\right], \tag{43}$$

where $e = e(u)$ is given by

$$e = \frac{1}{2}(1+u^2) \qquad a = 1 \tag{44}$$

$$e = \frac{1}{3}\left(2\sqrt{3u^2+1} - 1\right), \quad a = 2. \tag{45}$$

2.4.1 - A numerical example

After the splitting the solution of the collision step consists of the set of ODEs

$$\frac{\partial f_1}{\partial t} = \frac{1}{\varepsilon}(f_3^2 - f_1 f_2) \tag{46}$$

$$\frac{\partial f_2}{\partial t} = \frac{1}{\varepsilon}(f_3^2 - f_1 f_2) \tag{47}$$

$$\frac{\partial f_3}{\partial t} = \frac{1}{a\varepsilon}(f_1 f_2 - f_3^2). \tag{48}$$

Due to the stiffness of the system for small values of ε an implicit scheme, such as the implicit Euler method, is needed. Here, since the system of equation can be solved analytically, we used directly the exact solution.

Next we need a numerical scheme for solving the transport phase for f_1 and f_2

$$\frac{\partial f_1}{\partial t} + \frac{\partial f_1}{\partial x} = 0 \tag{49}$$

$$\frac{\partial f_2}{\partial t} - \frac{\partial f_2}{\partial x} = 0 \tag{50}$$

$$\frac{\partial f_3}{\partial t} = 0. \tag{51}$$

In our numerical experiments we considered the first order upwind scheme with uniform mesh spacing Δx in the spatial grid points x_i given by

$$f_j^{n+1/2}(x_i) = f_j^n(x_i) + \eta\,[\,f_j^n(x_{i+i_j}) - f_j^n(x_i)], \quad j = 1, 2$$

and its second order TVD (total variation diminishing) extension [**79**]

$$\begin{aligned} f_j^{n+1/2}(x_i) = {} & f_j^n(x_i) + \eta[f_j^n(x_{i+i_j}) - f_j^n(x_i)] \\ & - i_j \frac{\eta(1-\eta)}{2}\,[\,F_j^n(x_{i+i_j})\Delta x - F_j^n(x_i)\Delta x], \quad j = 1, 2 \end{aligned}$$

where $\eta = \Delta t/\Delta x$, $i_j = (-1)^j$,

$$F_j(x_i) = \frac{[f_j(x_{i-j+2}) - f_j(x_{i-j+1})]}{\Delta x}\,\phi(\theta_j(x_i)), \quad \theta_j(x_i) = \left[\frac{f_j(x_i) - f_j(x_{i-1})}{f_j(x_{i+1}) - f_j(x_i)}\right]^{i_j},$$

and ϕ is the particular slope-limiter function. For example, the "superbee" limiter of Roe

$$\phi_{RS}(\theta) = \max\{0, \min\{1, 2\theta\}, \min(\theta, 2)\},$$

or Van Leer's limiter function

$$\phi_{VL}(\theta) = (|\theta| + \theta)/(1 + |\theta|).$$

The initial data is characterized by two local Maxwellian with mass and velocity

$$\underline{\rho},\ \underline{u}, \quad x \le 0$$

$$\overline{\rho},\ \overline{u}, \quad x > 0.$$

where the macroscopic quantities $\overline{\rho}, \overline{u}$ are computed in terms of $\underline{\rho}, \underline{u}$ from the classical Rankine-Hugoniot relations.

The test case we consider is a shock wave problem for $a = 2$ characterized by

$$\overline{\rho} = 4.0 \quad , \quad \overline{u} = 0 \quad ; \qquad \underline{\rho} = 1.0 \quad , \quad \underline{u} = 1.0.$$

In this situation, corresponding to a shock wave travelling with speed $s = 1/3$, the problem can be solved exactly

$$\rho(x,t) = \frac{4 + e^{\xi/\varepsilon}}{1 + e^{\xi/\varepsilon}}, \quad u(x,t) = \frac{e^{\xi/\varepsilon}}{1 + e^{\xi/\varepsilon}}, \tag{52}$$

where $\xi = [3x - t]/2$.

The results are reported in Figure 7 and show very good agreement with the analytic solution. In particular the Euler shock is well captured by the scheme.

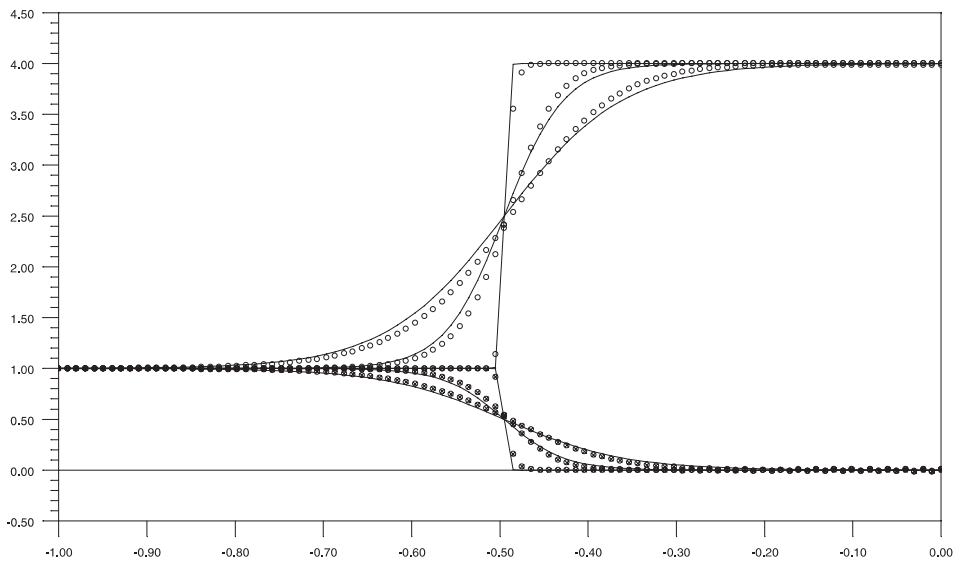

Fig. 7. – Profiles for different Knudsen numbers $\varepsilon = 0.1,\ 0.05,\ 10^{-6}$ of $\rho\,(\circ)$ and $u\,(\otimes)$ for the second order scheme with $\phi = \phi_{VL}$ and $\eta = 0.5$, $\Delta x = 0.01$ at $t = 1.5$

3 - Spectral Fourier methods

3.1 - *Preliminaries*

The use of spectral methods to develop fast algorithms for collisional integral equations was inspired by the theoretical results of A.V. Bobylev on Maxwell molecules. In [**10**] Bobylev showed that for such model a sort of convolution theorem holds, namely the Fourier transformed collision operator is defined through a simple integration over the sphere $\mathbb{S}^2$ instead of $\mathbb{R}^3 \times \mathbb{S}^2$. For this reason the first attempts aimed at deriving a similar simplification at a discrete level [**58**], [**50**]. The development of a general spectral method based on a Fourier-Galerkin approach has been subsequently done in [**102**], [**106**]. The presentation here follows [**98**], [**106**], [**48**].

After a splitting of the equation we focus our attention on the development of spectral methods for the space homogeneous collision step. Let us denote the collision operator as the difference of a gain and a loss part $Q(f,f) = Q^+(f,f) - fL[f]$, with

$$Q^+(f,f)(v) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B(v, v_*, \omega)\, f(v')\, f(v'_*)\, d\omega\, dv_*, \tag{53}$$

$$L[f](v) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B(v, v_*, \omega)\, f(v_*)\, d\omega\, dv_*. \tag{54}$$

In the sequel we will restrict to inverse power forces interactions and thus we will use the notation $B = B(|v - v_*|, \theta)$.

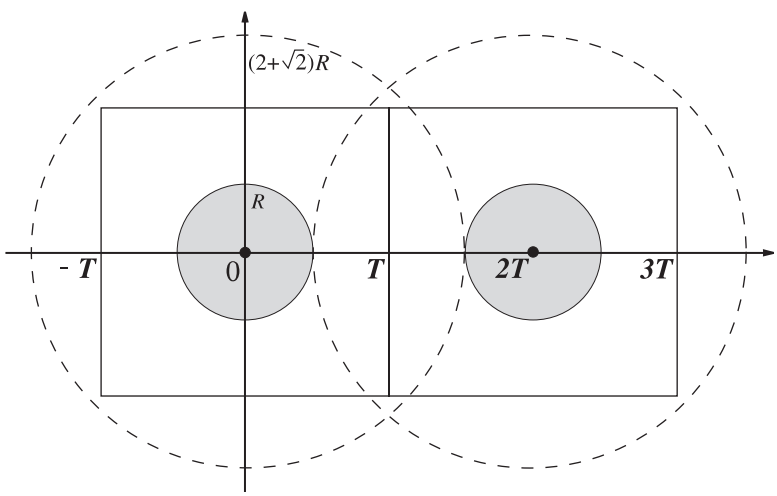

Fig. 8. – Periodization in 2D

3.2 - *Fourier-Galerkin projection*

We start from the homogeneous equation

$$\frac{\partial f}{\partial t} + f\,L(f) = Q^+(f,f)$$

with initial condition

$$f(x,v,t=0) = f_0(x,v).$$

The method can be described as the results of three main steps.

1. Change of variables ($v_* \to g = v - v_*$)

Integrating over $g = v - v_*$ gain and loss parts can be rewritten as

$$Q^+(f,f) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B(|g|,\theta)\, f(v')\, f(v'_*)\, d\omega\, dg, \tag{55}$$

$$L(f) = \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} B(|g|,\theta)\, f(v-g)\, d\omega\, dg, \tag{56}$$

where now

$$v' = v - \frac{1}{2}(g - |g|\omega), \quad v'_* = v - \frac{1}{2}(g + |g|\omega). \tag{57}$$

Note that what is crucial here is that we have rewritten the post-collisional velocities in the form $v' = v + \Theta(x,y)$ where x and y are the integration variables.

2. Reduction to a bounded domain

The following proposition characterizes the action of the collision operator on compactly supported functions. This is an essential step to understand the effect of aliasing in the final scheme. Let $\mathcal{B}_R$ denote the ball of radius R centered in the origin, then we have

Proposition 1. *Let* $\mathrm{Supp}\,(f(v)) \subset \mathcal{B}_R$ *then*

$$\mathrm{Supp}\,(Q(f,f)(v)) \subset \mathcal{B}_{\sqrt{2}R}, \tag{58}$$

and

$$Q(f,f)(v) = \int_{\mathcal{B}_{2R}}\int_{\mathbb{S}^2} B(|g|,\theta)[\,f(v')\,f(v'_*) - f(v)\,f(v-g)]\,d\omega\,dg, \tag{59}$$

with $v', v'_*, v-g \in \mathcal{B}_{(2+\sqrt{2})R}$.

The first statement follows from energy conservation since if $v', v'_* \in \mathcal{B}_R$ then $v^2 \le v^2 + (v-g)^2 = (v')^2 + (v'_*)^2 \le 2R^2$. Next we observe that if $v, v_* \in \mathcal{B}_R$ we have $|g| = |v - v_*| \le |v| + |v_*| \le 2R$. Finally $v \in \mathcal{B}_{\sqrt{2}R}$ and $g \in \mathcal{B}_{2R}$ imply $|v-g| \le |v| + |g|$ $\le (2+\sqrt{2})R$.

3. Fourier-Galerkin projection

We can consider the distribution function $f(v)$ restricted on $[-T,T]^3$ by assuming $f(v) = 0$ on $[-T,T]^3 \setminus \mathcal{B}_R$, and extend it by periodicity in the whole space $\mathbb{R}^3$. If the period T satisfies the following condition

$$T \ge R + \frac{(2+\sqrt{2})R - R}{2} = \frac{(3+\sqrt{2})}{2} R = R/\lambda, \tag{60}$$

then the value of $Q(f,f)$ remains unchanged. For simplicity we set $T = \pi$ and hence $R = \lambda\pi$.

Then we approximate f by the truncated Fourier series

$$f_N(v) = \sum_{k=-N}^{N} \hat{f}_k \, e^{ik\cdot v}, \tag{61}$$

$$\hat{f}_k = \frac{1}{(2\pi)^3} \int\limits_{[-\pi,\pi]^3} f(v)\, e^{-ik\cdot v}\, dv. \tag{62}$$

Here we use the vector notation $k = (k_1, k_2, k_3)$ and the compact expression

$$\sum_{k=-N}^{N} := \sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} \sum_{k_3=-N}^{N}.$$

The fundamental unknowns are the coefficients $\hat{f}_k$. To obtain the Fourier-Galerkin method we require that

$$\int\limits_{[-\pi,\pi]^3} \left(\frac{\partial f_N}{\partial t} + f_N\, L(f_N) - Q^+(f_N, f_N) \right) e^{-ik\cdot v}\, dv = 0.$$

By substituting expression for f_N in the Boltzmann equation we get

$$f_N\, L(f_N) = \sum_{l=-N}^{N} \sum_{m=-N}^{N} \hat{f}_l \hat{f}_m \hat{B}(m,m)\, e^{i(l+m)\cdot v},$$

and

$$Q^+(f_N, f_N) = \sum_{l=-N}^{N} \sum_{m=-N}^{N} \hat{f}_l \hat{f}_m \hat{B}(l,m)\, e^{i(l+m)\cdot v},$$

where the *kernel modes* $\hat{B}(l,m)$ are given by

$$\hat{B}(l,m) = \int_{\mathcal{B}_{2R}} \int_{\mathbb{S}^2} B(|g|,\theta)\, e^{-ig\cdot\frac{(l+m)}{2} - i|g|\omega\cdot\frac{(m-l)}{2}}\, d\omega\, dg. \tag{63}$$

The final scheme is charaterized by a set of ODE's for the Fourier coefficients

$$\frac{\partial \hat{f}_k}{\partial t} + \sum_{m=k-N}^{N} \hat{f}_{k-m}\, \hat{f}_m\, \hat{B}(m,m) = \sum_{m=k-N}^{N} \hat{f}_{k-m}\, \hat{f}_m\, \hat{B}(k-m,m). \tag{64}$$

with the initial condition

$$\hat{f}_k(0) = \frac{1}{(2\pi)^3} \int_{[-\pi,\pi]^3} f_0(v)\, e^{-ik\cdot v}\, dv.$$

Remark 3. *The following remarks clarify some aspects of the implementation and the cost of the method.*

• *Setting $\bar{N} = N^3$ the evaluation of the whole spectral scheme requires exactly $O(\bar{N}^2)$ operations. Thus the cost of the spectral method is smaller then the cost of DVM. In fact, for DVM we have a cost of $O(\bar{N}^{(2d+1)/d})$ where $d = 2, 3$ is the dimension of the velocity space* [**83**].

• *In the VHS model, the kernel does not depend on the angle θ: $B = C_a|g|^a$. In this case $\hat{B}(l,m)$ is a function of $|l-m|$, $|l+m|$ and thus can be precomputed and stored in a matrix. This can be done in $O(N^4)$ storage, by replacing the $N^3 \times N^3$ matrix $\hat{B}(l,m)$ by a smaller $N^2 \times N^2$ matrix $\hat{b}(i,j)$, where the integers i, j are given by (see* [**39**]*, chapter 4)*

$$0 \le i = |k|^2 \le 3N^2, \quad 0 \le j = |l-m|^2 < 12N^2.$$

• *The scheme preserves mass (Fourier mode of order zero) but not momentum and energy exactly. However, we will see that these quantities, except for the error on the initial data, are approximated with spectral accuracy in time.*

3.3 - *Analysis of the kernel modes*

Let us assume $B = C_a|g|^a$ (VHS model). Then we can reduce the computation of the kernel modes to a 1D integral. One has

$$\hat{B}(l,m) = C_a 16\pi^2 (2\lambda\pi)^{3+a} F_a(\xi,\eta),$$

where $\xi = |l+m|\lambda\pi$, $\eta = |l-m|\lambda\pi$ and $r = |g|/2\lambda\pi$ and

$$F_a(\xi,\eta) = \int_0^1 r^{2+a}\, \mathrm{Sinc}\,(\xi r)\, \mathrm{Sinc}\,(\eta r)\, dr, \quad \mathrm{Sinc}\,(x) \equiv \frac{\sin x}{x}. \tag{65}$$

Note that the five fold integral which defines the $\hat{B}(l,m)$ has been reduced to a one-dimensional integral, and hence the value of the coefficients can be easily computed numerically by an accurate quadrature formula, and stored in an array at the beginning of the calculation.

Expressions of F_a

We have the symmetry relation $F_a(\xi,\eta) = F_a(\eta,\xi)$. For integer values of a, F_a has an explicit analytical expression.

We give the expressions for $a = 0$ (Maxwellian gas) and $a = 1$ (H. S. gas)

$$F_0(\xi,\eta) = \frac{p\sin(q) - q\sin(p)}{\xi\eta pq}, \tag{66}$$

$$F_1(\xi,\eta) = \frac{p^2(q\sin(q)+\cos(q)) - q^2(p\sin(p)+\cos(p)) - 4\xi\eta}{2\xi\eta pq}, \tag{67}$$

where $p = (\xi+\eta)$, $q = (\xi-\eta)$.

In the 2D case we have

$$\hat{B}(l,m) = C_a 4\pi^2 (2\pi\lambda)^{2+a} \int_0^1 F_a(\xi,\eta), \quad F_a(\xi,\eta) = r^{1+a} J_0(\xi r) J_0(\eta r)\, dr. \tag{68}$$

3.4 - *Properties of the spectral method*

Notations

For any $t \geq 0$, $f_N(t) \in \mathbb{P}^N$ where

$$\mathbb{P}^N = span\{e^{ik\cdot v} \mid -N \leq k_j \leq N, j = 1,2,3\}.$$

Let $\mathcal{P}_N : L^2([-\pi,\pi]^3) \to \mathbb{P}^N$ be the orthogonal projection upon $\mathbb{P}^N$ in the inner product of $L^2([-\pi,\pi]^3)$, hence $f_N = \mathcal{P}_N f$. We have the following [**106**]

Proposition 2. *Let $f \in L^2([-\pi,\pi]^3)$, $f \geq 0, \forall\ v$, and let us define*

$$\begin{pmatrix} \rho \\ \rho u \\ \rho e \end{pmatrix} := \int_{[\pi,\pi]^3} f \begin{pmatrix} 1 \\ v \\ |v|^2 \end{pmatrix} dv.$$

Then the following relations hold

$$\rho = \rho_N, \quad |\rho u - \rho u_N| \leq \frac{C_1}{N}\|f\|_2, \quad |\rho e - \rho e_N| \leq \frac{C_2}{N^2}\|f\|_2. \tag{69}$$

This estimates can be strongly improved if f is smooth. If $f \in H^r_p([-\pi,\pi]^3)$, where $r \geq 0$ is an integer and $H^r_p([-\pi,\pi]^3)$ is the subspace of the Sobolev space $H^r([-\pi,\pi]^3)$, which consist of periodic functions, for each $\varphi \in L^2([-\pi,\pi]^3)$ we have [**106**]

$$| < f, \varphi > - < f, \varphi_N > | \leq \|\varphi\|_2 \, \|f - f_N\|_2 \leq \frac{C}{N^r} \|\varphi\|_2 \|f\|_{H^r_p}, \tag{70}$$

where $\|\cdot\|_{H^r_p}$ denotes the norm in $H^r_p([-\pi,\pi]^3)$. This inequality shows that the projection error on the moments decays faster than algebraically when the solution is infinitely smooth.

Using the previous notations the spectral method becomes

$$\frac{\partial f_N}{\partial t} = Q_N(f_N), \tag{71}$$

where $Q(f)$ denotes the Boltzmann operator with the relative velocity restricted on $\mathcal{B}_{2R}$.

We have the following consistency result in the L^2-norm for $Q_N(f_N)$ [**106**]

Theorem 2. *Let $f \in L^2([-\pi,\pi]^3)$, then*

$$\|Q(f) - Q_N(f_N)\|_2 \leq C \left(\|f - f_N\|_2 + \frac{\|Q(f_N)\|_{H^r_p}}{N^r} \right), \quad \forall\, r \geq 0, \tag{72}$$

where C depends on $\|f\|_2$.

The previous estimate states that the rate of convergence in the L^2-norm of $Q_N(f_N)$ to $Q(f)$ depends only on the speed of convergence of f_N to f. Hence if f_N is spectrally accurate so it is $Q_N(f_N)$. The following corollary states the spectral accuracy of the approximation of the collision operator [**106**]

Corollary 1. *Let $f \in H^r_p([-\pi,\pi]^3)$, $r \geq 0$ then*

$$\|Q(f) - Q_N(f_N)\|_2 \leq \frac{C}{N^r} \left(\|f\|_{H^r_p} + \|Q(f_N)\|_{H^r_p} \right). \tag{73}$$

3.5 - *Numerical results*

In our tests we use the spectral method coupled with an explicit fourth-order Runge-Kutta method in time. This scheme provides the high temporal accuracy needed to demonstrate spectral accuracy in velocity.

3.5.1 - Test #1 2D Maxwellian molecules

We consider the initial condition

$$f(v,0) = \frac{v^2}{\pi\sigma^2}\exp(-v^2/\sigma^2).$$

The integration time is $t_{\max} = 40$. This problem has an exact solution given by [**10**]

$$f(v,t) = \frac{1}{4\pi\mathbb{S}^2\sigma^2}\left(2S - 1 + \frac{1-S}{2S}\frac{v^2}{\sigma^2}\right)\exp\left(-\frac{v^2}{2S\sigma^2}\right), \tag{74}$$

where $S = 1 - \exp(-\sigma^2 t/8)/2$. This test is used to check spectral accuracy, by comparing the error at a given time, when using $N = 8, 16$, and 32 Fourier modes for each dimension.

3.5.2 - Test #2 3D Maxwell molecules

Next, we consider the initial condition

$$f(v,0) = \frac{1}{2(2\pi\sigma^2)^{3/2}}\left[\exp\left(-\frac{|v-2\sigma e_1|^2}{2\sigma^2}\right) + \exp\left(-\frac{|v+2\sigma e_1|^2}{2\sigma^2}\right)\right]$$

with $\sigma = \lambda\pi/6$. The integration time is $t_{\max} = 10$. Here $e_1 \equiv (1,0,0)$ denotes the unit vector in the direction v_x. This test is used to check the evolution law for the stress tensor which has an analytical expression [**105**].

3.5.3 - Test #3 2D VHS molecules

Finally we take as initial condition

$$f(v,0) = \frac{1}{2(2\pi\sigma^2)}\left[\exp\left(-\frac{|v-2\sigma e_1|^2}{2\sigma^2}\right) + \exp\left(-\frac{|v+2\sigma e_1|^2}{2\sigma^2}\right)\right]$$

with $\sigma = \lambda\pi/6$. The integration time is $t_{\max} = 10$. This test is used compare the relaxation to equilibrium of the stress tensor for Maxwellian molecules, with the relaxation of other VHS molecules.

In all the computations the scaling parameters are chosen in such a way that the numerical support of the initial condition is well approximated by $\mathcal{B}_R$.

The results are reported in Figures 9 and 10. Figure 9 clearly shows the spectral accuracy of the method as well as the aliasing effect. The latter requires a careful balance between resolution and choice of the computational domain. For example, the convergence rates at $t = 5$ are $\eta = 5.6$ passing from $N = 8$ to $N = 16$ and $\eta = 9.7$ passing from $N = 16$ to $N = 32$.

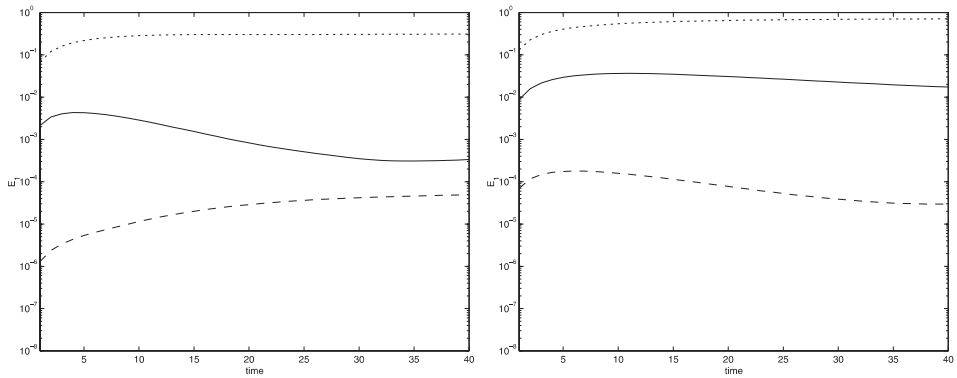

Fig. 9. – Test #1 2D Maxwell molecules: L_1 relative norm of the error vs time (left) and with a more compact initial condition (right) for $N = 8, 16$, and 32 modes per direction

Remark 4. *The same spectral Fourier strategy for the collision operator can be applied also to:*

• *Landau equations. In this case the overall computational cost of the spectral scheme can be reduced to* $O(N^3 \ln N)$ *thanks to the algebraic structure of the operator (see* [**109**], [**46**]*). Using further approximations fast methods for Landau equations were also derived in* [**22**], [**38**], [**78**].

• *Boltzmann equation for granular gases. In the spectral method, a suitable source of energy, for example a diffusive operator, has to be considered to avoid the Dirac delta stationary states of the operator (see* [**91**], [**47**]*). As an alternative, rescaling velocity techniques can be used (see Filbet and Russo in* [**110**] *and the references therein).*

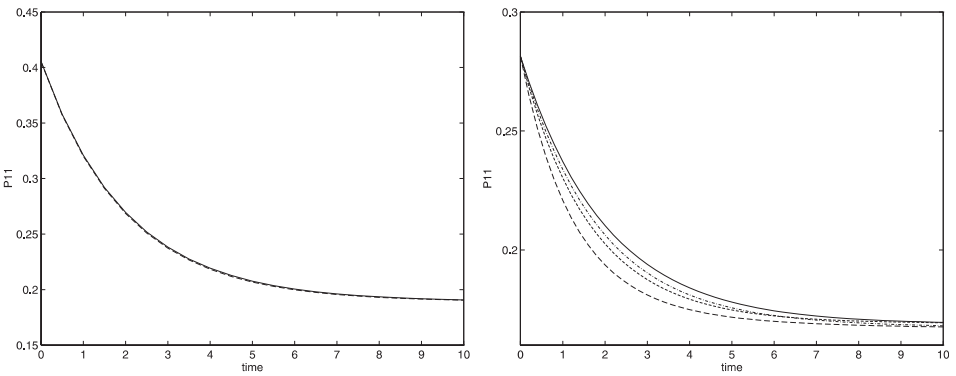

Fig. 10. – Test #2 and #3: Time evolution of the stress tensor component P_{11} in 3D Maxwell molecules for $N = 16$ modes (left). Exact reference solution (line) and numerical solution (dashed line). Relaxation of the stress tensor component P_{11} in 2D VHS molecules for different values of $a = 0, 0.5, 1, 2$ for $N = 32$ modes (right)

4 - Non homogeneous case

The solution of the transport equation can be obtained in a variety of ways accordingly to the particular application considered (see for example [**79**], [**49**], [**120**] and the references therein). A review of such a broad field is above the scope of this paper. Here we give the details of the methods developed originally in [**49**] which has several nice properties, among which to allow the use of large time steps even for large velocities.

4.1 - *Positive and conservative schemes*

Let us consider the transport equation written as

$$\partial_t f + \nabla_x (v f) = 0, \quad \forall (t,x) \in \mathbb{R}^+ \times \mathbb{R}^d. \tag{75}$$

Then, the solution of the transport equation at time t^{n+1} reads

$$f(t^{n+1},x) = f(t^n, x - v\,\Delta t), \quad \forall x \in \mathbb{R}^d.$$

For simplicity, let us restrict ourselves to a one dimensional problem and introduce a finite set of mesh points $\{x_{i+1/2}\}_{i\in I}$ on the computational domain. We will use the notations $\Delta x = x_{i+1/2} - x_{i-1/2}$, $C_i = [x_{i-1/2}, x_{i+1/2}]$ and x_i the center of C_i. Assume the values of the distribution function are known at time $t^n = n\,\Delta t$ on cells C_i, we compute the new values at time t^{n+1} by integration of the distribution function on each sub-interval. Thus, using the explicit expression of the solution, we have

$$\int_{x_{i-1/2}}^{x_{i+1/2}} f(t^{n+1},x)\,dx = \int_{x_{i-1/2}-v\,\Delta t}^{x_{i+1/2}-v\,\Delta t} f(t^n,x)\,dx,$$

then, setting

$$\Phi_{i+1/2}(t^n) = \int_{x_{i+1/2}-v\,\Delta t}^{x_{i+1/2}} f(t^n,x)\,dx,$$

we obtain the conservative form

$$\int_{x_{i-1/2}}^{x_{i+1/2}} f(t^{n+1},x)dx = \int_{x_{i-1/2}}^{x_{i+1/2}} f(t^n,x)dx \;+\; \Phi_{i-1/2}(t^n) \;-\; \Phi_{i+1/2}(t^n). \tag{76}$$

The evaluation of the average of the solution over $[x_{i-1/2}, x_{i+1/2}]$ allows to ignore fine details of the exact solution which may be costly to compute. The main step is now to

choose an efficient method to reconstruct the distribution function from the cell average on each cell C_i. We will consider a reconstruction via primitive function preserving positivity and maximum values of f [**49**]. Let $F(t^n,x)$ be a primitive of the distribution function $f(t^n,x)$, if we denote by

$$f_i^n = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} f(t^n,x)dx,$$

then $F(t^n,x_{i+1/2}) - F(t^n,x_{i-1/2}) = \Delta x f_i^n$ and

$$F(t^n,x_{i+1/2}) = \Delta x \sum_{k=0}^{i} f_k^n = w_i^n.$$

First we construct an approximation of the primitive on the small interval $[x_{i-1/2},x_{i+1/2}]$ using the stencil $\{x_{i-3/2},x_{i-1/2},x_{i+1/2},x_{i+3/2}\}$

$$\tilde{F}_h(t^n,x) = w_{i-1}^n + (x-x_{i-1/2})\,f_i^n + \frac{1}{2\Delta x}(x-x_{i-1/2})(x-x_{i+1/2})[f_{i+1}^n - f_i^n]$$

$$+\frac{1}{6\Delta x^2}(x-x_{i-1/2})(x-x_{i+1/2})(x-x_{i+3/2})[f_{i+1}^n - 2f_i^n + f_{i-1}^n],$$

where we use the relation $w_i^n - w_{i-1}^n = \Delta x f_i^n$. Thus, by differentiating $\tilde{F}_h(x)$, we obtain a third order accurate approximation of the distribution function on the interval $[x_{i-1/2},x_{i+1/2}]$

$$\tilde{f}_h(t^n,x) = \frac{\partial \tilde{F}_h}{\partial x}(t^n,x) = f_i^n$$

$$+\frac{1}{6\,\Delta x^2}\Big[2\,(x-x_i)(x-x_{i-3/2}) + (x-x_{i-1/2})(x-x_{i+1/2})\Big](f_{i+1}^n - f_i^n)$$

$$-\frac{1}{6\,\Delta x^2}\Big[2\,(x-x_i)(x-x_{i+3/2}) + (x-x_{i-1/2})(x-x_{i+1/2})\Big](f_i^n - f_{i-1}^n).$$

Unfortunately, this approximation does not preserve positivity of the distribution function f. Then, in order to satisfy a maximum principle and to avoid spurious oscillations we introduce slope correctors

$$f_h(t^n,x) = f_i^n \tag{77}$$

$$+\frac{\varepsilon_i^+}{6\,\Delta x^2}\Big[2\,(x-x_i)(x-x_{i-3/2}) + (x-x_{i-1/2})(x-x_{i+1/2})\Big](f_{i+1}^n - f_i^n)$$

$$-\frac{\varepsilon_i^-}{6\,\Delta x^2}\Big[2\,(x-x_i)(x-x_{i+3/2}) + (x-x_{i-1/2})(x-x_{i+1/2})\Big](f_i^n - f_{i-1}^n),$$

with

$$\varepsilon_i^{\pm} = \begin{cases} \min\left(1, 2f_i^n/(f_{i\pm1}^n - f_i^n)\right) & \text{if } f_{i\pm1}^n - f_i^n > 0, \\ \min\left(1, -2\,(f_\infty - f_i^n)/(f_{i\pm1}^n - f_i^n)\right) & \text{if } f_{i\pm1}^n - f_i^n < 0, \end{cases} \tag{78}$$

where $f_\infty = \max_{j\in I}\{f_j^n\}$ is a local maximum.

The theoretical properties of this reconstruction can be summarized by the following [**49**]

Proposition 3. *The approximation of the distribution function $f_h(x)$, defined by (77)-(78), satisfies*

- *The conservation of the average: for all $i \in I$,* $\int_{x_{i-1/2}}^{x_{i+1/2}} f_h(x)dx = \Delta x f_i$.
- *The maximum principle: for all $x \in (x_{min}, x_{max})$,* $0 \le f_h(x) \le f_\infty$.

Moreover, if we assume that the Total Variation of the distribution function $f(x)$ is bounded, then we obtain the global estimate

$$\int_{x_{min}}^{x_{max}} |\, f_h(x) - \tilde{f}_h(x)|\, dx \le 4\, TV(f)\, \Delta x,$$

where $\tilde{f}_h$ denotes the third order approximation of f without slope corrector.

Remark 5. *If the solution is smooth, we can check numerically that the scheme is third order. In several dimensions we can perform reconstruction dimension by dimension using tensor product.*

4.2 - *Time discretization*

The time discretization in the collision step is more delicate than in the transport step, because of the severe restriction on the time step when the system is close to the *stiff* (fluid) regimes.

For not small Knudsen numbers, we can use Runge-Kutta or multistep schemes. For small Knudsen numbers, implicit schemes are practically unusable due to the high computational cost required to solve the large system of nonlinear equation originated by the collision operator. Here we will use a different approach based on the Time Relaxed (TR) scheme proposed in [**51**]. For more details on the derivation and the properties of the schemes we refer to Section 8 of this paper and to [**51**], [**48**].

These time discretizations are based on the *Wild sum* expansion of the solution of the equation

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon}(P(f,f) - \mu f)$$

where P is a nonegative bilinear operator. The general schemes of order m read

$$f^{n+1}(v) = (1-\tau)\sum_{k=0}^{m} \tau^k f_k^n(v) + \tau^{m+1} M(v), \tag{79}$$

where $f^n = f(n\Delta t)$, Δt is a small time interval, and the functions f_k^n are defined recursively as

$$f_{k+1}^n(v) = \frac{1}{k+1}\sum_{h=0}^{k} \frac{1}{\mu} P(f_h^n, f_{k-h}^n), \quad k = 0,1,2,\ldots \tag{80}$$

The quantity $\tau = 1 - e^{-\mu\Delta t/\varepsilon}$ is called *relaxed time*.

The Boltzmann equation can be readily written in the above form in the case of Maxwellian kernels. In fact we have $P(f,f) = Q^+(f,f)$ and $\mu = L[f] = 4\pi C_0 \rho$. In the case of the Boltzmann equation with non constant $L[f]$ we can proceed as follows. If the loss term is bounded, then the equation can always be written in the above form, where μ is a positive constant such that

$$\mu \geq L[f](v) \equiv \int_{\mathbb{R}^3}\int_{\mathbb{S}^2} \sigma(|v - v_*|, \omega) f(v_*)\, d\omega\, dv_*,$$

and $P(f,f) = Q_+(f,f) + (\mu - L[f](v))f$.

The drawback of this simple choice, namely of taking a fixed μ, is that the truncation error of the schemes is proportional to $(\mu\Delta t)^m$, and therefore it can be large for large μ. A better choice would be to compute at each time step

$$\mu = \max_{\Omega}(L[f](v) - Q_+(f,f)/f),$$

where Ω is some relevant region of velocity space (e.g. $[-T/2, T/2]^3$). This is the smallest μ that makes the first order scheme positive.

Note that for $\mu\Delta t/\varepsilon \to 0$ the schemes project over the local Maxwellian M. For intermediate regions, an adaptive time step technique can be used.

4.3 - *Implementation issues*

Several improvements to the performance of the schemes can be done as a consequence of the splitting strategy. Here we describe briefly two of the most relevant.

Parallelization

As already mentioned, the code can be implemented in a parallel environment since collisions take place simultaneously at all space points during the relaxation step. More precisely, let $f_{i,j}^n$ denote f at time t^n, position x_i and velocity v_j.

- The transport step works on the space variables (velocity v_j is just a parameter), thus $T_{\Delta t}$ acts on the columns of matrix $f_{i,j}^n$.
- The collision step works on the velocity variables (position x_i is just a parameter), thus $C_{\Delta t}$ acts on the rows of matrix $f_{i,j}^n$.

Then a parallel algorithm works as follows

(a) $f^* = T_{\Delta t}(f^n)$ [computation]
(b) $f^* \to f^{*,T}$ [communication]
(c) $f^{n+1,T} = C_{\Delta t}(f^{*,T})$ [computation]
(d) $f^{n+1,T} \to f^{n+1}$ [communication]

Multi-resolution

Transport is more efficient and less accurate than the collision step, thus a finer grid can be used for transport than for collision.

Let f_h denote the function on the fine grid in velocity and f_H the function on the coarse grid (see Figure 11). Then the multiresolution algorith works as follows

(a) $f_h^* = T_{\Delta t}(f_h^n)$
(b) $\hat{f}_h^*$ is computed by FFT on the fine grid
(c) higher modes are neglected to get $\hat{f}_H^*$ (on empty circles)
(d) $\hat{f}_H^{n+1} = C_{\Delta t}(\hat{f}_H^*)$
(e) f_h^{n+1} (on green nodes) is computed by trigonometric interpolation using IFFT from $\hat{f}_H^{n+1}$

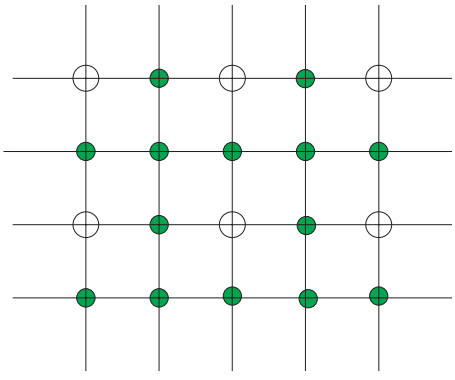

Fig. 11. – The fine (all nodes) and the coarse grids (empty circles)

4.4 - *Numerical tests*

The tests are performed using a parallel machine, and a 2D model in velocity and 1D or 2D in space.

4.4.1 - Test #1: time accuracy

We use first order TR scheme with adaptive time stepping and μ given by

$$\mu(t) = 3/2 \max_{v\in\Omega}\left(L(v) - \frac{Q_+}{f}\right),$$

where $\Omega = [-T/2, T/2]^3$. We use 64^2 modes in 2D so that the velocity error can be neglected. We compare the results with a reference solution obtained with a 4th order Runge-Kutta with a fixed time step. The results are reported in Table 3.

TABLE 3. – *Test #1 Comparison between fixed Δt and $\Delta t = \tau/\mu(t)$*

$\tau = \mu(t)\,\Delta t$	n_{Tot}	Numerical error $\varepsilon^{(1)}$ with a fixed $\Delta t = T/n_{Tot}$	Numerical error $\varepsilon^{(2)}$ with $\Delta t = \tau/\mu(t)$
0.010	50	0.0036	0.002
0.025	20	0.0080	0.007
0.050	11	0.0160	0.014
0.100	08	0.0300	0.025
0.200	07	0.0520	0.045
0.500	05	0.2000	0.090
1.000	03	XXXX	0.150
3.000	02	XXXX	0.040
5.000	01	XXXX	0.006

The time evolution of μ is given in Figure 12. It shows that the method becomes more and more accurate as the solution approaches the Maxwellian state.

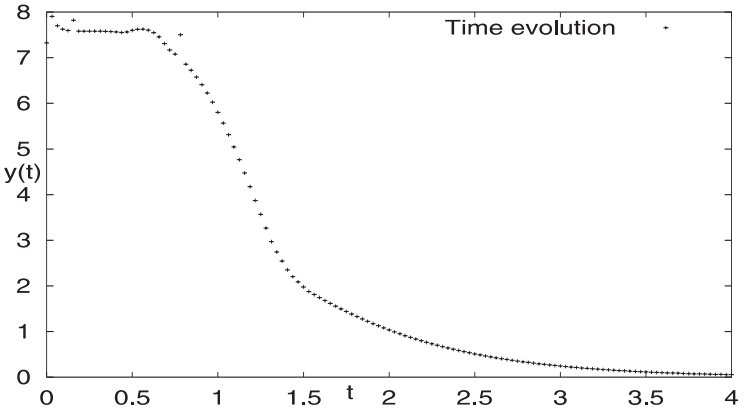

Fig. 12. – Time evolution of the value $\mu(t)$

4.4.2 - Test #2: space accuracy

Next we test the overall space accuracy of the method using as initial condition

$$f_0(x,v) = (1+\beta\,\cos(k_0\,x))\,\exp(-v^2/2), \quad (x,v)\in[0,L]\times\mathbb{R}^2,$$

with periodic boundary conditions. The error is computed as

$$\varepsilon_{2h} = \max_{t\in(0,T)} (|f_h(t) - f_{2h}(t)|_1)/|f_0|_1,$$

and the results are given in Table 4. Second order accuracy is observed.

TABLE 4. – *Test #2 Convergence results*

Numerical parameters	relative l^1 error norm	$\varepsilon_{2h}/\varepsilon_h$
$n_x = 032$, $n_{v_x} = n_{v_y} = 08$, $\Delta t = 0.100$	$\varepsilon_{4h} = 0.3835$	5.20
$n_x = 064$, $n_{v_x} = n_{v_y} = 16$, $\Delta t = 0.050$	$\varepsilon_{2h} = 0.0738$	4.35
$n_x = 128$, $n_{v_x} = n_{v_y} = 32$, $\Delta t = 0.025$	$\varepsilon_h = 0.0169$	X

4.4.3 - Test #3: stationary shock profile

We consider a stationary shock wave problem for the Boltzmann equation solved on a finite domain $-L < x < L$ with boundary conditions that the incoming flux of particles at $x = \pm L$ is distributed according to the Maxwellian flux $vM^{\pm}(v)$. As initial data, we take $f(x,v,0) = M(\rho,u,T)$, with

$$\rho = 1.0, \quad T = 1.0, \quad \mathcal{M} = 2.0, \qquad L > x > 0,$$

where $\mathcal{M}$ is the Mach number. The mean velocity is

$$u_x = -\mathcal{M}\sqrt{\gamma T}, \quad u_y = 0,$$

with $\gamma = 5/3$.

The values for ρ, u and T for $x < 0$ are given by the Rankine-Hugoniot conditions [**131**].

The profiles are shown in Figure 13 for different Knudsen numbers. As a reference solution we report also the solution obtained by Monte Carlo methods.

4.4.4 - Test #4 $2+2$ dimensional BE: the ghost effect

Consider a gas between two plates at rest in a finite domain. In this situation, the stationary state at a uniform pressure (the velocity is equal to zero and the pressure is constant) is an obvious solution of the Navier-Stokes equations; the temperature

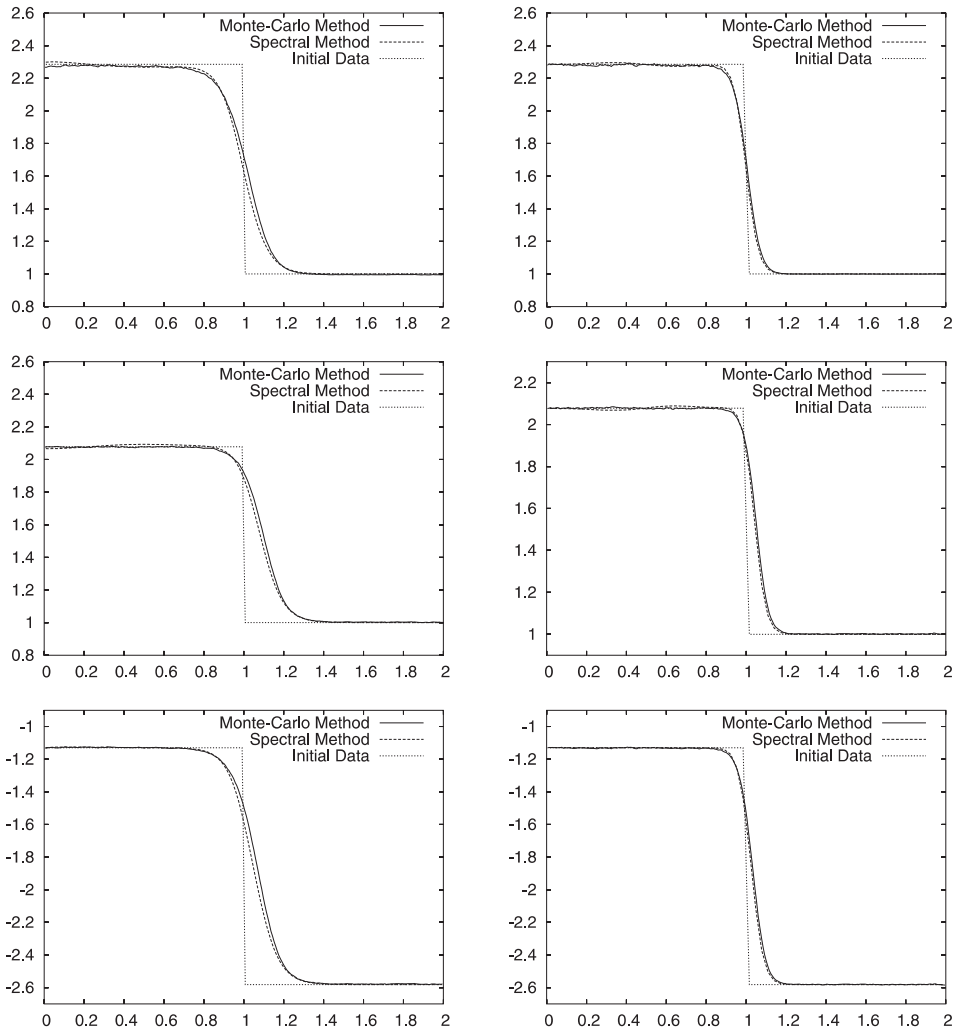

Fig. 13. – Test #3 Shock profiles at Mach 2: $\varepsilon = 10^{-1}$ (left) and $\varepsilon = 0.05$ (right). From top to bottom: density ρ, mean velocity u and temperature T

field is determined by the heat conduction equation

$$u = 0, \quad T = C \quad -\nabla_x(T^{1/2}\nabla_x T) = 0.$$

On the other hand, if we move the plate by a velocity proportional to the Knudsen number, then the macroscopic fields (density and temperature profiles) will be affected by the flow, even for vanishing Knudsen number. This effect, called "ghost effect", is predicted by the Hilbert expansion of the Boltzmann equation in terms of the Knudsen number, and it is rather difficult to capture numerically, since the flow

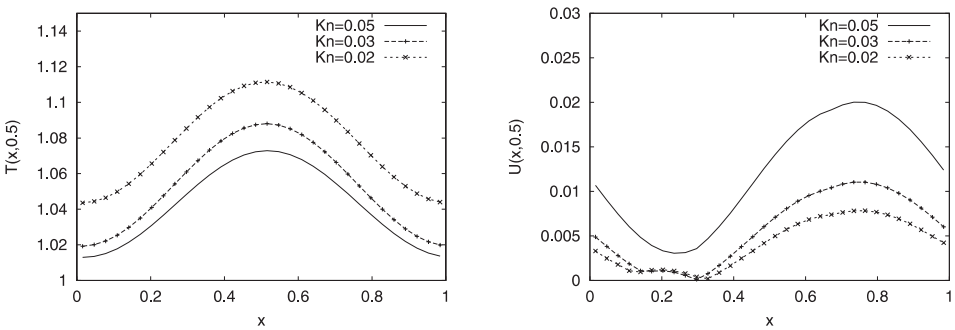

Fig. 14. – Ghost effect: temperature and mean velocity along $y = const$ for various Knudsen numbers $\varepsilon = 0.05, 0.02, 0.01$

velocity is very small [**123**]. The results show that the numerical solution agrees with one obtained by the asymptotic theory and not with the one obtained from the heat conduction equation; this result is a confirmation of the validity of the asymptotic theory.

Thus consider two parallel plates, both with temperature distribution

$$T_w(x) = 1 - 0.5 \cos(2\pi x); \quad \forall x \in (0,1),$$

in slow motion with velocity

$$u_w(x) = (\varepsilon, 0).$$

We use the hard spheres model with diffusive b.c. on the walls and periodic in x. The cross section of temperature and velocity profile are shown in Figure 14 for various values of the Knudsen number, while velocity field and isothermal lines are reported in Figure 15, for Knudsen number $\varepsilon = 0.02$.

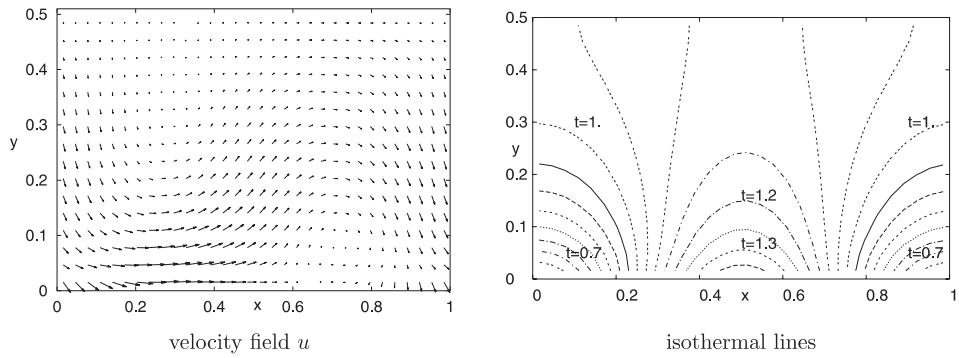

Fig. 15. – Ghost effect: velocity field, and isothermal lines; Knudsen number $\varepsilon = 0.02$

5 - Fast methods

In this section, for a particular class of interactions, using a Carleman-like representation of the collision operator together with a suitable angular approximation, we are able to derive spectral methods that can be evaluated through fast algorithms. Related methods based on FFT have been derived in [**14**], [**15**]. The class of interactions includes *Maxwellian molecules* in dimension two and *hard spheres* molecules in dimension three. The methods are strictly related to the spectral methods described in the previous sections, but on the contrary to the latter the new methods are able to take advantage of the structure of the Boltzmann equation in order to decouple the kernel modes. This is the central issue in the development of fast schemes. For more details and extensions of this technique to more general collision interactions and discrete velocity models we refer to [**90**], [**89**].

5.1 - *Carlemann representation and reduction to bounded domains*

Truncation of the Boltzmann collision operator in a bounded domain is the necessary preliminary stage of any deterministic method. Here we shall approximate the collision operator starting from a Carlemann-like representation which conserves more symmetries of the collision operator when one truncates it in a bounded domain.

The basic identity we shall need is

$$\frac{1}{2}\int\limits_{S^{d-1}} F(|u|\sigma - u)\,d\sigma = \frac{1}{|u|^{d-2}}\int\limits_{\mathbb{R}^d} \delta(2\,x\cdot u + |x|^2)\,F(x)\,dx. \tag{81}$$

Using (81) with $u = g = v - v_*$ and performing the change of variables $x \to x/2$ and $v_* \to y = v_* - v - x$ we can write

$$Q(f,f)(v) = 2^{d-1}\int\limits_{x\in\mathbb{R}^d}\int\limits_{y\in\mathbb{R}^d} B\left(|x+y|, -\frac{x\cdot(x+y)}{|x||x+y|}\right)\frac{1}{|x+y|^{d-2}}\,\delta(x\cdot y) \cdot [\,f(v+y)\,f(v+x) - f(v+x+y)\,f(v)]\,dx\,dy.$$

Now let us consider the bounded domain $\mathcal{D}_T = [\,-T,T]^d$ $(0 < T < +\infty)$. From now on let us write

$$\tilde{B}(x,y) = \frac{2^{d-1}}{|x+y|^{d-2}}\,B\left(|x+y|, -\frac{x\cdot(x+y)}{|x||x+y|}\right).$$

For VHS molecules we have

$$\tilde{B}(x,y) = \frac{2^{d-1}}{|x+y|^{d-2}} C_a |x+y|^a,$$

so that Maxwell molecules ($a = 0$) in dimension $d = 2$ and Hard Spheres ($a = 1$) in dimension $d = 3$ correspond simply to constant values.

Next we have to truncate the integration in x and y without affecting the action of the operator for compactly supported functions. Thus we set them to vary in $\mathcal{B}_S$, the ball of center 0 and radius S. For a compactly supported function f with support $\mathcal{B}_R$, we take $S = 2R$ in order to obtain all possible collisions. In fact we have

$$|x|^2 \le |x|^2 + |y|^2 = |x+y|^2 = |g|^2 \le (2R)^2,$$

thus $|x| \le 2R$ and similarly we get $|y| \le 2R$.

The operator now reads

$$Q^R(f,f)(v) = \int_{x \in \mathcal{B}_{2R}} \int_{y \in \mathcal{B}_{2R}} \tilde{B}(x,y)\, \delta(x \cdot y)\, [f(v+y) f(v+x) - f(v+x+y)\, f(v)]\, dx\, dy, \tag{82}$$

with $v \in \mathcal{B}_{\sqrt{2}R}$. The interest of this representation is to preserve the real collision kernel and its invariance properties.

Remark 6. *Let us recall that in the classical spectral method it is enough to take $T \ge (3+\sqrt{2})R/2$ to prevent intersections of the regions where f is different from zero. The operator (82) requires a slightly larger periodization since the arguments of the integrands are contained into $\mathcal{B}_{3\sqrt{2}R}$ instead of $\mathcal{B}_{(2+\sqrt{2})R}$. In fact we have that $|x| \le 2R$ and $|y| \le 2R$ imply $|x+y|^2 = |x|^2 + |y|^2 \le 8R^2$ (thanks to the orthogonality condition $x \cdot y = 0$ consequence of the δ function) and then $|x+y| \le 2\sqrt{2}R$. From this we get $|v+x+y| \le |v| + |x+y| \le \sqrt{2}R + 2\sqrt{2}R = 3\sqrt{2}R$. In this case we need to take $T \ge (3+\sqrt{2})R/\sqrt{2}$ as a bound for the periodization.*

5.2 - *Spectral methods*

Now we use the representation Q^R to derive the spectral methods. The main difference compared to the usual spectral method is in the way we truncate the collision operator.

Following the same computation as in the classical spectral method but using representation (82) we obtain the following set of ordinary differential equations on the Fourier coefficients

$$\frac{d\hat{f}_k(t)}{dt} = \sum_{\substack{l,\, m=-N \\ l+m=k}}^{N} \hat{\beta}(l,m)\hat{f}_l\hat{f}_m, \quad k=-N,\dots,N \tag{83}$$

where now $\hat{\beta}(l,m) = \beta(l,m) - \beta(m,m)$ with

$$\beta(l,m) = \int\limits_{x\in\mathcal{B}_{2R}} \int\limits_{y\in\mathcal{B}_{2R}} \tilde{B}(x,y)\,\delta(x\cdot y)\,e^{il\cdot x}\,e^{im\cdot y}\,dx\,dy.$$

In the sequel we shall focus on β, and one easily checks that $\beta(l,m)$ depends only on $|l|$, $|m|$ and $|l\cdot m|$.

Note that the classical way to truncate the Boltzmann collision operator for periodic functions, presented in Section 3, originates the following form of the kernel modes (63) in the x, y notation

$$\hat{\beta}_{\text{usual}}(l,m) = \int\limits_{x\in\mathcal{B}_{2R}} \int\limits_{y\in\mathcal{B}_{2R}} \tilde{B}(x,y)\,\delta(x\cdot y)\,\chi_{\{|x+y|\le 2R\}}\left[e^{il\cdot x}\,e^{im\cdot y} - e^{im\cdot(x+y)}\right] dx\,dy.$$

One can notice that here x and y are also restricted to the ball $\mathcal{B}_{2R}$ but the condition $|x+y|\le 2R$ couples the two modulus, such that the ball is not completely covered (for instance, if x and y are orthogonal both with modulus $2R$, the condition is not satisfied, since $|x+y| = 2\sqrt{2}R$).

5.3 - *Fast algorithms*

The search for fast deterministic algorithms for the collision operator, i.e. algorithms with a cost lower than $O(N^{2d+\kappa})$ (with typically $\kappa = 1$), consists mainly in identifying some convolution structure in the operator (see for example [**109**], [**98**]). If this is trivial for the loss part of the operator, for the gain part this is rather contradictory with the search for a conservative scheme in a bounded domain, since the boundary condition needed to prevent for the outgoing or ingoing collisions breaks the invariance.

The aim is to approximate each $\hat{\beta}(l,m)$ by a sum

$$\beta(l,m) \simeq \sum_{p=1}^{A} a_p(l)a'_p(m). \tag{84}$$

This gives a sum of A discrete convolutions and so the algorithm can be computed in $O(A\,N^d \log_2 N)$ operations by means of standard FFT techniques [**26**]. To this purpose we shall use a further approximated collision operator where the number of possible directions of collision is reduced to a finite set.

We start from representation (82) and write x and y in spherical coordinates

$$Q^R(f,f)(v) = \frac{1}{4} \int\limits_{e\in\mathbb{S}^{d-1}} \int\limits_{e'\in\mathbb{S}^{d-1}} \delta(e\cdot e')\,de\,de'$$

$$\cdot \left\{ \int\limits_{-R}^{R} \int\limits_{-R}^{R} \rho^{d-2}\,(\rho')^{d-2}\,\tilde{B}(\rho,\rho')[f(v+\rho' e')\,f(v+\rho e) - f(v+\rho e+\rho' e')\,f(v)]\,d\rho\,d\rho' \right\}.$$

Let us denote with $\mathcal{A}$ a discrete set of orthogonal couples of unit vectors (e,e'), which is even, i.e. $(e,e')\in\mathcal{A}$ implies that $(-e,e')$, $(e,-e')$ and $(-e,-e')$ belong to $\mathcal{A}$ (this property on the set $\mathcal{A}$ is required to preserve the conservation properties of the operator). Now we define $Q^{R,\mathcal{A}}$ to be

$$Q^{R,\mathcal{A}}(f,f)(v) = \frac{1}{4} \int\limits_{(e,e')\in\mathcal{A}} \left\{ \int\limits_{-R}^{R} \int\limits_{-R}^{R} \rho^{d-2}\,(\rho')^{d-2}\,\tilde{B}(\rho,\rho') \big[f(v+\rho' e')\,f(v+\rho e) \right. \tag{85}$$

$$\left. - f(v+\rho e+\rho' e')\,f(v)\big]\,d\rho\,d\rho' \right\} d\mathcal{A} \tag{86}$$

where $d\mathcal{A}$ denotes a discrete measure on $\mathcal{A}$ which is also even in the sense that $d\mathcal{A}(e,e') = d\mathcal{A}(-e,e') = d\mathcal{A}(e,-e') = d\mathcal{A}(-e,-e')$. It is easy to check that $Q^{R,\mathcal{A}}$ has the same conservation properties as Q_R.

We make the *decoupling assumption* that

$$\tilde{B}(x,y) = a(|x|)\,b(|y|). \tag{87}$$

This assumption is obviously satisfied if $\tilde{B}$ is constant. This is the case of Maxwellian molecules in dimension two, and hard spheres in dimension three (the most relevant kernel for applications). Extensions to more general interactions are discussed in [**89**].

We describe the method in dimension $d=3$ with $\tilde{B}$ satisfying the decoupling assumption (87) (see [**89**] for other dimensions). First we change to spherical coordinates

$$\beta(l,m) = \frac{1}{4} \int\limits_{e\in\mathbb{S}^2} \int\limits_{e'\in\mathbb{S}^2} \delta(e\cdot e') \left[\int\limits_{-R}^{R} \rho\,a(\rho)\,e^{i\rho(l\cdot e)}\,d\rho \right] \left[\int\limits_{-R}^{R} \rho'\,b(\rho')\,e^{i\rho'(m\cdot e')}\,d\rho' \right] de\,de'$$

and then we integrate first e' on the intersection of the unit sphere with the plane $e^\perp$,

$$\beta(l,m) = \frac{1}{4} \int\limits_{e\in\mathbb{S}^2} \phi^3_{R,a}(l\cdot e) \left[\int\limits_{e'\in\mathbb{S}^2\cap e^\perp} \phi^3_{R,b}(m\cdot e')\, de' \right] de$$

where

$$\phi^3_{R,a}(s) = \int\limits_{-R}^{R} \rho\, a(\rho)\, e^{i\rho s}\, d\rho, \qquad \phi^3_{R,b}(s) = \int\limits_{-R}^{R} \rho\, b(\rho)\, e^{i\rho s}\, d\rho.$$

Thus we get the following decoupling formula with two degrees of freedom

$$\beta(l,m) = \int\limits_{e\in\mathbb{S}^2_+} \phi^3_{R,a}(l\cdot e)\, \psi^3_{R,b}\big(\Pi_{e^\perp}(m)\big)\, de$$

where $\mathbb{S}^2_+$ denotes the half-sphere and

$$\psi^3_{R,b}\big(\Pi_{e^\perp}(m)\big) = \int\limits_0^\pi \phi^3_{R,b}\big(|\Pi_{e^\perp}(m)|\,\cos\theta\big)\, d\theta.$$

In the particular case where $\tilde{B} = 1$ (hard spheres), we can compute the functions ϕ^3_R and ψ^3_R:

$$\phi^3_R(s) = R^2\left[2\,\mathrm{Sinc}(Rs) - \mathrm{Sinc}^2(Rs/2)\right], \qquad \psi^3_R(s) = \int\limits_0^\pi \phi^3_R\big(s\,\cos\theta\big)\, d\theta.$$

Now the function $e \to \phi^3_{R,a}(l\cdot e)\,\psi^3_{R,b}\big(\Pi_{e^\perp}(m)\big)$ is periodic on $\mathbb{S}^2_+$. Taking a spherical parametrization (θ,φ) of $e\in\mathbb{S}^2_+$ and taking for the set $\mathcal{A}$ uniform grids of respective size M_1 and M_2 for θ and φ we get

$$\beta(l,m) \simeq \frac{\pi^2}{M_1 M_2} \sum_{p,q=0}^{M_1,M_2} a_{p,q}(l) a'_{p,q}(m)$$

where

$$a_{p,q}(l) = \phi^3_{R,a}(l\cdot e_{(\theta_p,\varphi_q)}), \qquad a'_{p,q}(m) = \psi^3_{R,b}(\Pi_{e^\perp_{(\theta_p,\varphi_q)}}(m))$$

and $(\theta_p,\varphi_q) = \Big(p\,\pi/M_1, q\,\pi/M_2\Big)$.

We shall consider this expansion with $M = M_1 = M_2$ to avoid anisotropy in the computational grid. The computational cost of the algorithm is then $O(M^2N^3\log_2 N)$, compared to $O(N^6)$ of the usual spectral method. Thus we require $M^2\log N \ll N^3$ in order to speed up the schemes.

5.4 - *Error estimates*

In order to give a consistency result, the first step will be to prove a consistency result for the approximation of Q^R by $Q^{R,M} = Q^{R,\mathcal{A}_\mathcal{M}}$ where $\mathcal{A}_\mathcal{M}$ is the uniform spherical grid with M points for each angular coordinate (see [**89**]). Here H^k_p denotes the periodic Sobolev space on $\mathcal{D}_\pi$.

Lemma 1. *The error on the approximation of the collision operator is spectrally small, i.e. for all $k > d-1$ such that $f \in H^k_p$,*

$$\|Q^R(f,f) - Q^{R,M}(f,f)\|_{L^2} \leq C_1 \frac{R^k \|f\|^2_{H^k_p}}{M^k}.$$

For the second step we use the consistency result ([**106**], Corollary 5.4) on the operator Q^R, which we quote here for the sake of clarity.

Lemma 2. *For all $k \in \mathbb{N}$ such that $f \in H^k_p$*

$$\|Q^R(f,f) - \mathcal{P}_N Q^R(f_N,f_N)\|_{L^2} \leq \frac{C_2}{N^k}\left(\|f\|_{H^k_p} + \|Q^R(f_N,f_N)\|_{H^k_p}\right).$$

Combining these two results, one gets the following consistency result of spectral accuracy [**89**]

Theorem 3. *For all $k > d-1$ such that $f \in H^k_p$,*

$$\|Q^R(f,f) - \mathcal{P}_N Q^{R,M}(f_N,f_N)\|_{L^2} \leq C_1 \frac{R^k \|f_N\|^2_{H^k_p}}{M^k} + \frac{C_2}{N^k}\left(\|f\|_{H^k_p} + \|Q^R(f_N,f_N)\|_{H^k_p}\right).$$

Finally let us focus briefly on the macroscopic quantities. In fact here no additional error (related to M) occurs, compared with the usual spectral method, since the approximation of the collision operator that we are using is still conservative. Following the method of [**106**], Remark 5.4, we have the following spectral accuracy result

$$\left|\langle Q^{R,M}(f,f),\varphi\rangle - \langle \mathcal{P}_N Q^{R,M}(f_N,f_N),\varphi\rangle\right|_{L^2} \leq \frac{C_3}{N^k}\|\varphi\|_{L^2}\left(\|f\|_{H^{k+d}_p} + \|Q^{R,M}(f_N,f_N)\|_{H^k_p}\right),$$

where φ can be replaced by $v, |v|^2$. Thus the error on momentum and energy is independent on M and it is spectrally small with respect to N.

5.5 - *Numerical results*

In this section we will present numeral results for the space homogeneous equation which show the improvement of the fast spectral algorithms with respect to the classical spectral methods. Extension to nonhomogeneous situations follows straightforwardly (see [**45**] for more details).

5.5.1 - Test #1: 2D Maxwell molecules

We consider the same test problem #1, namely the exact solution for Maxwell molecules in 2D, as for the standard spectral method. We present the results obtained by the classical spectral method and the fast spectral method with different number of discrete angles. In Table 5, we give a quantitative comparison of the numerical error at time $T_{end} = 1$.

TABLE 5. – *Comparison of the L^1 error in 2D between the classical spectral method and the fast spectral method with different numbers of discrete angles and with a second-order Runge-Kutta time discretization at time $T_{end} = 1$*

Number of points	Classical spectral	Fast spectral with $M = 4$	Fast spectral with $M = 6$	Fast spectral with $M = 8$
8	0.02013	0.02778	0.02129	0.02112
16	0.00204	0.00329	0.00238	0.00224
32	1.405E-5	2.228E-5	1.861E-5	1.772E-5

5.5.2 - Test #2-#3 : 2D Maxwell molecules and 3D VHS molecules

We now consider again the same tests as for the classical spectral method, namely the sum of two Gaussian, both in 2D and 3D velocity space. This tests are used to check the speed-up of the fast algorithm compared to the classical one. The results are given in Tables 6 and 7.

TABLE 6. – *Comparison of the computational time in 2D between the classical spectral method and the fast spectral method with different numbers of discrete angles and with a second order Runge-Kutta time discretization*

Number of points	Classical spectral	Fast spectral with $M = 4$	Fast spectral with $M = 6$	Fast spectral with $M = 8$
16	2 *sec.* 40	1 *sec.*15	1 *sec.* 70	2 *sec.* 30
32	38 *sec.* 01	5 *sec.* 55	8 *sec.* 47	11 *sec.* 10
64	616 *sec.*	35 *sec.* 50	54 *sec.* 66	71 *sec.* 27

TABLE 7. – *Comparison of the computational time in 3D between the classical spectral method and the fast spectral method with different numbers of discrete angles and with a second-order Runge-Kutta time discretization*

Number of points	Classical spectral	Fast spectral with $M = 4$	Fast spectral with $M = 6$	Fast spectral with $M = 8$
16	1 *min.* 14 *sec.*	3 *min.* 31 *sec.*	7 *min.* 45 *sec.*	13 *min.* 44 *sec.*
32	118 *min.* 02 *sec.*	50 *min.* 31 *sec.*	105 *min.* 19 *sec.*	186 *min.* 18 *sec.*
64	125 *h* 54 *min.*	8 *h* 45 *min.* 22 *sec.*	21 *h* 39 *min.*	35 *h* 01 *min.* 28 *sec.*

PART II
Probabilistic methods

6 - Random sampling

In this section we give a brief review of random sampling, which is at the basis of several Monte Carlo methods [**73**]. The most commonly used Monte Carlo methods for the simulation of the collision step, namely Nanbu-Babosky's and Bird's schemes [**9**], [**92**] will be introduced later.

6.1 - *Random number generators*

Random sampling assumes that one is able to sample from a uniformly distributed random number between 0 and 1. In this section we briefly review some on the main techniques to generate sequences of (pseudo) random numbers with the aid of a computer.

Real random numbers can not be generated by a computer because of several reasons, among which we mention

- floating points are used as approximation or real numbers;
- a really random sequence can not be generated even at a discrete level (it would require an infinite memory).

Random number generators produce a sequence of numbers which satisfies some statistical properties of true random sequences. In particular, one wishes to generate a sequence ξ_n which is

- uniformly distributed (approximate Lebesgue measure in $[0, 1]$, the frequency histogram approximates a uniform distribution, see Figure 16);

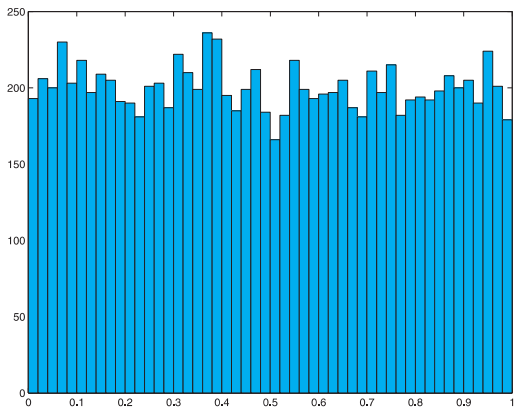

Fig. 16. – Histogram of a pseudorandom sequence. $N = 10000$, number of bins (subintervals of $[0, 1]$ used to realize the frequency histogram) is 50

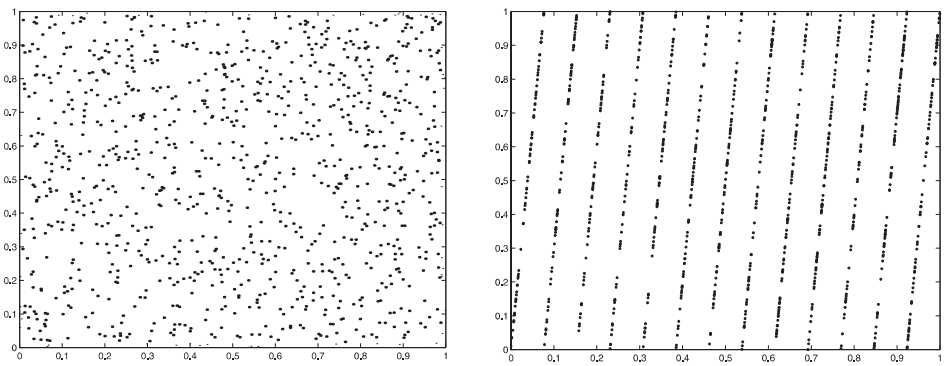

Fig. 17. – Examples of good (left) and bad (right) pseudo random number generators

• the elements of the sequence are uncorrelated (for example absence of pairwise correlation means that (ξ_n, ξ_{n+1}) should approximate Lebesgue measure in $[0,1]^2$, see Figure 17), absence of three-term correlation means that $(\xi_n, \xi_{n+1}, \chi_{n+2})$ should approximate Lebesgue measure in $[0,1]^3$), and so on);

• they have to be computed quickly.

A family of commonly used generators are the so-called Linear Congruential Generators (LCG) defined by the recurrent relation

$$x_{n+1} = (a x_n + c) \bmod m, \quad n \geq 0, \tag{88}$$

with $x_0, a, c, m \in \mathbb{N}$.

Dividing x_n by m one obtains an approximation of the uniform distribution in $[0,1]$. Clearly x_n is a sequence with period at most m, therefore m has to be large enough.

The quality of the result depends on the choice of a, c, m. A good choice, used by several computer codes is $m = 2^{31} - 1, a = 7^5, c = 0$. Further information on random number generators can be found, for example, in [**73**] or in [**75**].

6.2 - *Monovariate distribution*

The first sampling technique we describe is the so-called inverse transform method. Let $x \in \mathbb{R}$ be a random variable with density $p_x(x)$, i.e. $p_x(x) \geq 0$, $\int_\Omega p_x(x)\,dx = 1$, and let ξ be a uniformly distributed random variable (number) in $[0,1]$.

Then the relation between x and ξ can be found by equating the infinitesimal probabilities

$$p_x(x)\,dx = 1 \cdot d\xi.$$

By integration one has

$$P_x(x) = \int_{-\infty}^{x} p_x(y)\,dy = \xi, \tag{89}$$

where $P_x(x)$ is the distribution function corresponding to the random variable x, i.e. the primitive of $p_x(x)$. Then the random variable x can be sampled by sampling a uniformly distributed variable ξ, and then solving equation (89) with respect to x

$$x = P_x^{-1}(\xi). \tag{90}$$

Example 3. (Sampling from an exponential)
Let $p_x(x) = \exp(-x)$, $x \geq 0$ *(see Figure* (18)*). Then*

$$P_x(x) = \int_0^x \exp(-y)\,dy = 1 - \exp(-x) = \xi,$$

and therefore

$$x = -\ln(1-\xi),$$

or $x = -\ln\xi$, *because* $1-\xi$ *is also uniformly distributed in* $[0,1]$.

Unless otherwise stated, in this chapter we denote by ξ, ξ_1, ξ_2, and so on, samples from uniformly distributed random numbers in $[0,1]$.

We will consider density functions defined on the whole real line. If the support of the density is strictly contained in $\mathbb{R}$, as in the previous example, the density function can be defined by using the Heaviside Θ-function. In the example above one could define $p_x(x) = \exp(-x)\Theta(x), x \in \mathbb{R}$.

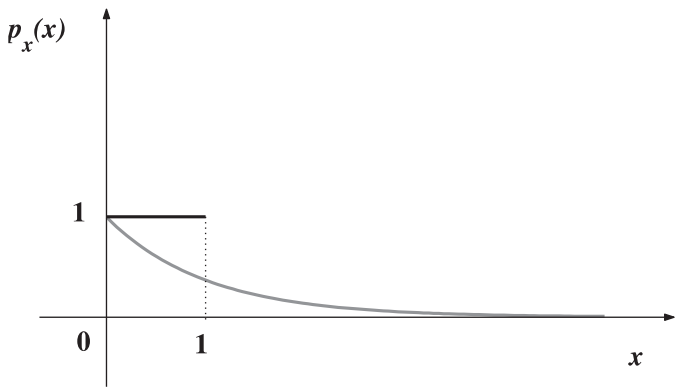

Fig. 18. – Sampling from an exponential distribution

It may be expensive to compute the inverse function, since in general a nonlinear equation has to be solved. In such cases, a different technique, the so-called acceptance-rejection, can be used.

Let x be a random variable with density $p_x(x)$, $x \in \mathbb{R}$. We look for a function $w(x) \geq p_x(x)\, \forall x \in \mathbb{R}$ whose primitive $W(x)$ is easily invertible. Let $A = \int_{-\infty}^{\infty} w(x)\, dx$ and denote with ξ_1 and ξ_2 uniformly $[0,1]$ random numbers.

Algorithm 1 (acceptance-rejection)

1. *Sample from* $w(x)/A$ *by solving the equation* $W(x) = A\xi_1$*;*
2. *if* $w(x)\xi_2 < p_x(x)$ *then accept the sample, else reject the sample and repeat step* 1.

In Figure 19, η has been sampled from $w(x)$. It will be accepted with probability equal to the ratio $p(\eta)/w(\eta)$.

The efficiency of the scheme depends on how easy it is to invert the function $W(x)$ and how frequently we accept the sample. The fraction of accepted samples equals the ratio of the areas below the two curves $p_x(x)$ and $W(x)$ and it is therefore equal to $1/A$.

Sometimes a density function is given as a convex combination of simpler density functions

$$p(x) = \sum_{i=1}^{M} w_i p_i(x),$$

where w_i are probabilities (i.e. $w_i \geq 0$, $\sum_{i=1}^{M} w_i = 1$), and $p_i(x)$ are probability densities. In that case the sampling can be performed as follows

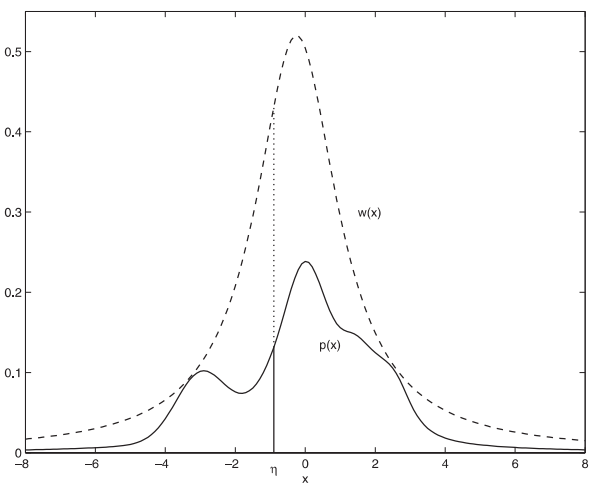

Fig. 19. – Use of the acceptance-rejection technique to sample from a random variables with a complicated density function $p(x)$

Algorithm 2

1. *select an integer* $i \in \{1, \dots, M\}$ *with probability* w_i*;*
2. *sample* x *from a random variable with density* $p_i(x)$.

Because of the relevance in several applications, step 1 of the previous algorithm deserves an extended discussion.

6.3 - *Discrete sampling*

Let us suppose that $k \in \{1, \dots, M\}$ is an integer random number, with probabilities $\{w_k\}$. In order to sample k with probability w_k we can proceed as follows. Divide interval $[0, 1]$ in M intervals, i-th interval being of length w_i, extract a uniform $[0, 1]$ random number ξ, and detect the interval k to which ξ belongs in the following way (see figure (20).

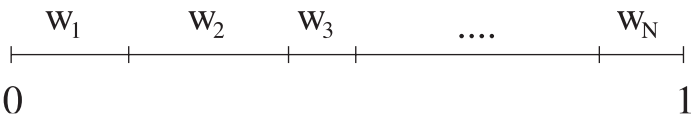

Fig. 20. – Discrete sampling from given probabilities by binary search

Algorithm 3 (discrete sampling 1)

1. *Compute* $W_k = \sum_{i=1}^{k} w_k$, $k = 1, \dots, M$, $\quad W_0 = 0$*;*
2. *find the integer* k *such that* $W_{k-1} \leq \xi < W_k$.

For an arbitrary set of probabilities $\{w_i\}$, once W_i have been computed, step 2 can be performed with a binary search, in $O(\ln M)$ operations.

Example 4. (Sampling a geometrically distributed variable)

As an application of the above procedure, let us show how to sample from a geometrically distributed random variable. Let us take $\tau \in (0, 1)$*, and assume that*

$$w_k = (1 - \tau)\tau^{k-1}, \quad k = 1, \dots, \infty. \tag{91}$$

Then it is

$$W_k = \sum_{j=1}^{k} w_j = 1 - \tau^k. \tag{92}$$

The integer k *that satisfies the condition* $W_{k-1} \leq \xi < W_k$ *is obtained as*

$$k = \left\lfloor \frac{\ln(1 - \xi)}{\ln \tau} \right\rfloor + 1,$$

where $\lfloor x \rfloor$ *denotes the integer part of* x.

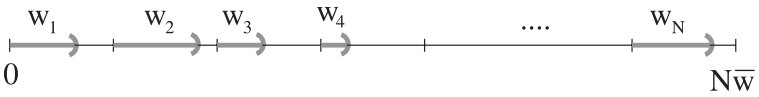

Fig. 21. – Discrete sampling from given probabilities by the rejection technique

In the general case in which a simple explicit expression of W_k is not known, the approach outlined above is efficient if the numbers W_k can be computed once, and then used several times, and if M is not too large. If the number M is very large and the numbers w_k may change, a more efficient acceptance-rejection technique can be used, if at least an estimate $\overline{w}$ such that $\overline{w} \geq w_i,\ i = 1, \ldots, M$ is known (see Figure 21).

Algorithm 4 (discrete sampling 2)

1. *Select an integer random number uniformly in* $[1, \ldots, M]$, *as*
$$k = \lfloor M\xi_1 \rfloor + 1.$$
2. *if* $w_k\xi_2 < \overline{w}$ *then accept the sampling else reject it and repeat point* 1.

Clearly the procedure can be generalized to the case in which the estimate depends on k, i.e. $\overline{w}_k \geq w_k,\ k = 1, \ldots, M$.

The above technique is of crucial relevance in the Monte Carlo simulation of the scattering in the Boltzmann equation.

6.3.1 - Sampling without repetition

Sometimes it is useful to extract n numbers, $n \leq N$, from the sequence $1, \ldots, N$ without repetition. This sampling is often used in several Monte Carlo simulations. A simple and efficient method to perform the sampling is the following.

Algorithm 5 (sampling without repetition)

1. *set* $\mathsf{ind}_i = i,\ i = 1, \ldots, N$;
2. $M = N$;
3. *for* $i = 1$ *to* n
   - *set* $j = \lfloor M\xi \rfloor + 1$, $\mathsf{seq}_i = \mathsf{ind}_j$,
   - $\mathsf{ind}_j = \mathsf{ind}_M$, $M = M - 1$

   end for

At the end the vector seq will contain n distinct integers randomly sampled from the first N natural numbers. Of course if $n = N$, the vector seq contains a random permutation of the sequence $1, \ldots, N$.

6.4 - *Multivariate distributions*

Suppose we want to sample a n-dimensional random variable $x = (x_1, \ldots, x_n)$, whose probability density is $p_x(x)$.

If the density can be written as a product of densities of scalar random variables (marginal probability densities), i.e. if

$$p_x(x_1, \dots, x_n) = p_1(x_1) p_2(x_2) \cdots p_n(x_n),$$

then the n scalar random variables $x_1, \dots, x_n$ are independent, and the problem is equivalent to sampling n monovariate random variables.

If this is not the case, then one may first look for a transformation $T : x \to \eta = T(x)$ such that in the new variables the probability density is factorized, i.e.

$$p_x(x_1, \dots x_n) dx_1 dx_2 \dots dx_n = p_{\eta_1}(\eta_1) p_{\eta_2}(\eta_2) \cdots p_{\eta_n}(\eta_n) d\eta_1 d\eta_2 \dots d\eta_n, \tag{93}$$

then sample the variables $\eta_1, \dots \eta_n$, and finally compute x by inverting the map T, i.e. $x = T^{-1}(\eta)$.

In some cases such transformation can be readily found. In other cases it is more complicated. There is a general technique to find a mapping $T : x \to \xi$ where $\xi = (\xi_1, \dots, \xi_n)$ denotes a uniformly random variable in $[0, 1]^n$. Of course such transformation is not unique, since we only impose that its Jacobian determinant $J = |\partial \xi / \partial x|$ is a given function $p_x(x_1, \dots, x_n)$.

An explicit transformation is constructed as follows

$$\begin{aligned}
T_1(x_1) &= \int_{-\infty}^{x_1} d\eta \int_{\mathbb{R}^{n-1}} dx_2 \dots dx_n \, p_x(\eta, \dots, x_n), \\
T_2(x_1, x_2) &= \frac{\displaystyle\int_{-\infty}^{x_2} d\eta \int_{\mathbb{R}^{n-2}} dx_3 \dots dx_n \, p_x(x_1, \eta, \dots, x_n)}{\displaystyle\int_{-\infty}^{\infty} d\eta \int_{\mathbb{R}^{n-2}} dx_3 \dots dx_n \, p_x(x_1, \eta, \dots, x_n)}, \\
T_3(x_1, x_2, x_3) &= \frac{\displaystyle\int_{-\infty}^{x_3} d\eta \int_{\mathbb{R}^{n-3}} dx_4 \dots dx_n \, p_x(x_1, x_2, \eta \dots, x_n)}{\displaystyle\int_{-\infty}^{\infty} d\eta \int_{\mathbb{R}^{n-3}} dx_4 \dots dx_n \, p_x(x_1, x_2, \eta, \dots, x_n)}, \\
&\vdots \\
T_n(x_1, \dots, x_n) &= \frac{\displaystyle\int_{-\infty}^{x_n} d\eta \, p_x(x_1, \dots, \eta)}{\displaystyle\int_{-\infty}^{\infty} d\eta \, p_x(x_1, \dots, \eta)}.
\end{aligned} \tag{94}$$

It is straightforward to check that $|\partial\xi/\partial x| = p_x(x_1, \dots x_n)$. Furthermore, the computation of the inverse requires the solution of a triangular system: find x_1 by solving the first equation of the system, substitute x_1 in the second equation, and solve it for x_2, substitute x_1 and x_2 in the third equation and solve for x_3, and so on, therefore the systems of equations is equivalent to n single nonlinear equations for $x_1, \dots, x_n$.

Remark 7. *This transformation is used to map an arbitrary measure with density $p_x(x)$ into the Lebesgue measure on the unit cube $[0,1]^n$. This can be used, for example, to approximate a continuous measure by a discrete measure, once a good discrete approximation of the Lebesgue measure is known.*

Let us assume that we have a "good" approximation of the uniform measure obtained by N suitably chosen points $\eta_{(i)} \in [0,1]^n$, $i = 1, \dots, N$,

$$\frac{1}{N}\sum_{i=1}^{N} \delta(\eta - \eta_{(i)}) \approx \chi_{[0,1]^n}(\eta),$$

then

$$\frac{1}{N}\sum_{i=1}^{N} \delta(x - x_{(i)}) \approx p_x(x),$$

where $x_{(i)} = T^{-1}(\eta_{(i)})$, $i = 1, \dots, N$. This technique can be effectively used to obtain good quadrature formulae to compute integrals in highly dimensional space, and is the basis of the so called quasi-Monte Carlo *integration.*

If the inverse transform map is too expensive, then the acceptance-rejection technique can be used, exactly as in the case of the monovariate distribution. More precisely, let x be a random variable with density $p_x(x)$, $x \in \mathbb{R}^n$. Then we look for a function $w(x) \geq p_x(x) \, \forall x \in \mathbb{R}^n$ which is "easy to sample". Let $A = \int_{\mathbb{R}^n} w(x)\, dx$. Then the algorithm works as follows

Algorithm 6

1. *sample x from $w(x)/A$ by any known method,*
2. *if $w(x)\xi < p_x(x)$ then accept the sample, else reject it and repeat step* 1.

As a first example we show how to sample from a Gaussian distribution in different ways.

Example 5. (Sampling from a Gaussian)

• *Let x be a normally distributed random variable with zero mean and unit variance,*

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

In order to sample from p one could invert the distribution function $P(x) = (1 + \operatorname{erf}(x/\sqrt{2}))/2$, where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt$$

denotes the error function. *However the inversion of the error function may be expensive.*

• *An alternative procedure is obtained by the so called* Box-Muller method *described below. Let us consider a two dimensional normally distributed random variable. Then*

$$p(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) = p(x)p(y).$$

If we use polar coordinates

$$x = \rho \cos\theta,\ y = \rho \sin\theta, \tag{95}$$

then we have

$$\frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) dx\, dy = \frac{1}{2\pi} \exp\left(-\frac{\rho^2}{2}\right) \rho\, d\rho\, d\theta.$$

Therefore in polar coordinates the density function is factorized as $p_\rho\, d\rho\, p_\theta\, d\theta$, with

$$p_\rho = \exp\left(-\frac{\rho^2}{2}\right) \rho, \quad \rho \geq 0$$

$$p_\theta = \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi$$

The random variables ρ and θ are readily sampled by inverting P_ρ and P_θ, i.e.

$$\rho = \sqrt{-2 \ln \xi_1},\ \theta = 2\pi\xi_2,$$

and, from these x and y are easily obtained.

• *A slightly more efficient way to sample from a* 2D *Maxwellian is obtained by avoiding the use of trigonometric functions.*

More precisely, the algorithm is obtained as follows and it is based on sampling from the unit circle using rejection (see Figure 22*).*

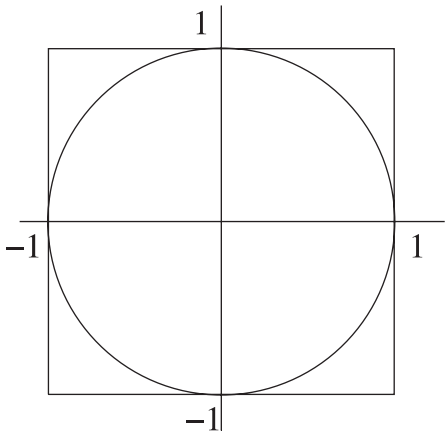

Fig. 22. – Square and unit circle

Algorithm 7 (Gauss 2)

1. *set* $v_x = 2\xi_1 - 1, \quad v_y = 2\xi_2 - 1, \quad s = v_x^2 + v_y^2$
2. *if* $= s < 1$ then *(accept)*
   $r = \sqrt{-\log s / s}$
   $v_x = v_x\, r, \quad v_y = v_y\, r$
   else *(reject) return to point* 1.

At the end of the procedure we have two points sampled from a Normal(0,1) distribution (i.e. a Gaussian distribution with zero mean and unit variance). Of course, if the random variable has mean μ and standard deviation σ, then x and y will be scaled accordingly as

$$x = \mu + \sigma\rho\cos\theta, \; y = \mu + \sigma\rho\sin\theta.$$

Finally we show how to sample a point uniformly from the surface of a sphere.

Example 6 (Sampling the unit sphere).
A point on a unit sphere is identified by the two polar angles (φ, θ),

$$\begin{aligned} x &= \sin\theta\cos\varphi, \\ y &= \sin\theta\sin\varphi, \\ z &= \cos\theta. \end{aligned}$$

Because the distribution is uniform, the probability of finding a point in a region is proportional to the solid angle

$$dP = \frac{d\omega}{4\pi} = \frac{\sin\theta\, d\theta}{2} \cdot \frac{d\varphi}{2\pi},$$

and therefore

$$\frac{d\varphi}{2\pi} = d\xi_1,$$

$$\frac{\sin\theta\, d\theta}{2} = d\xi_2.$$

Integrating the above expressions we have

$$\varphi = 2\pi\xi_1,$$

$$\theta = \arccos(1 - 2\xi_2).$$

7 - Direct Simulation Monte Carlo (DSMC) methods

In this paragraph we will describe the classical DSMC methods due to Bird and Nanbu in the case of the spatially homogeneous Boltzmann equation.

We assume that the kinetic equations we are considering can be written in the form

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon}\,[P(f,f) - \mu f], \tag{96}$$

where $\mu > 0$ is a constant and $P(f,f)$ is a non negative bilinear operator. In particular, for both Kac equation and Boltzmann equation for Maxwellian molecules we have $P(f,f) = Q^+(f,f)$.

Example 7.

- *For the Kac equation we have*

$$P(f,f)(v) = \frac{1}{2\pi}\int_0^{2\pi}\int_{-\infty}^{+\infty} f(v')\,f(v'_*)\,dv_*\,d\theta, \quad \mu = \rho = \int_{-\infty}^{\infty} f(v)\,dv. \tag{97}$$

therefore it it is of the form (96).

- *The Boltzmann equation for Maxwell molecules is also of similar form. We have, in fact,*

$$P(f,f)(v) = \int_{\mathbb{R}^3}\int_{S^2} b_0(\cos\theta)\,f(v')\,f(v'_*)\,d\omega\,dv_*. \tag{98}$$

The case of general kernels with cut-off will be discussed later.

7.1 - *Nanbu's method (*DSMC *no time counter)*

We assume that f is a probability density, i.e.

$$\rho = \int_{-\infty}^{+\infty} f(v,t)\,dv = 1.$$

Let us consider a time interval $[0, t_{\max}]$, and let us discretize it in n_{TOT} intervals of size Δt. Let us denote by $f^n(v)$ an approximation of $f(v, n\Delta t)$. The forward Euler scheme writes

$$f^{n+1} = \left(1 - \frac{\mu\Delta t}{\varepsilon}\right) f^n + \frac{\mu\Delta t}{\varepsilon} \frac{P(f^n, f^n)}{\mu}. \tag{99}$$

Clearly if f^n is a probability density then $P(f^n, f^n)/\mu$ is a probability density. Moreover if $\Delta t \leq \varepsilon/\mu$ ([1]) then also f^{n+1} is a probability density since it is a convex combination of probability densities.

From a physical point of view we can give the following interpretation to the above equation. A particle with velocity v_i will not collide with probability $(1 - \mu\Delta t/\varepsilon)$, and it will collide with probability $\mu\Delta t/\varepsilon$, according to the collision law described by $P(f^n, f^n)(v)$.

7.1.1 - Maxwellian case

Let us consider first kinetic equations for which $P(f, f) = Q^+(f, f)$, i.e. the collision kernel does not depend on the relative velocity of the particles.

An algorithm based on this probabilistic interpretation was proposed by Nanbu.

Algorithm 8 (Nanbu for Maxwell molecules)

1. *compute the initial velocity of the particles,* $\{v_i^0, i = 1, \ldots, N\}$,
   by sampling them from the initial density $f_0(v)$
2. for $n = 1$ to n_{TOT}
   for $i = 1$ to N
   with probability $1 - \mu\Delta t/\varepsilon$
   ∘ *set* $v_i^{n+1} = v_i^n$
   with probability $\mu\Delta t/\varepsilon$
   ∘ *select a random particle* j
   ∘ *compute* v_i' *by performing the collision between particle* i *and particle* j
   ∘ *assign* $v_i^{n+1} = v_i'$
   end for
   end for

([1]) This is related to (and is more restrictive than) the stability condition for Forward Euler method for evolutionary equations.

Note that in the algorithm we check for a possible collision of each individual particle. It would be more efficient to compute the expected number of collisions, and then to perform them. Furthermore, in this way we reduce the fluctuations in the number of collisions, obtaining a simulation with a reduced variance. The expected number of particles that collide in a small time step Δt is $N\mu\Delta t/\varepsilon$, and then the expected number of collision pairs is $N\mu\Delta t/(2\varepsilon)$.

The algorithm above is not exactly conservative, i.e. energy is conserved only in the mean, but not at each collision. A conservative version of the algorithm was introduced by Babovsky [4]. Instead of selecting single particles, independent particle pairs are selected, and conservation is maintained at each collision. A conservative version of Nanbu-Babovsky algorithm is illustrated below.

Algorithm 9 (Nanbu-Babovsky for Maxwell molecules)

1. *compute the initial velocity of the particles,* $\{v_i^0, i = 1, \dots, N\}$, *by sampling them from the initial density* $f_0(v)$
2. for $n = 1$ to n_{TOT}
   given $\{v_i^n, i = 1, \dots, N\}$
   - *set* $N_c = \mathrm{Iround}\,(\mu N \Delta t/(2\varepsilon))$
   - *select* N_c *pairs* (i, j) *uniformly among all possible pairs, and for those*
     - *perform the collision between* i *and* j, *and compute* v_i' *and* v_j' *according to the collision law*
     - *set* $v_i^{n+1} = v_i'$, $v_j^{n+1} = v_j'$
   - *set* $v_i^{n+1} = v_i^n$ *for all the particles that have not been selected*

   end for

In the above algorithm, the N_c pairs are selected by sampling $2N_c$ integers without repetition among the first N natural numbers. Here by $\mathrm{Iround}\,(x)$ we denote a suitable integer rounding of a positive real number x. For example

$$\mathrm{Iround}\,(x) = \begin{cases} \lfloor x \rfloor + 1 & \text{with probability} \quad x - \lfloor x \rfloor \\ \lfloor x \rfloor & \text{with probability} \quad \lfloor x \rfloor + 1 - x \end{cases}$$

where $\lfloor x \rfloor$ denotes the integer part of x.

Remark 8.

• *Clearly the above algorithm can be applied to the Kac equation and to the homogeneous Boltzmann equation with Maxwellian molecules. The only difference in the two cases consists in the computation of the post-collisional velocities.*

• *Note that in Nanbu-Babovsky algorithm we are forced to take collision by pairs. This may represent a problem since the CFL condition* $\Delta t \leq \varepsilon/\mu$ *does not guarantee that* $2N_c \leq N$ *when* N *is an odd number. This drawback, which can be significant when we have a small number of particles, can be overcome simply taking the slighty more restrictive CFL condition* $\Delta t \leq \varepsilon(N-1)/(\mu N)$ *when* N *is odd.*

• *The probabilistic interpretation breaks down if* $\Delta t/\varepsilon$ *is too large, because the coefficient of* f^n *on the right hand side may become negative. This implies that the time step becomes extremely small when approaching the fluid dynamic limit. Therefore Nanbu method becomes very inefficient near the fluid regime.*

Post-collisional velocities

When the above scheme is applied to the Kac equation, the new velocities v_i' and v_j' are computed as

$$v_i' = v_i \cos\theta - v_j \sin\theta, \quad v_j' = v_i \sin\theta + v_j \cos\theta,$$

where $\theta = 2\pi\xi$ and ξ denotes a random number, uniformly distributed in $[0,1]$.

For Maxwell molecules one has

$$v_i' = \frac{v_i + v_j}{2} + \frac{|v_i - v_j|}{2}\omega, \quad v_j' = \frac{v_i + v_j}{2} - \frac{|v_i - v_j|}{2}\omega, \tag{100}$$

where ω is chosen uniformly, according to:

2D

$$\omega = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \theta = 2\pi\xi, \tag{101}$$

3D

$$\omega = \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix}, \quad \theta = \arccos(2\xi_1 - 1), \quad \phi = 2\pi\xi_2, \tag{102}$$

where ξ_1, ξ_2 are uniformly distributed random variables in $[0,1]$.

7.1.2 - Variable Hard Sphere case

The Nanbu-Babovsky algorithm has to be modified when the scattering cross section is not constant. To this aim we shall assume that the collision kernel satisfies some cut-off hypothesis, which is essential from a numerical point of view.

We will denote by $Q_\Sigma(f,f)$ the collision operator obtained by replacing the kernel B with the kernel B_Σ

$$B_\Sigma(|v - v_*|) = \min\{B(|v - v_*|), \Sigma\}, \quad \Sigma > 0.$$

Thus, for a fixed Σ, let us consider the homogeneous problem

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon} Q_\Sigma(f,f). \tag{103}$$

The operator $Q_\Sigma(f,f)$ can be written in the form $P(f,f) - \mu f$ taking

$$P(f,f) = Q_\Sigma^+(f,f) + f(v) \int_{\mathbb{R}^3} \int_{S^2} [\Sigma - B_\Sigma(|v - v_*|)]\, f(v_*)\, d\omega\, dv_*, \tag{104}$$

with $\mu = 4\pi\Sigma\rho$ and

$$Q_\Sigma^+(f,f) = \int_{\mathbb{R}^3} \int_{S^2} B_\Sigma(|v - v_*|)\, f(v')\, f(v'_*)\, d\omega\, dv_*.$$

In this case, a simple scheme is obtained by using the acceptance-rejection technique to sample the post collisional velocity according to $P(f,f)/\mu$, where $\mu = 4\pi\Sigma\rho$ and Σ is an upper bound of the scattering cross section.

Such an approach, however, has the drawback of requiring a choice of Σ, and an estimate of the approximation introduced by Σ. In practical Monte Carlo simulations, such a problem is automatically solved, since one can compute an estimate of the maximum scattering cross section, based on the actual particle velocities.

The conservative DSMC algorithm for VHS collision kernels can be written as

Algorithm 10 (Nanbu-Babovsky for VHS molecules)

1. *compute the initial velocity of the particles,* $\{v_i^0, i = 1, \dots, N\}$, *by sampling them from the initial density* $f_0(v)$
2. for $n = 1$ to n_{TOT}
   given $\{v_i^n, i = 1, \dots, N\}$
   - *compute an upper bound* Σ *of the cross section*
   - *set* $N_c = \mathrm{Iround}\,(N\rho\Sigma\Delta t/(2\varepsilon))$
   - *select* N_c *dummy collision pairs* (i,j) *uniformly among all possible pairs, and for those*
     - *compute the relative cross section* $B_{ij} = B(|v_i - v_j|)$
     - if $\Sigma\,\xi < B_{ij}$
       - *perform the collision between* i *and* j, *and compute* v_i' *and* v_j' *according to the collisional law*
       - *set* $v_i^{n+1} = v_i'$, $v_j^{n+1} = v_j'$
   - *set* $v_i^{n+1} = v_i^n$ *for all the particles that have not collided*

   end for

7.1.3 - Computational considerations

The upper bound Σ should be chosen as small as possible, to avoid inefficient rejection, and it should be computed fast. The ideal choice of Σ would be

$$\Sigma = B_{\max} \equiv \max_{ij} B(|v_i - v_j|).$$

However, this choice is too expensive, since this computation would require an $O(N^2)$ operations.

An upper bound of $B_{\max}$ is obtained by taking $\Sigma = B(2\Delta v)$, where

$$\Delta v = \max_i |v_i - \bar{v}|, \quad \bar{v} := \frac{1}{N}\sum_i v_i. \tag{105}$$

For general collision kernel, the algorithm is slightly modified by introducing the angular dependence.

The cost of the method is proportional to the number of *dummy collision pairs*, that is $O(\mu N \Delta t/2)$. Thus for a fixed final time T the total cost is independent of the choice of $\Delta t = T/n$. However this is true only if we do not had to compute Σ (like in the Maxwellian case). In fact, the cost of the computation of the upper bound Σ is $O(N)$, which may be much larger than $O(\mu N \Delta t/2)$.

A possible way to overcome this difficulty is to update at each time step the value of the upper bound Σ only if it increases. This may be done as follows.

During the computation of the collision between particles i and j, let $\tilde{v}_i$ and $\tilde{v}_j$ denote the new particle velocities. Then the quantity Δv is updated according to

$$\Delta v = \max\left(\Delta v, |\tilde{v}_i - \bar{v}|, |\tilde{v}_j - \bar{v}|\right). \tag{106}$$

At the end of the collision loop, the upper bound on the cross section is computed as

$$\Sigma = \sigma(2\Delta v).$$

In space nonhomogeneous calculations, assuming that there are several collisional time steps during a convection time step, usually the estimate Σ is computed at the beginning of the collision step, and then updated with the formula above. Using this formula, however, Δv can not decrease, in contrast with what one expects when a system relaxes to equilibrium. This expression can become larger than necessary. For very small ε it could be worth computing it again, in order to have a better estimate.

7.2 - *Bird's method (DSMC time counter)*

The method is currently the most popular method for the numerical solution of the Boltzmann equation. It has been derived according to physical considerations for the simulation of particle collisions.

7.2.1 - Maxwellian case

Let us consider first the Maxwellian case. The average number of collisions in a short time step Δt is given by

$$N_c = \frac{N\mu\Delta t}{2\varepsilon}.$$

This means that the average time between collisions Δt_c is given by

$$\Delta t_c = \frac{\Delta t}{N_c} = \frac{2\varepsilon}{\mu N}.$$

Now it is possible to set a time counter, t_c, and to perform the calculation as follows

Algorithm 11 (Bird for Maxwell molecules)

1. *compute the initial velocity of the particles,* $\{v_i^0, i = 1, \dots, N\}$, *by sampling them from the initial density* $f_0(v)$
2. *set time counter* $t_c = 0$
3. *set* $\Delta t_c = 2\varepsilon/(\mu N)$
4. for $n = 1$ to n_{TOT}
   - repeat
     -- *select a random pair* (i, j) *uniformly within all possible* $N(N-1)/2$ *pairs*
     -- *perform the collision and produce* v_i', v_j'
     -- *set* $\tilde{v}_i = v_i', \tilde{v}_j = v_j'$
     -- *update the time counter* $t_c = t_c - \Delta t_c \log(\xi)$
     until $t_c \geq (n+1)\Delta t$
   - *set* $v_i^{n+1} = \tilde{v}_i, i = 1, \dots, N$

   end for

Remark 9.

• *The time counter is updated by a random time step, which has an exponential distribution, and an average* Δt_c. *A very similar result would be obtained if the time counter is updated by the average time step* Δt_c.

• *The above algorithm is very similar to the Nanbu-Babovsky (NB) scheme for Maxwellian molecules or for the Kac equation. The main difference is that in NB scheme the particles can collide only once per time step, while in Bird's scheme multiple collisions are allowed. This has a profound influence on the time accuracy of the method. In fact, while the solution of the NB scheme converges in probability to the solution of the time discrete Boltzmann equation, Bird's method converges to the solution of the space homogeneous Boltzmann equation. In this respect it may be considered a scheme of infinite order in time. For the space homogeneous Boltzmann equation, the time step* Δt, *in fact, can be chosen to be the full time span* $t_{\max}$.

7.2.2 - Variable Hard Sphere case

When a more general gas is considered, Bird's scheme has to be modified to take into account that the average number of collisions in a given time interval is not constant, and that the collision probability on all pairs is not uniform. This can be done as follows. The expected number of collisions in a time step Δt is given by

$$N_c = \frac{N\rho\overline{B}\Delta t}{2\varepsilon},$$

where $\overline{B}$ denotes the average collision frequency.

Once the expected number of collisions is computed, then the mean collision time can be computed as

$$\Delta t_c = \frac{\Delta t}{N_c} = \frac{2\varepsilon}{N\rho\overline{B}}.$$

The N_c collisions have to be performed with probability proportional to $B_{ij} = B(|v_i - v_j|)$. In order to do this, one can use the same acceptance-rejection technique as in Nanbu-Babovsky scheme. The drawback of this procedure is that computing $\overline{B}$ would be too expensive. A solution to this problem is to compute a local time counter Δt_c as follows. First select a collision pair (i,j) using rejection. Then compute

$$\Delta t_{ij} = \frac{2\varepsilon}{N\rho B_{ij}}.$$

This choice gives the correct expected value for the collision time

$$\Delta t_c = \sum_{1\le i<j\le N} \Delta t_{ij} \frac{B_{ij}}{\sum_{1\le i<j\le N} B_{ij}} = \frac{2\varepsilon}{N\rho\overline{B}}.$$

Bird's algorithm for general VHS molecules can therefore be summarized as:

Algorithm 12 (Bird for VHS molecules)

1. *compute the initial velocity of the particles,* $\{v_i^0, i = 1,\dots,N\}$, *by sampling them from the initial density* $f_0(v)$
2. *set time counter* $t_c = 0$
3. for $n = 1$ to n_{TOT}
   - *compute an upper bound* Σ *of the cross section*
   - repeat
     -- *select a random pair* (i,j) *uniformly within all possible* $N(N-1)/2$ *pairs*
     -- *compute the relative cross section* $B_{ij} = B(|v_i - v_j|)$
     -- if $\Sigma\,\xi < B_{ij}$

- *perform the collision between i and j, and compute v_i' and v_j' according to the collisional law*
- *set* $\tilde{v}_i = v_i', \tilde{v}_j = v_j'$
- *set* $\Delta t_{ij} = 2\varepsilon/(N\rho B_{ij})$
- *update the time counter* $t_c = t_c + \Delta t_{ij}$

until $t_c \geq (n+1)\Delta t$

∘ *set* $v_i^{n+1} = \tilde{v}_i, i = 1, \ldots, N$

end for

Remark 10.

• *At variance with NB scheme, there is no restriction on the time step Δt for Bird's scheme. For space homogeneous calculations Δt could be chosen to be the total computation time $t_{\max}$. However, the scheme requires an estimate of $B_{\max}$, and this has to be updated in time. This can be done either by performing the estimate at certain discrete time steps as in NB scheme, or by updating its value at every collision process. A possible solution in $O(1)$ operations is to check whether the new particles v_i', v_j' generated at each collision increase the quantity $\Delta v = \max_j |v_j - \bar{v}|$.*

• *As for Nanbu-Babovsky method even Bird's method becomes very expensive and practically unusable near the fluid regime. In this case the collision time between the particles Δt_{ij} becomes very small, and a huge number of collisions is needed in order to reach a fixed final time $t_{\max}$.*

8 - Time Relaxed discretizations

A different family of schemes can be constructed, with the purpose of obtaining a Monte Carlo method which is effective even near the fluid dynamic limit. Such schemes are based on a different approach for the time discretization of the equation. These schemes are based on the Wild sum expansion of the solution, and have been briefly described in Section 4.2, when discussing time discretization techniques for deterministic schemes. Here we give a more detailed analysis of such schemes.

8.1 - *Wild sum expansion*

It is the starting point for the construction of TR schemes.

Consider the differential system

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon}[P(f,f) - \mu f] \tag{107}$$

with initial condition

$$f(v,0) = f_0(v),$$

$\mu > 0$ is a constant and P a bilinear operator.

It is possible to show that the function f satisfies the following formal expansion [**132**]

$$f(v,t) = e^{-\mu t/\varepsilon} \sum_{k=0}^{\infty} \left(1 - e^{-\mu t/\varepsilon}\right)^k f_k(v). \tag{108}$$

The functions f_k are given by the recurrence formula for $k = 0, 1, \ldots$

$$f_{k+1}(v) = \frac{1}{k+1} \sum_{h=0}^{k} \frac{1}{\mu} P(f_h, f_{k-h}). \tag{109}$$

In order to prove such an expansion, we observe that (here we assume that $\varepsilon = 1$)

$$\frac{\partial}{\partial t}\left(e^{\mu t} f\right) = e^{\mu t}\left(\frac{\partial f}{\partial t} + \mu f\right) = e^{\mu t} P(f,f).$$

Now let us define $F(v,\tau) = e^{\mu t} f$, with $\tau = 1 - e^{-\mu t}$. It is

$$\frac{\partial F}{\partial \tau} = \frac{\partial}{\partial t}\left(e^{\mu t} f\right)\frac{dt}{d\tau} = e^{\mu t}\,\frac{P(f,f)}{\mu}.$$

Using the bilinearity of operator P, one has

$$\frac{\partial F}{\partial \tau} = \frac{1}{\mu} P(F,F). \tag{110}$$

Then one can write the Taylor series in τ of the function F. The easiest way is to assume a formal power series

$$F = \sum_{k=0}^{\infty} f_k \tau^k. \tag{111}$$

By inserting this expansion into Eq. (110), one has

$$\sum_{k=0}^{\infty} (1+k) f_{k+1} \tau^k = \frac{1}{\mu} P\Big(\sum_{h=0}^{\infty} f_h \tau^h, \sum_{m=0}^{\infty} f_m \tau^m\Big)$$

and, equating the coefficients in τ^k, one obtains the recurrence formulas (109). Expressing f in terms of F, and making use of the expansion (111), one obtains the Wild expansion (108).

The coefficient functions f_k that appear in the Wild sum expansion have several interesting properties, that make them suitable for the construction of numerical methods that preserve some of the qualitative properties of the exact solution.

Properties of the coefficients f_k

i) *Conservation*: If the collision operator preserves some moments, then all the functions f_k will have the same moments, i.e. if, for some function $\phi(v)$

$$\int_{\mathbb{R}^3} P(f,f)\phi(v)\,dv = \mu \int_{\mathbb{R}^3} f\,\phi(v)\,dv, \tag{112}$$

then the coefficients f_k defined by (109) are nonnegative and satisfy $\forall\, k > 0$

$$\int_{\mathbb{R}^3} f_k\phi(v)\,dv = \int_{\mathbb{R}^3} f_0\phi(v)\,dv.$$

ii) *Asymptotic behavior*: If the sequence $\{f_k(v)\}_{k\geq 0}$ defined by (109) is convergent, then (108) is well defined for any value of ε. Moreover, if we denote by $M(v) = \lim_{k\to\infty} f_k$ then

$$\lim_{t\to\infty} f(v,t) = M(v).$$

Here $M(v)$ is the local (Maxwellian) equilibrium.

If $B(|v - v_1|, \omega) \leq \Sigma$ is a bounded scattering kernel section then the BE can be always written in the above form, where $\mu = 4\pi\rho\Sigma$ is a positive constant s.t.

$$\mu \geq L_\Sigma[f](v) \equiv \int_{\mathbb{R}^3}\int_{S^2} B(|v - v_1|, \omega)\, f(v_*)\, d\omega\, dv_*$$

and $P(f,f) = Q^+_\Sigma(f,f) + (\mu - L_\Sigma[f])\, f$.

8.2 - *Time relaxed (TR) schemes*

From the previous representation, it appears natural to introduce the following class of numerical schemes for the time discretization of equation (107) [**51**], [**105**].

$$f^{n+1}(v) = (1-\tau)\sum_{k=0}^{m} \tau^k f_k^n(v) + \tau^{m+1} M(v), \tag{113}$$

where $f^n = f(n\Delta t)$, Δt is a small time interval, and $\tau = 1 - e^{-\mu\Delta t/\varepsilon}$ (*relaxed time*).

The schemes are obtained by truncating the Wild series (108), and by replacing the reminder with a Maxwellian, as suggested by the asymptotic behavior of the Wild sum coefficients f_k. Such schemes have the following properties:

i) *Conservation*: If the initial condition f^0 is a non negative function, then f^{n+1} is nonnegative for any $\mu\Delta t/\varepsilon$, and satisfies

$$\int_{R^3} f^{n+1}\phi(v)\,dv = \int_{R^3} f^n\phi(v)\,dv.$$

ii) *Asymptotic preservation*: For any $m \geq 1$, we have

$$\lim_{\mu\Delta t/\varepsilon\to\infty} f^{n+1} = M(v). \tag{114}$$

iii) *Accuracy*: If $\sup_{k>m}\{|f_k^n - M|\} \leq C$ for some constant $C = C(v)$ then

$$|f(v,\Delta t) - f^1(v)| \leq C(v)\tau^{m+1}.$$

The first two properties are obvious. The third can be proven by observing that

$$(1-\tau)\sum_{k=m+1}^{\infty}\tau^k = \tau^{m+1}$$

and therefore

$$|f(v,\Delta t) - f^1(v)| = (1-\tau)\left|\sum_{k=m+1}^{\infty}(f_k - M)\tau^k\right| \leq (1-\tau)\sum_{k=m+1}^{\infty}|f_k - M|\tau^k \leq C(v)\tau^{m+1}.$$

Finally le us notice that the case $m = 0$ is equivalent to the BGK approximation of the Boltzmann equation [**6**].

8.3 - *Generalizations*

Generalized TR schemes can be derived, of the following form

$$f^{n+1}(v) = \sum_{k=0}^{m} A_k f_k^n(v) + A_{m+1}M(v).$$

The weights $A_k = A_k(\tau)$ are nonnegative function that satisfy the following properties

i) *Conservation*:

$$\sum_{k=0}^{m+1} A_k(\tau) = 1 \quad \tau \in [0,1],$$

ii) *Asymptotic preservation*:

$$\lim_{\tau\to 1} A_k(\tau) = 0, \quad k = 0,\dots,m$$

iii) *Consistency*:

$$\lim_{\tau\to 0} A_1(\tau)/\tau = 1, \quad \lim_{\tau\to 0} A_k(\tau)/\tau = 0, \quad k = 2,\dots,m+1.$$

Some Generalized TR schemes have been considered in [**105**], [**104**]. However, what is the "best" choice of the functions $A_k(\tau)$ is still an open problem, and subject of research. Quite remarkably, this optimization problem can be solved for Maxwellian molecules. Explicit expressions of the weight for the generalized first order TRMC scheme for the Kac equation have been obtained recently in [**29**] and read

$$(115)\quad A_0(t)=e^{-\mu\Delta t/\varepsilon},\ A_1(t)=\frac{4}{3}(e^{-\mu\Delta t/(4\varepsilon)}-e^{-\mu\Delta t/\varepsilon}),\ A_2(t)=1+\frac{1}{3}(e^{-\mu\Delta t/\varepsilon}-4e^{-\mu\Delta t/(4\varepsilon)}).$$

9 - Time Relaxed Monte Carlo (TRMC) methods

Just as the probabilistic interpretation of the Forward-Euler scheme for the Boltzmann equation leads to the Nanbu-Babovsky scheme, so the probabilistic interpretation of the TR time discretization yields the Time Relaxed Monte Carlo Methods.

In this section we shall describe how to construct first, second and third order TRMC schemes, while we leave to a future section the description of infinite order TR schemes.

As we shall see, the TR approach will lead to schemes that give basically the same results as Nanbu-Babovsky scheme in cases in which the Knudsen number is not too small. Their advantage becomes evident in those regions in which the system is near the fluid dynamic limit. In such cases, the asymptotic property of TRMC schemes guarantees the correct fluid dynamic limit without the excessive cost required by standard Monte Carlo methods.

9.1 - *First order TR scheme (TRMC1)*

For $m = 1$ the generalized TR scheme writes

$$f^{n+1} = A_0 f^n + A_1 f_1 + A_2 M.$$

The probabilistic interpretation of the above equation is clear since f^n, $f_1 = P(f,f)/\mu$ and M are probability densities and then also f^{n+1} is a probability density since it is a convex combination of them independently of Δt. From a physical viewpoint we have that a particle extracted from f^n

- does not collide with probability A_0,
- collides with another particle extracted from f^n with probability A_1,
- is thermalized, namely is replaced by a particle sampled from a Maxwellian, with probability A_2.

Sampling two particles from f_1 is equivalent to perform a *dummy collision*, and it can be done exactly as in the case of the Nanbu-Babovsky scheme.

As in the case of NB scheme, it is preferable to compute the expected number of collision pair, $< N_c >= NA_1/2$, and perform the collisions among those pairs, in order to maintain conservation of momentum and energy.

The conservative version of the methods can be formalized in the following algorithm

Algorithm 13 (First order TRMC for VHS molecules)

1. *compute the initial velocity of the particles,* $\{v_i^0, i = 1, \dots, N\}$, *by sampling them from the initial density* $f_0(v)$
2. for $n = 1$ to n_{TOT}
   given $\{v_i^n, i = 1, \dots, N\}$,
   - *compute an upper bound* Σ *of the cross section*
   - *set* $\tau = 1 - \exp(-\mu \Delta t/\varepsilon)$
   - *compute* $A_1(\tau)$, $A_2(\tau)$
   - *set* $N_c = \mathrm{Iround}(NA_1/2)$
   - *perform* N_c *dummy collisions, as in Algorithm (*10*)*
   - *set* $N_1 = N - 2N_c$ *and* $N_M = \mathrm{Iround}(N_1 A_2/(A_0 + A_2))$
   - *select* N_M *particles among those that have not collided, and compute their mean momentum and energy*
   - *sample* N_M *particles from the Maxwellian with the above momentum and energy, and replace the* N_M *selected particles with the sampled ones*
   - *set* $v_i^{n+1} = v_i^n$ *for all the* $N - 2N_c - N_M$ *particles that have not been selected*

   end for

In this formulation the probabilistic interpretation holds uniformly in $\mu\Delta t$, at variance with NB, which requires $\mu\Delta t < 1$. Note that since we take collisions by pairs in intermediate regimes it may happen that $2N_c > N$ when N is odd. In this case a slightly smaller time step, such that $A_1 \leq (N-1)/N$ has to be chosen. This is particularly relevant when we use a small number of particles.

9.1.1 - Particle Euler solvers

As $\mu\Delta t \to \infty$, the distribution at time $n+1$ is sampled from a Maxwellian. In a space non homogeneous case, this would be equivalent to a stochastic particle method for Euler equations [**116**]. The sampling from the Maxwellian requires some explanation. If we replace a certain number of particles (approximately NA_2 for TRMC1) by particles sampled from a Maxwellian, the momentum and energy of the particles will be conserved only on average.

If we want to be exactly conserved we can proceed as follows. First select $N_M = 2\,\mathrm{Iround}\,(NA_2/2)$ particles randomly and uniformly among the $N - 2N_c$ particles, and compute their moments (mean velocity u and variance $2T$) (at least two particles are needed to have a non zero sample variance). Then sample from a Maxwellian with the same mean and variance. Finally, scale the sampled particles, so that mean and variance of the sample are u and $2T$ [**112**]. One may argue that, by doing so, one is introducing some correlation among the particles, so that the sampled velocities will not be independent. As an alternative, once N_M particles have been selected, one can use an algorithm to sample a given number of normally distributed particles, with assigned moments u and $2RT$ of the sample. Such procedure has been introduced by Pullin [**117**], and used in the TRMC codes that produced the numerical results shown in the next section.

9.2 - *Second order TR scheme (TRMC2)*

A second order scheme in time can be obtained starting from the generalized TR schemes with $m = 2$

$$f^{n+1} = A_0 f^n + A_1 f_1 + A_2 f_2 + A_3 M,$$

with $f_1 = P(f^n, f^n)/\mu$, $f_2 = P(f^n, f_1)/\mu$. Since also f_2 is a probability density, again we have a clear probabilistic interpretation of f^{n+1} as a convex combination of probability densities. From a particle viewpoint we can say that a particle extracted from f^n

- does not collide with probability A_0,
- collides with another particle extracted from f^n with probability $A_1 + A_2$,
  - if the collision takes place then the particle collides again with another particle extracted from f^0 with probability $A_2/(A_1 + A_2)$,
- is thermalized, namely is replaced by a particle sampled from a Maxwellian, with probability A_3.

The algorithm now reads

Algorithm 14 (second order TRMC for VHS molecules)

- compute the initial velocity of the particles, $\{v_i^0, i = 1, \ldots, N\}$, by sampling them from the initial density $f_0(v)$
- for $n = 1$ to n_{TOT}
  given $\{v_i^n, i = 1, \ldots, N\}$,
  - *compute an upper bound* Σ *of the cross section*
  - *set* $\tau = 1 - \exp(-\mu \Delta t/\varepsilon)$

- *compute* $A_1(\tau)$, $A_2(\tau)$, $A_3(\tau)$
- *set* $N_c = \mathrm{Iround}\,(N(A_1+A_2)/2)$, $N_2 = \mathrm{Iround}\,(N_c A_2/(A_1+A_2))$
- *perform* N_c *dummy collisions, as in Algorithm (10), and store* N_2 *of them*
- *select* N_2 *particles from* f^n
- *perform the dummy collisions, as in Algorithm 10, of these selected particles with the stored set of* N_2 *particles that have collided once*
- *set* $N_1 = N - 2N_c$ *and* $N_M = \mathrm{Iround}\,(N_1 A_3/(A_0+A_3))$
- *replace* N_M *particles with samples from Maxwellian, as in Algorithm 13*
- *set* $v_i^{n+1} = v_i^n$ *for all the* $N - 2N_c - N_M$ *particles that have not been selected*

end for

Similarly, higher order TRMC methods can be constructed. For example, a third order scheme is obtained from

$$f^{n+1} = A_0 f^n + A_1 f_1 + A_2 f_2 + A_3 f_3 + A_4 M,$$

with $f_1 = P(f^n, f^n)/\mu, f_2 = P(f^n, f_1)/\mu$ and $f_3 = [2P(f^n, f_2) + P(f_1, f_1)]/(3\mu)$.

9.3 - *Numerical results*

In this section we report some numerical results for the space-homogeneous Boltzmann equation, obtained by some of the numerical schemes described above. In particular, we compare Nanbu-Babovsky (NB) scheme with time relaxed Monte Carlo schemes of various orders.

As a test problem we use the exact solution for the Kac equation [**10**], [**76**]

$$f(v,t) = \frac{1}{2C^{3/2}}\left[\frac{3}{2}(C-1) + (3-C)\frac{v^2}{C^2}\right]\exp(-v^2/C),$$

with $C(t) = 3 - 2\exp(-\sqrt{\pi}t/16)$.

The density function f has been reconstructed on a regular grid, by convolving the particle distribution by a suitable mollifier

$$f(V_I) = \frac{1}{N}\sum_{j=1}^{N} W_H(V_I - v_j), \tag{116}$$

where $\{V_I = V_{\min} + I\Delta V, I = 1, \ldots, N_g\}$.

The smoothing function W_{H} is given by

$$W_{\mathrm{H}}(x) = \frac{1}{H}W\left(\frac{x}{H}\right), \quad W(x) = \begin{cases} 3/4 - x^2 & \text{if } |x| \le 0.5, \\ (x-3/2)^2/2 & \text{if } 0.5 < |x| \le 1.5, \\ 0 & \text{otherwise.} \end{cases}$$

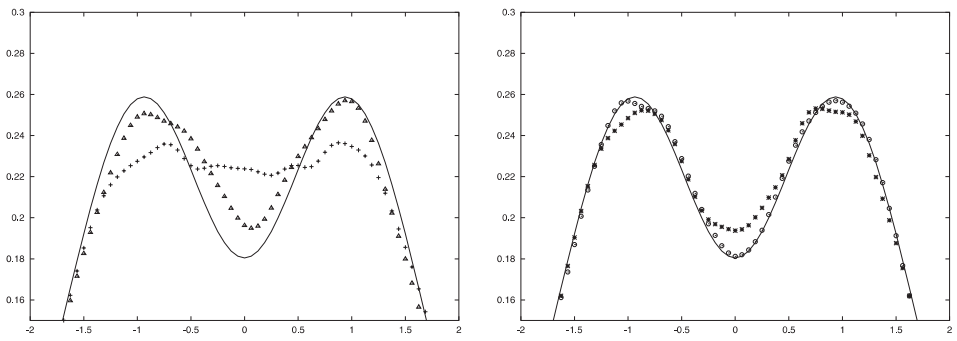

Fig. 23. – Kac equation: Details of the distribution function at time $t = 2.0$ for $\Delta t = 1.0$. Exact (line), NB(+), first order TRMC ($\triangle$), second order TRMC ($*$), third order TRMC ($\circ$)

The value $H = 0.2$ has been selected as a good compromise between fluctuations and resolution. The simulations are performed with $N = 5 \times 10^4$ particles.

Next we perform the same kind of accuracy test in the case of the Boltzmann equation for Maxwell molecules in the two-dimensional velocity case. An exact solution to this equation is given by (74). In this case the comparison with the exact solution is obtained by reconstructing the function on a regular grid of spacing $\Delta v = 0.25$ by the "weighted area rule" (i.e. bilinear interpolation).

All the simulations have been performed for $t \in [0, 16]$ by starting with $N = 10^5$ particles.

In Fig. 25 we show the L^2 norm of the error in time for both NB and TRMC schemes. In the first picture we report the results obtained with the same time step $\Delta t = 1.0$. The results confirm the gain of accuracy of the second order TRMC method on the transient time scale.

Using a time step of $\Delta t = 1.0$ for the second order TRMC method and $\Delta t = 0.5$ for the NB method the solutions are almost equivalent.

10 - Space non homogeneous case

The solution of the Boltzmann equation is performed in two separate steps, using a splitting approach. The computational domain is divided into a certain number N_c of cells. In each cells there are particles with position and velocity (x_i, v_i), $x_i \in \mathbb{R}^d$, $v_i \in \mathbb{R}^3$. The time interval of the computation $t_{\max}$ is divided into intervals of size Δt. Let us denote by

$$f^n(x, v) = m \sum_{i=1}^{N} \delta(x - x_i^n)\delta(v - v_i^n),$$

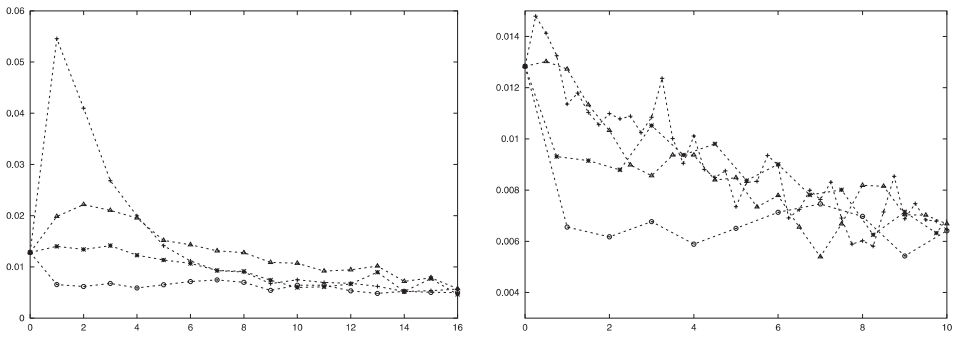

Fig. 24. – Kac equation: L^2 norm of the error vs time. NB(+), first order TRMC ($\triangle$), second order TRMC ($*$), third order TRMC ($\circ$). Left: Time step $\Delta t = 1.0$. Right: NB with $\Delta t = 0.25$, first order TRMC with $\Delta t = 0.5$, second order TRMC with $\Delta t = 0.75$, third order TRMC with $\Delta t = 1.0$

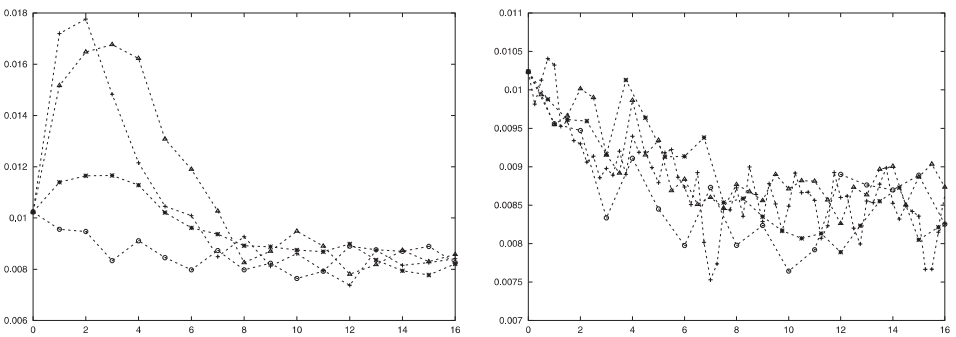

Fig. 25. – Maxwellian case, 2D in velocity: L^2 norm of the error vs time. NB (+), first order TRMC ($\triangle$), second order TRMC ($*$), third order TRMC ($\circ$). Left: Time step $\Delta t = 1.0$. Right: NB with $\Delta t = 0.25$, first order TRMC with $\Delta t = 0.5$, second order TRMC with $\Delta t = 0.75$, third order TRMC3 with $\Delta t = 1.0$

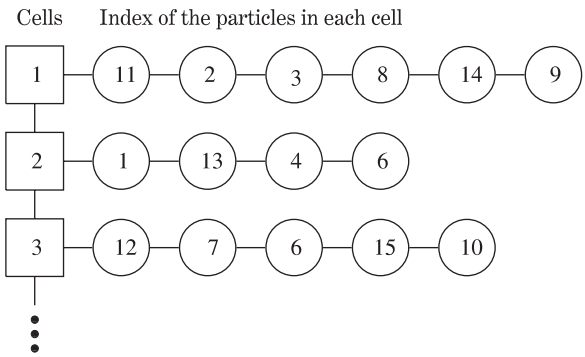

Fig. 26. – An efficient way to store the index of the particles belonging to each cell is through the use of a linked list. The latter, in turn, is efficiently managed by an array of indices: the indices from $I(j)$ to $I(j+1)-1$ correspond to the indices of the particles belonging to cell j

the approximation of the distribution function at time $n\Delta t$. It is completely determined by position and velocity of the particles: $\{(x_i, v_i),\ i = 1, \ldots, N\}$.

To compute (x_i^{n+1}, v_i^{n+1}) we first apply the collision step. In each cell j, we perform the collisions of all particles which belong to cell j, by one of the techniques seen before, for a time Δt. An easy way to find the particles which belong to a given cell j in 1D is to sort all particles according to their position, before starting the collision step, and to store, for each cell, the index of the last particle of the cell. For more complicated geometries, the indices of the particles associated to each cell can be stored in a list (see Fig. 26). Clearly the collision step will change the velocities of the particles, leaving their position unchanged.

10.1 - *Transport step*

Each particle moves according to free flow for a time step Δt. A tentative position $\tilde{x}_i^{n+1}$ is computed

$$\tilde{x}_i^{n+1} = x_i^n + \hat{v}_i^{n+1}\Delta t,$$

where $\hat{v}$ denotes the particle velocity after collision. If the particle does not exit the computational domain, then

$$x_i^{n+1} = \tilde{x}_i^{n+1}, \quad v_i^{n+1} = \hat{v}_i^{n+1},$$

otherwise boundary conditions are applied.

The convection time step Δt is the same for all particles. It is assigned in such a way that the fast particles travel approximately one cell per time step. In each cell, the space homogeneous Boltzmann equation is solved for one time step Δt. This can be obtained in a single step ($\Delta t_c = \Delta t$), or, more generally, with several steps. In the method of Bird, the collision step ends when the time counter becomes larger than $t + \Delta t$.

Typically, a single step is possible if the Knudsen number is very large, or, in the case of TR schemes, if it is extremely small. For intermediate cases one has $\Delta t_c = \Delta t/n_t$, where n_t is chosen by accuracy and positivity considerations.

10.1.1 - Boundary conditions

Inflow and outflow boundary conditions

If a particle leaves the computational domain (absorbing boundary) it is just deleted from the list of particles, and the particle number is decreased by one.

In some simulations (e.g. shock wave calculation, flow past a body) particles have to be injected into the domain, in order to simulate the inflow of matter from regions external to the computational domain.

In this case one assumes that the distribution function of the incoming particles is known.

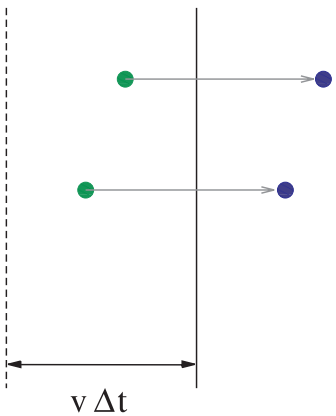

Fig. 27. – Inflow boundary condition on the left boundary: the distribution outside the computational domain is assumed to be a Maxwellian. The injected particles are sampled according to the flux of incoming particles

Let us consider a simple 2D problem. Assume that the computational domain is a 2D rectangle (see Figure 27). In this case one assumes that particles enter from the left part of the domain, with a distribution function which is a Maxwellian with mean velocity $u = (u_{Lx}, 0, 0)$, density ρ_L, and temperature T_L. Let is we by $f_L(v)$ such a Maxwellian. Notice that if the problem is 1D in space, then $\int f_L(v)\,dv$ denotes the density of particles per unit length, if it is 2D in space $\int f_L(v)\,dv$ denotes the density of particles per unit surface.

The expected number of particles which enter the domain in a time step Δt, through one cell edge of size Δx, will be equal to

$$< N_p >= \Delta t \Delta x \int\limits_{v_x>0} v f_L(v)\,dv.$$

To reduce fluctuations it is better to sample the particles that enter from each cell edge, rather than sampling a total number of incoming particles entering from a random position of the left boundary. Then one can chose an integer $N_p = \mathrm{Iround}\,(<N_p>)$, samples N_p particles from the Maxwellian f_L, and inject them into the computational domain.

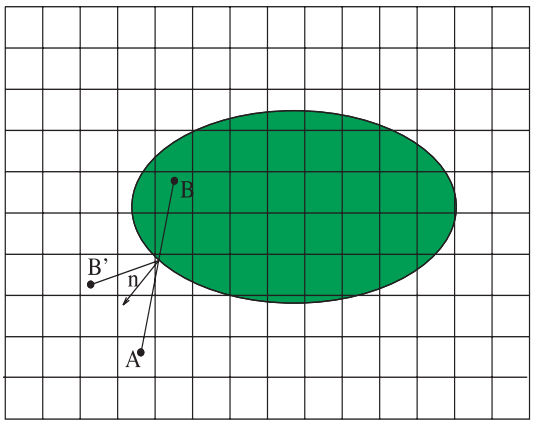

Fig. 28. – Cells subdivision of the computational domain

The distribution of sampled particles is obtained by the following consideration. In a time step Δt, particles with x-velocity v_x will enter the domain only if they lie in a strip of width $v_x \Delta t$. Therefore the density of particles entering the domain in a time span Δt with velocity v has to be proportional to

$$|v_x| f_L(v),$$

and not just to $f_L(v)$.

Solid walls

If the particle hits a solid wall, then it is injected into the domain according to the scattering law of a solid wall. The most commonly used boundary conditions are specular reflection, diffusive boundary condition, or a convex combination of the two. For diffusive boundary condition, the particle is emitted according to a (half) Maxwellian distribution, with zero mean, and temperature equal to the (prescribed) wall temperature.

This is easily obtained as follows (see Figure 28). The normal component is sampled from a distribution proportional to the flux, as in the case of inflow boundary conditions, while the two transversal components are sampled from a 2D Maxwellian, with mean velocity equal to the transverse wall velocity, and temperature equal to the wall temperature. In practice, one considers an auxiliary systems of coordinates in which the z-component is normal to the surface. One samples the z-component by 1D Maxwellian flux with zero mean ([2]), and the x and y component

([2]) We assume that the boundaries of the computational domain do not change, therefore in the frame of reference of the computational domain, the normal component of the wall is zero. If this is not true, then the normal component of the wall velocity has to be taken into account in the sampling.

from a 2D Maxwellian. A final rotation of coordinates will assign the three components of the velocity to the laboratory frame.

The location $\tilde{x}$ (and time $\tilde{t}$) of the impact between the particle and the boundary can be obtained by solution a nonlinear scalar equation. With reference to Figure 28, for example,

$$\phi(\zeta A + (1-\zeta)B) = 0,\ \zeta \in \mathbb{R}, A, B \in \mathbb{R}^2$$

if $\phi(P) = 0, P \in \mathbb{R}^2$, denotes the boundary of the wall. Then we set $\tilde{x} = \zeta A + (1-\zeta)B$.

After the particles is generated with the new velocity $\tilde{v}$, then the new location and velocity are set as

$$x_i^{n+1} = \tilde{x} + \tilde{v}(t_{n+1} - \tilde{t}).$$

Remark 11. *Monte Carlo methods can be used for both time dependent and stationary problems. In both cases, some averaging process is needed in order to decrease the fluctuations intrinsic in the stochastic nature of the method. They can be computed by solving the time dependent problem, and waiting until the system is close to the stationary state (a more efficient technique are based on the application of MC to the solution of the stationary problem directly, but we do not consider this approach here).*

In the case of stationary problems, time average can be used to improve the statistics, and decrease the fluctuation. Typically one sets two times: t_s and $t_{\max}$. Assuming that the solution reached the steady state at time $t = t_s$, then the rest of the time is used to accumulate statistics, and improve the estimation of the parameters we are interested in the calculation (typically moments of the distribution function, or sometimes the distribution function itself).

In the case of time dependent problem, this procedure can not be used, and one has to perform several runs of the same problem, with the same data, but with a different pseudo-random number sequence. Because of this reason, MC method is more effective for stationary problems than for time dependent ones.

10.2 - *Numerical results*

10.2.1 - 1D Shock wave profiles

As a prototype problem, we illustrate how to use MC methods for the computation of 1D shock wave profiles. Let us consider a stationary 1D shock travelling with constant speed V_S.

At the left and at the right of the shock one assumes that the gas is in local equilibrium, distributed according to a Maxwellian with parameters, respectively, (ρ_L, u_L, T_L), and (ρ_R, u_R, T_R).

These parameters and the shock speed V_S are not arbitrary, but satisfy the so called jump conditions, also known as Rankine-Hugoniot conditions [**131**].

Let $V = V_S - u$. Then the states on the far left and on the far right satisfy the condition

$$[\rho V] = 0$$

$$[p + \rho V^2] = 0$$

$$[V(\rho e + \frac{1}{2}\rho V + p)] = 0$$

where $\forall h$, $[h] \equiv h_L - h_R$ denotes the difference between the left and the right state, $p = R\rho T$ denotes the pressure, and e is the internal energy. For a monatomic gas $e = (3/2)RT$. Here $R = k_B/m$ is the gas constant, and k_B is the Boltzmann constant.

For a more general polytropic gas one has $e = RT/(\gamma - 1)$, where $\gamma = c_p/c_v$ is the polytropic constant, c_p and c_v are specific heat at constant pressure and volume respectively (see Table 8). For a perfect gas one has $c_p = c_v + R$, and

$$c_v = \frac{n_d}{2}R,$$

where n_d is the number or excited degrees of freedom of the molecule.

TABLE 8. – *Excited degree of freedom and polytropic constant of common gases*

n_d	gas	γ
3	monatomic gas in 3D	5/3
2	2D model of monatomic gas	2
5	biatomic gas (air)	7/5

Jump conditions are very general. They are a consequence of conservation laws, and they do not depend on the detailed relaxation process that happens in the shock zone.

For a polytropic gas the solutions of the Rankine-Hugoniot conditions can be expressed in terms of the Mach number

$$\mathcal{M} \equiv \frac{V_S - u_R}{a_R}$$

where $a = \sqrt{\gamma p/\rho}$ denotes the sound speed.

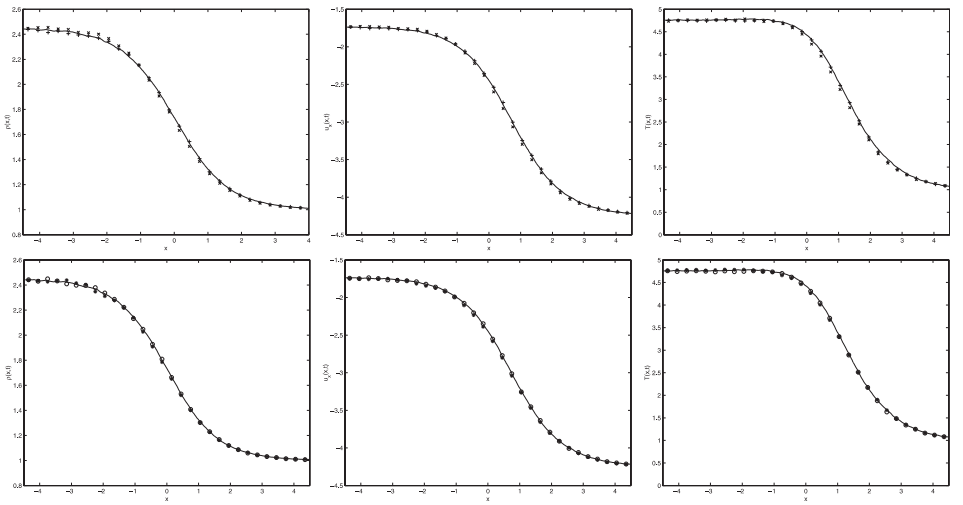

Fig. 29. – 1D shock profile: NB(+) and first order TRMC (×) (top), second order (∗) and third order (∘) TRMC (bottom) for $\varepsilon = 1.0$ and $\Delta t = 0.025$. From left to right: ρ, u, T. The line is the reference solution

Assuming one knows the state on the right of the shock (unperturbed state), and the Mach number, then the state on the left can be computed from the relations

$$\frac{u_L - u_R}{a_R} = \frac{2(\mathcal{M}^2 - 1)}{(\gamma + 1)\mathcal{M}},$$

$$\frac{\rho_L}{\rho_R} = \frac{(\gamma + 1)\mathcal{M}^2}{(\gamma - 1)\mathcal{M}^2 + 2},$$

$$\frac{p_L - p_R}{p_L} = \frac{2\gamma(\mathcal{M}^2 - 1)}{\gamma + 1}.$$

At each time step, the expected number of particles entering the domain is computed by integrating the flux over the interesting range of velocities. For example, on the left the average number of particles entering the domain in a time step Δt is given by

$$< N_L >= \int_0^\infty v_x M_L(v)\, dv \frac{\Delta t}{\Delta x} \bar{N}_L$$

where $\bar{N}_L$ is the number of particles per cell of the equilibrium distribution which is on the left of the computational domain. The theoretical fluxes

$$F_L = \int_0^\infty v_x M_L(v)\, dv, \quad F_R = \int_{-\infty}^0 |v_x| M_R(v)\, dv$$

are computed once at the beginning of the calculation.

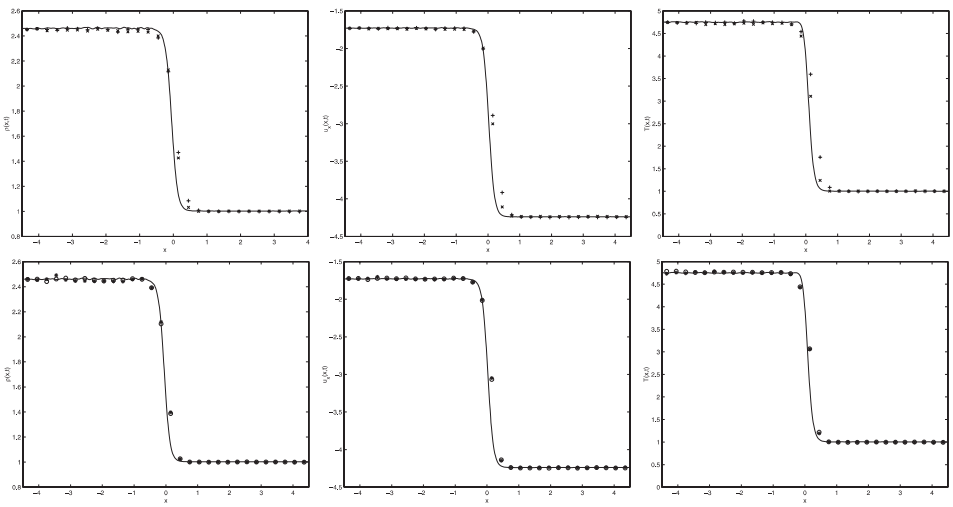

Fig. 30. – 1D shock profile: NB(+) and first order TRMC (×) (top), second order (∗) and third order (∘) TRMC (bottom) for $\varepsilon = 0.1$ and $\Delta t = 0.0025$ for DSMC, $\Delta t = 0.025$ for TRMC. From left to right: ρ, u, T. The line is the reference solution

A similar expression can be used for the particles entering from the right. The integral $\int_0^\infty v_x M_L(v)\, dv$ can be computed analytically.

The average is then converted into an integer: $N_L = \text{Iround}(< N_L >)$. The particles are injected with a velocity whose x component is sampled from $v_x M_L(v)$, and the other two components of the velocity are sampled from a Maxwellian with zero mean and temperature T_L. The position is given by $x = v_x \Delta t \xi$, where ξ is a uniformly distributed random number in $[0, 1]$. The presence of ξ is due to the fact that the time at which the particle enters the domain is uniformly distributed in $[t_n, t_n + \Delta t]$.

A similar procedure is used for the right boundary.

We compare here the following schemes: NB, TRMC1, TRMC2, TRMC2

The initial data is assumed to be a local Maxwellian $f(x, v, t) = M(\rho, u, T)$, in the whole computational domain, with

$$\rho = 1.0, \quad T = 1.0, \qquad x > 0.$$

The Mach number is $\mathcal{M} = 3.0$. The mean velocity is

$$u_x = -\mathcal{M}\sqrt{(\gamma T)}, \quad u_y = 0, \quad u_z = 0,$$

with $\gamma = 2$ since we consider a 2D velocity space.

The values for ρ, u and T for $x < 0$ are given by the Rankine-Hugoniot conditions.

In our test we use hard sphere molecules with 50 space cells on $[-7.5, 7.5]$ and 500 particles in each cell on $x > 0$. After the stationary stat has been reached the

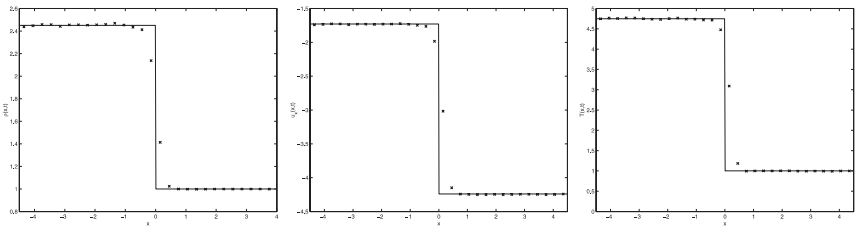

Fig. 31. – 1D shock profile: First order TRMC ($\times$) for $\varepsilon = 10^{-6}$ and $\Delta t = 0.025$. From left to right: ρ, u, T

solution is averaged for 1500 time steps. The reference solution is obtained with 200 space cells and 500 particles in each cell on $x > 0$, averaged for 8000 time steps.

The collision time step Δt_c is assigned empirically by $\Delta t_c = \Delta t / n_t$, with $n_t = 1, 10$ respectively for $\varepsilon = 1, 0.1$, in the case of NB schemes. For the TRMC scheme the collision step has been taken equal to the convection step. Observe that all the methods give essentially the same results for the shock profile. However, time relaxed methods are much more efficient for very small values of the parameter ε. The results shown in Figure 31 are produced only by first order TRMC, while the continuous line represents the shock profile for the limit $\varepsilon = 0$. It would be very expensive to produce such profile by standard MC methods (either NB or Bird).

10.2.2 - 2D Flow past an ellipse

In this section we show some computation of two dimensional stationary flow past an ellipse. The system configuration is illustrated in Figure 32. More detail about these computation are shown in [**112**].

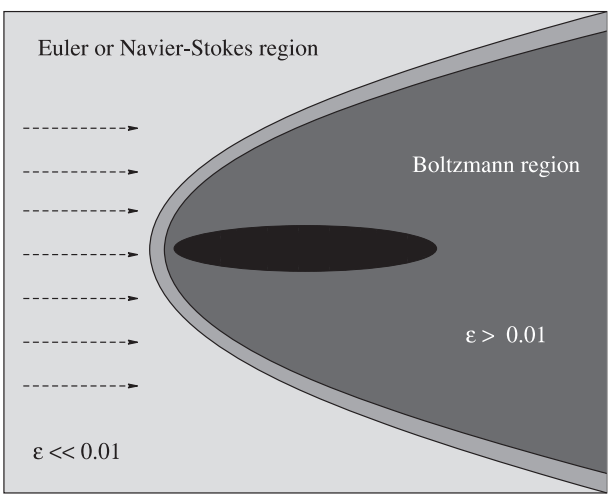

Fig. 32. – 2D flow past an ellipse

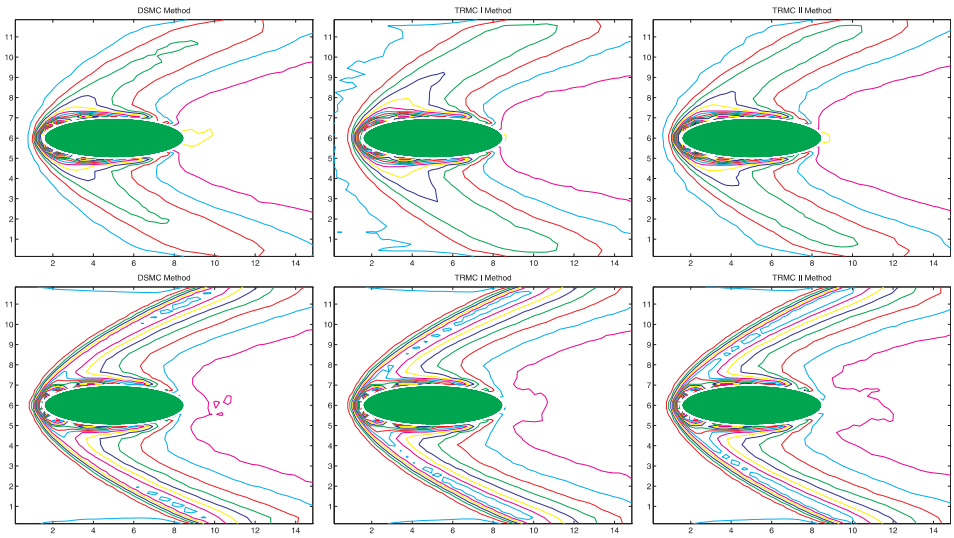

Fig. 33. – 2D flow: $\varepsilon = 0.1$ (top) and $\varepsilon = 0.01$ (bottom). NB (left), TRMC1 (center) and TRMC2 (right), solution for the mass ρ

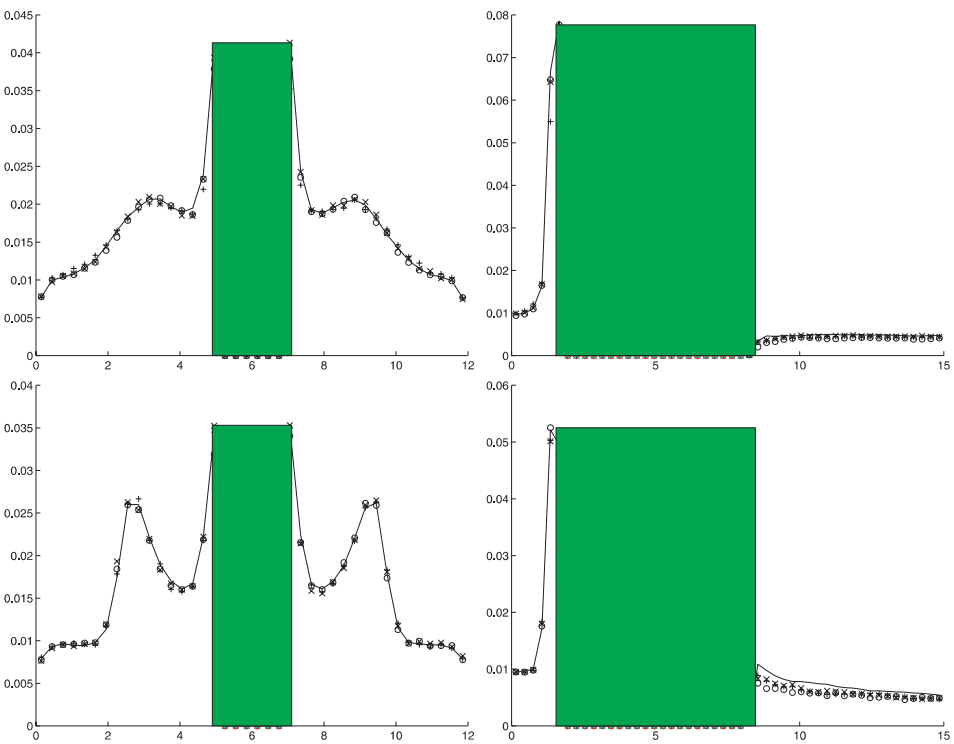

Fig. 34. – 2D flow: $\varepsilon = 0.1$ (top) and $\varepsilon = 0.01$ (bottom). Transversal and longitudinal sections for the mass ρ at $y = 6$ and $x = 5$ respectively; NB ($\circ$), TRMC1 ($+$), TRMC2 ($\times$)

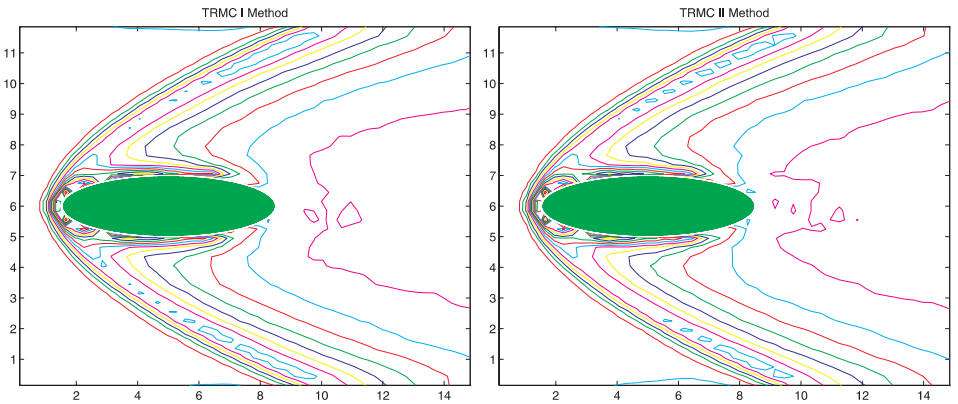

Fig. 35. – 2D flow: $\varepsilon = 0.001$. NB, TRMC1 and TRMC2 solution for the mass ρ

We compare NB, TRMC1 and TRMC2 schemes. The physical parameters of the flow are $\mathcal{M} = 20,\ \rho_{inf} = 0.01,\ T_{inf} = 200,\ T_{obj} = 1000,\ \varepsilon = 0.1, 0.01, 0.001$.

The model is the Boltzmann equation for hard sphere. The computational domain is a rectangle divided into 75×60 space cells and 100 particles in each cell at "infinity", i.e. at a large distance from the obstacle. Full accomodation boundary condition have been used.

As in the case of the stationary 1D shock, since we are computing a stationary shock wave, the accuracy of the methods can be increased by computing averages on the solution for large time.

The numerical results of the computation are illustrated in the figures, which show the 2D density distribution (Figures 33, 35) and the density profiles for longitudinal and transverse cross sections (Figures 34, 36), for various values of the Knudsen number ε (more precisely, for $\varepsilon = 0.1, 0.01, 0.001$).

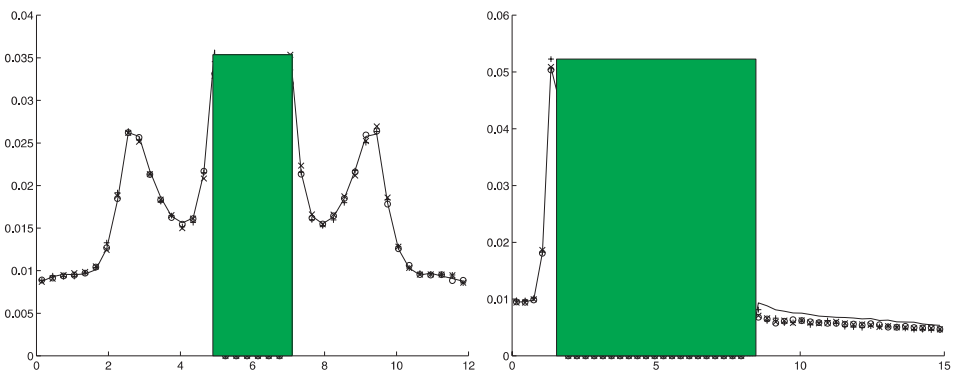

Fig. 36. – 2D flow: $\varepsilon = 0.001$. Transversal and longitudinal sections for the mass ρ at $y = 6$ and $x = 5$ respectively; NB ($\circ$), TRMC1 ($+$), TRMC2 ($\times$).

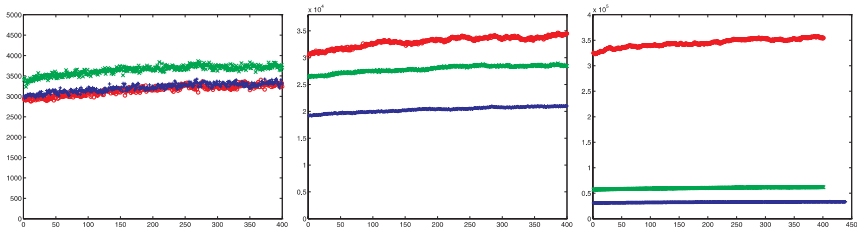

Fig. 37. – 2D flow: Number of "Collisions". Top to bottom and left to right $\varepsilon = 0.1, 0.01, 0.001$; NB ($\circ$), TRMC1 ($+$), TRMC2 (\green$\times$)

It appears that all the schemes provide approximately the same answer, although the computational cost of the various methods can be very different, as is illustrated in figure 37, which shows the total number of collisions per time step, as a function of time. The first picture corresponds to $\varepsilon = 0.1$, the second to $\varepsilon = 0.01$, and the third to $\varepsilon = 0.001$. When the system is far from local equilibrium, very few particles are sampled from the Maxwellian, and TRMC2 is the methods that requires more collisions, while NB and TRMC1 are of comparable cost. For $\varepsilon = 0.01$, the most expensive method is NB, while the least expensive is TRMC1. In the case $\varepsilon = 10^{-3}$, the number of collisions required by NB scheme (higher curve) is roughly one order of magnitude larger that the that required by TRMC (lower curve).

11 - Recursive Monte Carlo methods

11.1 - *Preliminaries*

There are several directions in which the effectiveness of Monte Carlo methods can be increased for the numerical solution of the Boltzmann equation. Here we do not describe extensions of the methods to other physical contexts, such as Boltzmann equation for polyatomic gases, Boltzmann and Boltzman-Vlasov transport equation for charge carriers in semiconductors, and so on, which are very important topics, but which are beyond the scope of the present paper.

Among the most popular strategies that can be used to improve the Monte Carlo methods for the standard Boltzmann equation for monatomic gas, we recall

- High order Monte Carlo in time
- Hybrid Monte Carlo methods
- Weighted particle methods

Here we will describe only some recent developments in the first direction. For the second topic we refer to [**99**], [**100**], [**101**], whereas for the last topic we refer to [**118**], [**119**].

11.2 - *Recursive* TRMC *methods*

The Wild sum expansion of the solution recalled in Section 8.1 has a natural probabilistic interpretation which leads to a recursive implementation of the Monte Carlo method. We rewrite here the starting equation (107)

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon}[P(f,f) - \mu f], \tag{117}$$

with $f(v,0) = f_0(v)$, $\mu > 0$, and $P(f,g)$ a positive bilinear operator.

Let now consider in a time interval $[0, \Delta t]$ the sum at the basis of TR schemes

$$f(v,\Delta t) = (1-\tau)\sum_{k=0}^{\infty} \tau^k f_k(v),$$

where $\tau = 1 - e^{-\mu\Delta t/\varepsilon}$ is the relaxed time. Because P is a positive operator, then all the coefficient f_k are positive functions, and, because of the properties of the Wild sum expansion, their integral is equal to the integral of $f_0(v)$, which we may assume to be 1 without loss of generality.

The sum has a very clear probabilistic interpretation.

- $f(\cdot,\Delta t)$ is the velocity distribution of particles at time Δt. Taking a particle at random from this distribution, it might happen that it has not collided one single time. The probability of this event is $e^{-\mu\Delta t/\varepsilon}$, and the velocity distribution given this is $f_0(v)$.
- The first term in the sum corresponds to those particles that have been involved in exactly one collision. The probability that a particle has such a history is $e^{-\mu\Delta t/\varepsilon}\left(1 - e^{-\mu\Delta t/\varepsilon}\right)$, and the velocity distribution is exactly $f_1(v)$.

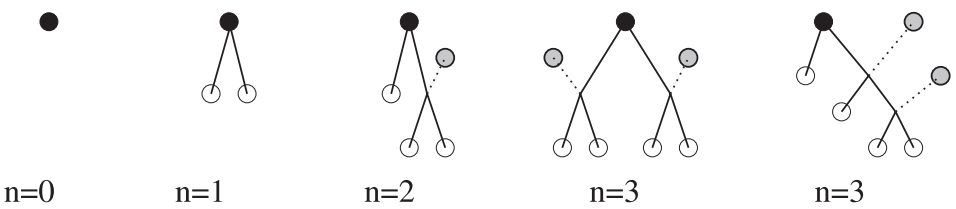

Fig. 38. – McKean graphs

- In the same way, f_n is the conditional velocity distribution given that exactly $n+1$ particles have been involved in its collision history back to the initial time. The coefficient in front of f_n represents the probability of such an event, i.e. that $n+1$ particles have been involved in the collision history. Such number of particles follows a geometric probability distribution.

An obvious algorithm follows from this probabilistic interpretation:

Algorithm 15

- *sample an integer k with geometric distribution $p_k = (1-\tau)\tau^k$ as is described in Example* 4.
- *sample from $f_k(v)$ (as shown below)*

Such an algorithm, however, would become extremely inefficient if $\tau \approx 1$, since the geometric series converges very slowly, and, very likely, one has to sample from f_k, with a large k, which may be very slow, as is evident from the recursive algorithm. Such recursive algorithm would sample from the "exact" solution of the Boltzmann equation. In this respect it is quite similar to the Bird algorithm, which has no time discretization error.

One can, however, make use of the asymptotic properties of the Wild sum coefficients, in order to provide a systematic way of truncating the series. This can be done as follows.

Let us start now from the TR schemes

$$f^{n+1}(v) = (1-\tau)\sum_{k=0}^{m} \tau^k f_k^n(v) + \tau^{m+1} M(v), \tag{118}$$

where

$$f_{k+1}(v) = \frac{1}{k+1}\sum_{h=0}^{k} \frac{1}{\mu} P(f_h, f_{k-h}). \tag{119}$$

Given a set of N particles distributed according to f^n, then, at time $(n+1)\Delta t$ one tries to split the particles into

- a set of $N(1-\tau)$ particles sampled from f_0,
- for each $k = 1, \dots, m$, a set of $N(1-\tau)\tau^k$ particles sampled from f_k, and
- $N\tau^{m+1}$ particles sampled from a Maxwellian.

This can be done in a recursive way making use of equation (118) and of the expression of the Wild sum coefficients (119). A naive implementation is obtained as follows

Algorithm 16

- *sample an integer n from the geometric distribution $p_n = (1-\tau)\tau^n$ (see Example* 4*)*
- *If $n \le m$ sample from f_n otherwise sample from M.*

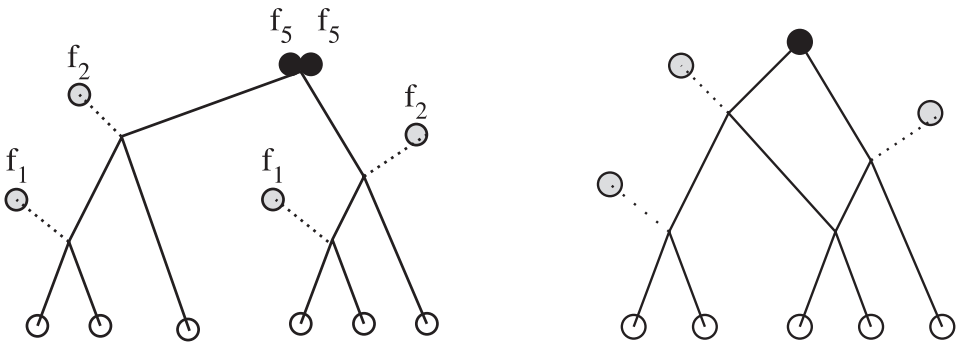

Fig. 39. – Generation of particles in the algorithm (left) and recollisions (right)

The sample from f_n is obtained recursively as

Algorithm 17 ($v =$ **sample from** f_n)

1. if $n = 0$
2. $v = sample\ from\ f_0$
   else
   choose k uniformly in $[0, \ldots, n-1]$
   $v_1 = sample\ from\ f_k$
   $v_2 = sample\ from\ f_{n-k-1}$
   $v = outcome\ of\ collision\ between\ v_1\ and\ v_2$
   end if

The very first step in this calculation is to compute a first pair of particles (velocities) distributed according to f_m. This involves the collisions of $m+1$ particles, and hence this number of particles are drawn at random from the initial distribution. Obviously this sets a limit of the number of terms of the Wild-sum that can be estimated with a finite number of particles at the initial time. All $m+1$ particles must be kept in order that the conserved quantities remain exact. From a practical viewpoint, the number m can be very large. Thus a maximum allowed value m_{max} is fixed (which represent the maximum depth of a possible tree) at the beginning of the computations. Note that, at variance with Bird's method, recollision of particles is not allowed.

The algorithm described above, however, is not conservative. Conservation can be preserved by a suitable use of *pairs of particles* rather than individual particles, in the recursive sampling. This can be done as is described in detail in the following recursive algorithm in the VHS case:

Algorithm 18 (conservative TRMCR for VHS molecules)

1. *For* $n \geq 0$ *compute* $N_n = \mathrm{Iround}\,(N \exp(-\mu\Delta t/\varepsilon)(1-\exp(-\mu\Delta t/\varepsilon))^n)$ *and stop for* $n = m+1$, *where* m *is s.t.* $N_p = \sum_{n=0}^{m} N_n \leq N$ *and* $N_p + N_{m+1} > N$. *Set* $N_{m+1} = N - N_p$.
2. *Set counters* $c_0 = N$ *and* $c_n = 0$ *for* $n = 1, \ldots, m+1$.
3. *For* $n = m, \ldots, 0$ *take* N_n *samples from the distribution with density* f_n. *If* $n = 0$, *take* v *from* f_0 *and decrease the counter* c_0 *by one. Else proceed as follows.*
   (a) *Choose* $k \in \{0, 1, 2, 3, \ldots, n-1\}$ *with equal probability.*
   (b) *Choose* v *from the density* f_k, *and* v_* *from the density* f_{n-k-1}. *This is done as follows.*
      i. *if* $c_k > 0$ *use a previously stored particle with a random choice and decrease the* c_k *counter by one. Otherwise take a sample from* f_k; *this will produce two particles distributed as* f_k, *so one is stored and the* c_k *counter increased by one.*
      ii. *if* $c_{n-k-1} > 0$ *use a previously stored particle with a random choice and decrease the* c_{n-k-1} *counter by one. Otherwise take a sample from* f_{n-k-1}; *this will produce two particles distributed as* f_{n-k-1}, *so one is stored and the* c_{n-k-1} *counter increased by one.*

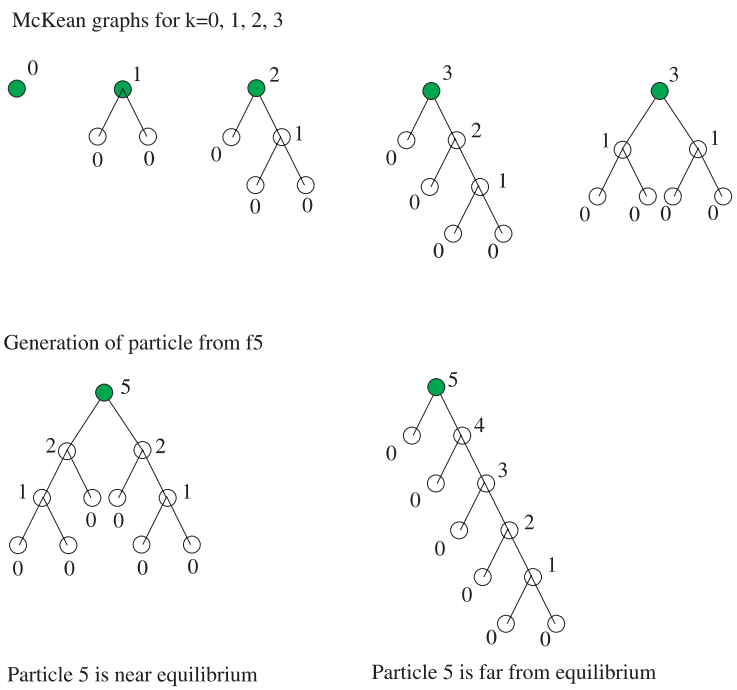

Fig. 40. – Different possible collision trees

(c) Compute a dummy collision between particles v and v_* to get two random variables distributed according to the density f_n.
(d) Sample N_{m+1} particles from the Maxwellian.

The above algorithm deserves some remarks.

Remark 12.

• *If the algorithm is used with no* a priori *bound on m_{max} then basically one is sampling from a representation of the exact solution. Therefore, in the space homogeneous case, there is virtually no error due to time discretization.*

• *If the Knudsen number ε is small, then τ is very large, and the geometric sequence decreases very slowly. Thus the computation becomes expensive, since many collisions are involved. On the other hand for large values of k it would be natural to sample from a Maxwellian, since $f_k \to M$ as $k \to \infty$, therefore one can fix an artificial m_{max}, smaller than the one given by the finite number of particles, and sample from $M(v)$ if $k > m_{max}$. Note however that a large number of collisions is not enough. To be sure that a particle thermalizes one has to check if the McKean graph is balanced (i.e. close to equilibrium) or unbalanced (i.e. not necessarily close to equilibrium). In fact f_k can come from the collision of particle from f_{k-1} with a particle from f_0.*

11.3 - *Numerical results*

In this section we present some numerical tests that show how the recursive TRMC methods can improve the efficiency (by reducing the number of collisions) when compared to traditional DSMC methods.

11.3.1 - Space homogeneous case

Here we consider the following test cases

1. Kac equation (comparison with exact BKW solution);
2. Maxwell molecules (comparison with exact BKW solution);
3. Hard spheres.

The exact solution used in the first two tests is the classical Bobylev-Krook-Wu solution [**10**], [**76**] that we already used to test other methods. The number of particles used in all the calculations is $N = 5 \times 10^4$.

We show the comparison between the recursive implementation of Time Relaxed Monte Carlo methods and Bird's scheme. We perform the computation up to a certain time, using just one single time step $\Delta t = t$. Then we compute the

TABLE 9. – *Kac equation*: *Relative L_2 error norm in time for TRMCR vs BIRD*

m_{max}	1000	100	25		Bird
$t = 0$	0.010390	0.010390	0.010390		0.010390
$t = 1$	0.007169	0.007169	0.007169	$m = 20$	0.007923
$t = 2$	0.005421	0.005421	0.006588	$m = 52$	0.009389
$t = 3$	0.006266	0.007841	0.010110	$m = 118$	0.006489
$t = 5$	0.005971	0.006217	0.021790	$m = 550$	0.004985
$t = 7$	0.008950	0.014841	0.019492	$m > 1000$	0.007836
$t = 10$	0.007481	0.010583	0.010218	$m > 1000$	0.007019
$t = 15$	0.006202	0.006202	0.006202	$m > 1000$	0.006637

relative L^2 norm of the error for the Bird method and the recursive TRMC scheme for several values of the maximum allowed number of terms ($m_{\max}$) in the Wild sum. The result of the comparisons are reported in Tables 9 and 10. For large enough values of $m_{\max}$, recursive TRMC and Bird's scheme give comparable results, while for small values of $m_{\max}$ a systematic difference is observed. The fifth column in the table represents the actual number of terms of the Wild sum required in the computation. Whenever $m < m_{\max}$ no particle is sampled from the Maxwellian.

TABLE 10. – *Maxwellian case: Relative L_2 error norm in time for TRMCR vs BIRD*

m_{max}	1000	100	25		Bird
$t = 0$	0.013312	0.013312	0.013312		0.013312
$t = 1$	0.012055	0.012055	0.012055	$m = 22$	0.011709
$t = 2$	0.012209	0.012209	0.012114	$m = 64$	0.012401
$t = 3$	0.012389	0.013130	0.013446	$m = 156$	0.012213
$t = 5$	0.012737	0.013762	0.016806	$m = 881$	0.012407
$t = 7$	0.011303	0.012881	0.014775	$m > 1000$	0.011802
$t = 10$	0.011896	0.012847	0.012998	$m > 1000$	0.013440
$t = 15$	0.012479	0.012479	0.012479	$m > 1000$	0.012513

The number of collisions required as a function of the final time is represented in Figure 41. Notice the gain in computational efficiency, even for quite large values of $m_{\max}$.

The last test case refers to the hard spheres model. The initial condition here is the sum of two Gaussian. A simple adaptive truncation techniques of the trees based

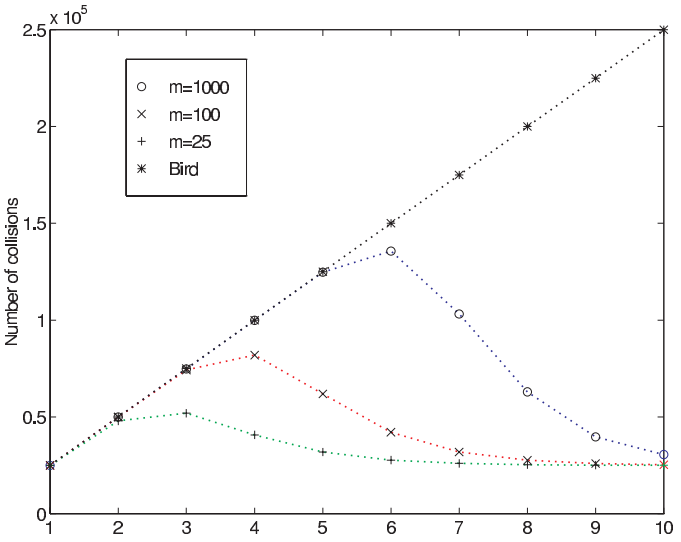

Fig. 41. – Kac equation: Number of collisions in time for TRMCR vs BIRD

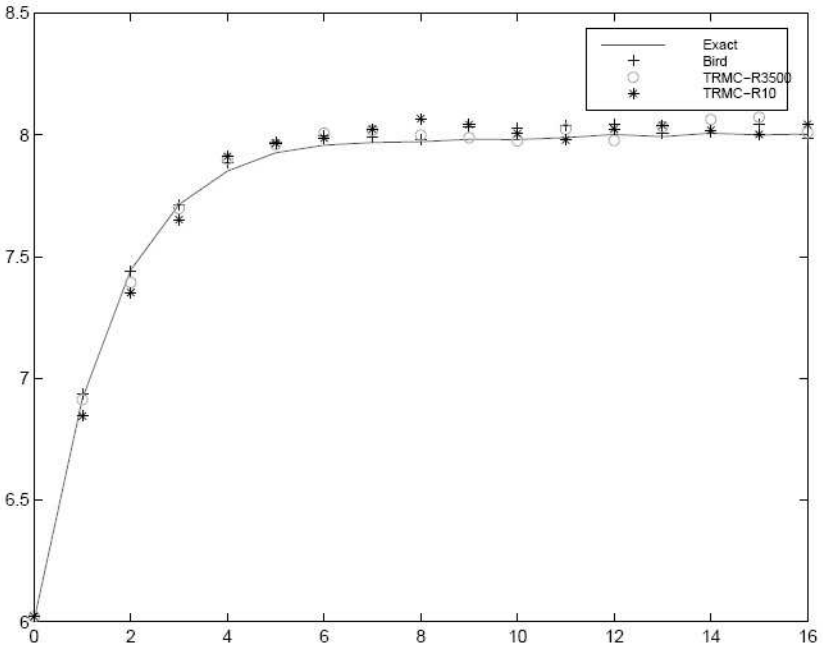

Fig. 42. – Hard spheres: 4th order moment in time. BIRD (+), TRMCR with $k_0 = 10$ (∗) and TRMCR with $k_0 = 3500$ (○)

on the evaluation of the fourth moment of the solution has been implemented. At each time step the method evaluate

$$E_4^{n+1,k} = \frac{|M_4^{n+1,k} - M_4^{n,k}|}{|M_4^{n+1,k}|},$$

where $M_4^{n,k}$ is the result for the fourth moment obtained with TRMCR of length k. If the variation is below a certain tolerance δ_1 then $k = 2k$, if it is above a tolerance δ_2 then $k = k/2$ and the solution is recomputed with the new value of k. Otherwise the result is kept. As a starting value we choose $k_0 = 10$ and $k_0 = 3500$. Note that in this simple technique if $E_4^{n+1,k} > \delta_2$ we discard the whole computation. A better strategy would consist in computing only the fractions of samples from $f_{k+1}, \ldots, f_{2k}$ taking advantage of the collision samples $f_1, \ldots, f_k$ already produced. We refer to [**113**] for further details, and other tests for adaptive strategies.

We compare Bird's method and recursive TRMC with a maximum length of the trees set to $m_{\max} = 3500$, which is approximatively the average collision length of the trees in Birds method in a time step $\Delta t = 1$. The results are reported in Figures 42 and 43.

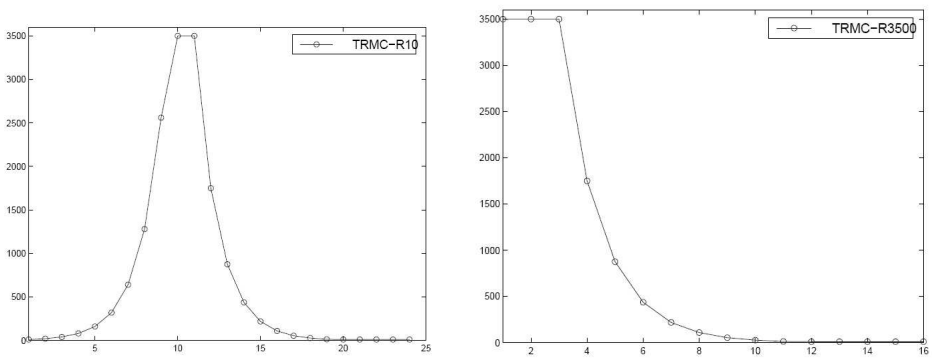

Fig. 43. – Hard spheres: 4th order moment in time. Maximum length of the trees in time for TRMCR with $k_0 = 10$ (left) and TRMCR with $k_0 = 3500$ (right)

11.3.2 - Space non homogeneous case

Next we apply the adaptive method for the homogeneous step to the computation of stationary shock waves. We consider the same initial data as in Section 10.2.1. The Knudsen number here is $\varepsilon = 0.01$ and we take 500 particles in each downstream cell.

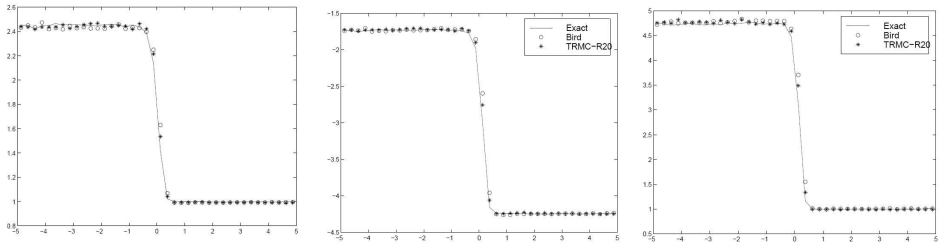

Fig. 44. – Stationary shock: Density (left), mean velocity (center) and temperature (right) for BIRD (∘) and TRMCR with $k_0 = 20$ (∗)

The reference solution has been computed with Bird's method and 1000 particles per cell. The results are reported in Figures 44 and 45. The gain of computational efficiency of TRMCR methods is evident.

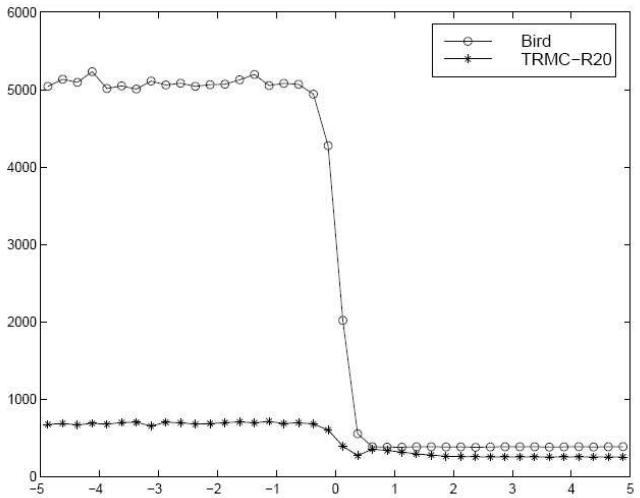

Fig. 45. – Stationary shock: number of collisions for BIRD (∘) and TRMCR with $k_0 = 20$ (∗)

References

[1] V. V. ARISTOV and F.G. CEREMISIN, *The conservative splitting method for solving Boltzmann's equation*, USSR Comput. Maths. Math. Phys. **20** (1980), 208-225.

[2] D. ARMBRUSTER, P. DEGOND and C. RINGHOFER, *Kinetic and fluid models for supply chains supporting policy attibutes*, submitted, Transp. Theory and Stat. Phys. (2004).

[3] A. ARNOLD and C. RINGHOFER, *An operator splitting method for the Wigner-Poisson problem*, SIAM J. Numer. Anal. **33** (1996), 1622-1643.

[4] H. BABOVSKY, *On a simulation scheme for the Boltzmann equation*, Math. Methos Appl. Sci. 8 (1986), 223-233.

[5] H. BABOVSKY and R. ILLNER, *A convergence proof for Nambu's simulation method for the full Boltzmann equation*, SIAM J. Numer. Anal. **26** (1989), 45-65.

[6] P. I. BHATNAGAR, E.P. GROSS and M. KROOK, *A model for collision processes in gases I. Small amplitute processes in charged and neutral one component systems*, Phys. Rev. **94** (1954), 511.

[7] A. BALDASSARRI, U. MARINI BETTOLO MARCONI and A. PUGLISI, *Kinetic models of inelastic gases*, Math. Models Methods Appl. Sci. **12** (2002), 965-983.

[8] N. BELLOMO and A. BELLOUQUID, *From a class of kinetic models to the macroscopic equations for multicellular systems in biology*, Mathematical models in cancer (Nashville, TN, 2002). Discrete Contin. Dyn. Syst. Ser. B **4** (2004), 59-80.

[9] G. A. BIRD, *Molecular gas dynamics and direct simulation of gas flows*, Clarendon Press, Oxford 1994.

[10] A. V. BOBYLEV, *Exact solutions of the Boltzmann equation*, Dokl. Akad. Nauk. S.S.S.R. **225** (1975), 1296-1299 (Russian).

[11] A. V. BOBYLEV, *The theory of the nonlinear spatially uniform Boltzmann equation for Maxwell molecules*, Math. Phys. Reviews **7** (1988), 111-233.

[12] A. V. BOBYLEV, J.A. CARRILLO and I.M. GAMBA, *On some properties of kinetic and hydrodynamics equations for inelastic interactions*, J. Statist. Phys. **98** (2000), 743-773.

[13] A. V. BOBYLEV, A. PALCZEWSKI and J. SCHNEIDER, *On approximation of the Boltzmann equation by discrete velocity models*, C. R. Acad. Sci. Paris Sér. I. Math. **320** (1995), 639-644.

[14] A. V. BOBYLEV and S. RJASANOW, *Difference scheme for the Boltzmann equation based on the fast Fourier transform*, Eur. J. Mech. B Fluids **16** (1997), 293-306.

[15] A. V. BOBYLEV and S. RJASANOW, *Fast deterministic method of solving the Boltzmann equation for hard spheres*, Eur. J. Mech. B Fluids **18** (1999), 869-887.

[16] F. BOUCHUT and B. PERTHAME, *A BGK model for small Prandtl numbers in the Navier-Stokes approximation*, J. Statist. Phys. **71** (1993), 191.

[17] J. F. BOURGAT, *Notice d'utilisation du logiciel BOL2D pour la simulation bidimensionnelle de l'equation de Boltzmann*, Rapport technique de l'INRIA-Rocquencourt, RT-0142 (1992).

[18] J. F. BOURGAT, P. LETALLEC, B. PERTHAME and Y. QIU, *Coupling Boltzmann and Euler equations without overlapping*, in Domain Decomposition Methods in Science and Engineering, Contemp. Math. **157**, AMS, Providence, RI 1994, 377-398.

[19] J. F. BOURGAT, P. LETALLEC and M.D. TIDRIRI, *Coupling Boltzmann and Navier-Stokes equations by friction*, J. Comput. Phys. **127** (1996), pag. 227-245.

[20] J. E. BROADWELL, *Shock structure in a simple discrete velocity gas*, Phys. Fluids **7** (1964), 1243-1247.

[21] C. BUET, *A discrete velocity scheme for the Boltzmann operator of rarefied gas dynamics*, Transport Theory Statist. Phys. **25** (1996), 33-60.

[22] C. BUET, S. CORDIER, P. DEGOND and M. LEMOU, *Fast algorithms for numerical, conservative, and entropy approximations of the Fokker-Planck equation*, J. Comput. Phys. **133** (1997), 310-322.

[23] R. E. CAFLISCH, *The fluid dynamical limit of the nonlinear Boltzmann equation*, Comm. Pure Appl. Math. **33** (1980), 651-666.

[24] R. E. CAFLISCH, *Monte Carlo and quasi-Monte Carlo methods*, Acta Numerica (1998), 1-49.

[25] R. E. CAFLISCH, S. JIN and G. RUSSO, *Uniformly accurate schemes for hyperbolic systems with relaxation*, SIAM J. Numer. Anal. **34** (1997), 246-281.

[26] C. CANUTO, M.Y. HUSSAINI, A. QUARTERONI and T.A. ZANG, *Spectral methods in fluid dynamics*, Springer Series in Computational Physics, Springer-Verlag, New York 1988.

[27] T. CARLEMAN, *Sur la théorie de l'équation intégrodifférentielle de Boltzmann*, Acta Math. **60** (1932).

[28] E. A. CARLEN, M.C. CARVALHO and E. GABETTA, *Central limit theorem for Maxwellian molecules and truncation of the Wild expansion*, Comm. Pure Appl. Math. **53** (2000), 370-397.

[29] E. A. CARLEN and F. SALVARANI, *On the optimal choice of coefficients in a truncated Wild sum and approximate solutions for the Kac equation*, J. Statist. Phys. **109** (2002), 261-277.

[30] C. CERCIGNANI, *The Boltzmann equation and its applications*, Springer Verlag, New York 1988.

[31] C. CERCIGNANI, R. ILLNER and M. PULVIRENTI, *The mathematical theory of dilute gases*, Springer-Verlag, New York 1994.

[32] C. CERCIGNANI and M. LAMPIS, *Kinetic models for gas-surface interactions*, Transport Theory Statist. Phys. **1** (1971), 101-114.

[33] F. G. CEREMISIN, *Fast solution of the Boltzmann Equation*, Proceedings of the 17th International Symposium on Rarefied Gas Dynamics, edited by A. Beylich, Weinheim (1991), 273-284.

[34] F. CHALUB, P.A. MARKOWICH, B. PERTHAME and C. SCHMEISER, *Kinetic models for chemotaxis and their drift-diffusion limits*, Monatsh. Math. **142** (2004), 123-141.

[35] A. J. CHORIN, *Numerical solution of Boltzmann's equation*, Comm. Pure Appl. Math. (1972), 171-186.

[36] S. CORDIER, L. PARESCHI and G. TOSCANI, *On a kinetic model for a simple market economy*, J. Stat. Phys. **120** (2005), 253-277.

[37] F. CORON and B. PERTHAME, *Numerical passage from kinetic to fluid equations*, SIAM J. Numer. Anal. **28** (1991), 26-42.

[38] P. DEGOND and B. LUCQUIN-DESREUX, *An entropy scheme for the Fokker-Planck collision operator of plasma kinetic theory*, Numer. Math. **68** (1994), 239-262.

[39] P. DEGOND, L. PARESCHI and G. RUSSO, *Modeling and Computational Methods for Kinetic Equations*, Series: Modeling and Simulation in Science, Engineering and Technology, Birkhauser, Boston 2004.

[40] S. M. DESHPANDE, *A Second Order Accurate, Kinetic-Theory Based, Method for Inviscid Compressible Flows*, NASA Langley Tech. Paper NO. 2613 (1986).

[41] L. DESVILLETTES, R. FERRIERE and C. PREVOST, *Infinite Dimensional Reaction-diffusion for Population Dynamics*, preprint (2004).

[42] L. DESVILLETTES and S. MISCHLER, *About the splitting algorithm for Boltzmann and BGK equations*, Math. Models Methods Appl. Sci. **6**, (1996), 1079-1101.

[43] L. DESVILLETTES and R.E. PERALTA HERRERA, *A vectorizable simulation method for the Boltzmann equation*, RAIRO Modél. Math. Anal. Numér. **28** (1994), 745-760.

[44] M. ESCOBEDO, PH. LAURENOT, S. MISCHLER and B. PERTHAME, *Gelation and mass conservation in coagulation-fragmentation models*, J. Differential Equations **195** (2003), 143-174.

[45] F. FILBET, C. MOUHOT and L. PARESCHI, *Solving the Boltzmann equation in* $O(N \log N)$, preprint (2005).

[46] F. FILBET and L. PARESCHI, *A numerical method for the accurate solution of the Fokker-Planck-Landau equation in the non homogeneous case*, J. Comput. Phys. **179** (2002), 1-26.

[47] F. FILBET, L. PARESCHI and G. TOSCANI, *Accurate numerical methods for the collisional motion of (heated) granular flows*, J. Comput. Phys. **202** (2005), 216-235.

[48] F. FILBET and G. RUSSO, *High order numerical methods for the space non-homogeneous Boltzmann equation*, J. Comput. Phys. **186** (2003), 457-480.

[49] F. FILBET, E. SONNENDRÜCKER and P. BERTRAND, *Conservative Numerical schemes for the Vlasov equation*, J. Comput. Phys. **172** (2001), 166-187.

[50] E. GABETTA and L. PARESCHI, *The Maxwell gas and its Fourier transform towards a numerical approximation*, Series on Advances in Math. for Appl. Sci. **23** (1994), 197-201.

[51] E. GABETTA, L. PARESCHI and G. TOSCANI, *Relaxation schemes for nonlinear kinetic equations*, SIAM J. Numer. Anal. **34** (1997), 2168-2194.

[52] I. M. GAMBA, S. RJASANOW and R. WAGNER, *Direct simulation to the uniformly heated granular Boltzmann equation*, Math. Comput. Model, to appear.

[53] A. L. GARCIA, J.B. BELL, W.Y. CRUTCHFIELD and B.J. ALDER, *Adaptive mesh and algorithm refinement using Direct Simulation Monte Carlo*, Journal of Computational Physics **154** (1999), 134-155.

[54] R. GATIGNOL, *Théorie cinétique d'un gas à répartition discrète de vitesses*, Lecture Notes in Phys. **36**, Springer Verlag, Heidelberg 1975.

[55] D. GOLDSTEIN, B. STURTEVANT and J.E. BROADWELL, *Investigation of the motion of discrete velocity gases*, Rar. Gas. Dynam., Progress in Astronautics e Aeronautics, 118, AIAA, Washington 1989.

[56] F. GOLSE and L. SAINT-RAYMOND, *The Navier-Stokes limit of the Boltzmann equation for bounded collision kernels*, Invent. Math. **155** (2004), 81-161.

[57] D. GOTTLIEB and S.A. ORSZAG, *Numerical analysis of spectral methods: theory and applications*, SIAM CBMS-NSF Series, Soc. Indust. Appl. Math., Philadelphia, Pa. 1977.

[58] Y. N. GRIGORIEV and A.N. MIKHALITSYN, *A spectral method of solving Boltzmann's kinetic equation numerically*, U.S.S.R. Comput. Maths. Math. Phys. **23** (1983), 105-111.

[59] N. G. HADJICOSTANTINOU, *Analysis of discretization in the direct simulation Monte Carlo*, Phys. Fluids **12** (2000), 2634.

[60] M. HERTY, A. KLAR and L. PARESCHI, *General kinetic models for vehicular traffic flows and Monte Carlo methods*, Comp. Meth. App. Math. **5** (2005), 154-169.

[61] R. W. HOCKNEY and J.W. EASTWOOD, *Computer Simulation Using Particles*, McGraw-Hill, New York 1981.

[62] L. H. HOLWAY, *New statistical models for kinetic theory: methods of construction*, Phys. Fluids **9** (1966), 1658-1673.

[63] I. IBRAGIMOV and S. RJASANOW, *Numerical solution of the Boltzmann equation on the uniform grid*, Computing **69** (2002), 163-186.

[64] R. ILLNER and H. NEUNZERT, *On simulation methods for the Boltzmann equation*, Transport Theory Statist. Phys. **16** (1987), 141-154.

[65] T. INAMURO and B. STURTEVANT, *Numerical study of discrete velocity gases*, Phys. Fluids A **2**, (1990), 2196-2203.

[66] V. V. IVANOV, VAL. V. IVANOV, YU. L. KALINOVSKY and P.V. ZRELOV, *Statistical and Kinetic Models of Internet Traffic Flows*, preprint (2004).

[67] M. S. IVANOV and S.V. ROGASINSKY, *Analysis of the numerical techniques of the direct simulation Monte Carlo method in the rarefied gas dynamics*, Soviet J. Numer. Anal. Math. Modelling **3** (1988), 453-465.

[68] S. JIN, *Runge-Kutta methods for hyperbolic conservation laws with stiff relaxation terms*, J. Comput. Phys. **122** (1995), 51-67.

[69] S. JIN and L. PARESCHI, *Discretization of the multiscale semiconductor Boltzmann equation by diffusive relaxation schemes*, J. Comput. Phys. **161** (2000), 312-330.

[70] S. JIN, L. PARESCHI and G. TOSCANI, *Diffusive relaxation schemes for multiscale discrete-velocity kinetic equations*, SIAM J. Numer. Anal. **35** (1998), 2405-2439.

[71] S. JIN, L. PARESCHI and G. TOSCANI, *Uniformly accurate diffusive relaxation schemes for multiscale transport equations*, SIAM J. Numer. Anal. **38** (2000), 913-936.

[72] M. KAC, *Probability and related topics in the physical sciences*, Interscience Publ., New York 1959.

[73] M. H. KALOS and P.A. WHITLOCK, *Monte Carlo methods*, I: Basics. John Wiley & Sons, New York 1986.

[74] A. KLAR and R. WEGENER, *Enskog-like kinetic models for vehicular traffic*, J. Statist. Phys. **87** (1997), 91-114.

[75] D. E. KNUTH, *The art of computer programming*, Addison-Wesley Publ. Company Reading, Mass. 1973-1981.

[76] M. KROOK and T.T. WU, *Formation of maxwellian tails*, Phys. Rev. Lett. **36** (1976), 1107-1109.

[77] L. D. LANDAU, *Die kinetische gleichung für den fall Coulombscher vechselwirkung*, Phys. Z. Sowjet. **154** (1936). Trad. *The transport equation in the case of the Coulomb interaction*, in D.ter Haar Ed. Collected papers of L.D. Landu, 163-170. Pergamon press, Oxford 1981.

[78] M. LEMOU, *Multipole expansions for the Fokker-Planck-Landau operator*, Numer. Math. **78** (1998), 597-618.

[79] R. J. LeVeque, *Numerical methods for conservation laws*, Birkhauser Verlag, Basel 1992.

[80] C. D. Levermore, *Moment closure hierarchies for kinetic theories*, J. Statist. Phys. **83** (1996), 1021-1065.

[81] Z. -H. Li and H.-X. Zhang, *Numerical investigation from rarefied flow to continuum by solving the Boltzmann model equation*, International Journal for Numerical Methods in Fluids Volume **42** (2003), 361-382.

[82] E. Longo and L. Preziosi, *On a conservative polar discretization of the Boltzmann equation*, Japan J. Indust. App. Math. **14** (1997), 399-435.

[83] P. A. Markowich and L. Pareschi, *Fast conservative and entropic numerical methods for the boson Boltzmann equation*, Numer. Math. **99** (2005), 509-532.

[84] P. A. Markowich, C. Ringhofer and C. Schmeiser, *Semiconductor equations*, Springer-Verlag, Wien-New York 1989.

[85] Y. -L. Martin, F. Rogier and J. Schneider, *Une méthode déterministe pour la résolution de l'équation de Boltzmann inhomogène*, C. R. Acad. Sci. Paris Sér. I Math. **314** (1992), 483-487.

[86] L. Mieussens, *Discrete velocity model and implicit scheme for the BGK equation of rarefied gas dynamics*, Math. Models Methods Appl. Sci. **10** (2000), 1121-1149.

[87] Y. Morchoisne, *Une méthode de différences finies pour la résolution de l'equation de Boltzmann: Traitement du terme de collision*, C. R. Acad. Sci. Paris Sér. II, **313** (1991), 1513-1518.

[88] C. Mouhot and L. Pareschi, *Fast methods for the Boltzmann collision integral*, C. R. Acad. Sci. Paris Sér I Math. **339** (2004), 71-76.

[89] C. Mouhot and L. Pareschi, *Fast algorithms for computing the Boltzmann collision operator*, Math. Comp. (2004), to appear.

[90] C. Mouhot and L. Pareschi, *An $O(N \log N)$ algorithm for computing discrete velocity models*, preprint (2005).

[91] G. Naldi, L. Pareschi and G. Toscani, *Spectral methods for one-dimensional kinetic models of granular flows and numerical quasi elastic limit*, ESAIM RAIRO Math. Model. Numer. Anal. **37** (2003), 73-90.

[92] K. Nanbu, *Direct simulation scheme derived from the Boltzmann equation I. Monocomponent Gases*, J. Phys. Soc. Japan **49** (1980), 2042-2049.

[93] T. Nishida, *Fluid dynamical limit of the nonlinear Boltzmann equation at the level of the compressible Euler equations*, Comm. Math. Phys. **61** (1978), 119-148.

[94] T. Ohwada, *Structure of normal shock waves: Direct numerical analysis of the Boltzmann equation for hard sphere molecules*, Phys. Fluids A, **5** (1993), 217-234.

[95] A. Palczewski, J. Schneider and A.V. Bobylev, *A consistency result for a discrete-velocity model of the Boltzmann equation*, SIAM J. Numer. Anal. **34** (1997), 1865-1883.

[96] A. Palczewski and J. Schneider, *Existence, stability, and convergence of solutions of discrete velocity models to the Boltzmann equation*, J. Statist. Phys. **91** (1998), 307-326.

[**97**] V. A. PANFEROV and A.G. HEINTZ, *A new consistent discrete-velocity model for the Boltzmann equation*, Math. Methods Appl. Sci. **25** (2002), 571-593.

[**98**] L. PARESCHI, *Computational methods and fast algorithms for Boltzmann equations*, Lecture Notes on the discretization of the Boltzmann equation, Chapter 7, Series on Advances in Mathematics for Applied Sciences **63**, World Scientific 2003.

[**99**] L. PARESCHI, *Hybrid multiscale methods for hyperbolic and kinetic problems*, ESAIM PROCEEDINGS, Vol. **15**, T. Goudon, E. Sonnendrucker and D. Talay Editors (2005), 87-120.

[**100**] L. PARESCHI and R. E. CAFLISCH, *An implicit Monte Carlo methods for rarefied gas dynamics I: The space homogeneous case*, J. Comput. Phys. **154** (1999), 90-116.

[**101**] L. PARESCHI and R. E. CAFLISCH, *Towards an hybrid Monte Carlo method for rarefied gas dynamics*, IMA Vol. App. Math. **135** (2004).

[**102**] L. PARESCHI and B. PERTHAME, *A Fourier spectral method for homogeneous Boltzmann equations*, Transport Theory Statist. Phys. **25** (1996), 369-382.

[**103**] L. PARESCHI and G. RUSSO, *An Introduction to Monte Carlo Methods for the Boltzmann equation*, Esaim Proceedings **10** EDP Sciences, SMAI (1999), 1-38.

[**104**] L. PARESCHI and G. RUSSO, *Asymptotic preserving Monte Carlo methods for the Boltzmann equation*, Transport Theory Statist. Phys. **29** (2000), 415-430.

[**105**] L. PARESCHI and G. RUSSO, *Time Relaxed Monte Carlo methods for the Boltzmann equation*, SIAM J. Sci. Comput. **23** (2001), 1253-1273.

[**106**] L. PARESCHI and G. RUSSO, *Numerical solution of the Boltzmann equation I. Spectrally accurate approximation of the collision operator*, SIAM J. Numer. Anal. **37** (2000), 1217-1245.

[**107**] L. PARESCHI and G. RUSSO, *On the stability of spectral methods for the homogeneous Boltzmann equation*, Transport Theory Statist. Phys. **29** (2000), 431-447.

[**108**] L. PARESCHI and G. RUSSO, *Implicit-Explicit Runge-Kutta methods and applications to hyperbolic systems with relaxation*, J. Sci. Comp., to appear.

[**109**] L. PARESCHI, G. RUSSO and G. TOSCANI, *Fast spectral methods for the Fokker-Planck-Landau collision operator*, J. Comput. Phys. **165** (2000), 216-236.

[**110**] L. PARESCHI, G. RUSSO and G. TOSCANI, *Modelling and numerical methods for dissipative kinetic equations*, Lipari Conference Proceedings (2004), Nova Science, to appear.

[**111**] L. PARESCHI, G. TOSCANI and C. VILLANI, *Spectral methods for the non cut-off Boltzmann equation and numerical grazing collision limit*, Numer. Math. **93** (2003), 527-548.

[**112**] L. PARESCHI and S. TRAZZI, *Numerical solution of the Boltzmann equation by Time Relaxed Monte Carlo (TRMC) methods*, Internat. J. Numer. Methods Fluids **48** (2005), 947-983.

[**113**] L. PARESCHI, S. TRAZZI and B. WENNBERG, *Recursive Monte Carlo methods for the Boltzmann equation*, preprint (2005).

[**114**] L. PARESCHI and B. WENNBERG, *A recursive Monte Carlo method for the Boltzmann equation in the Maxwellian case*, Monte Carlo Methods Appl. **7** (2001), 349-358.

[**115**] B. PERTHAME, *Introduction to the theory of random particle methods for Boltzmann equation*, in Advances in Kinetic Theory and Computing, B. Perthame Editor, World Scientific 1994.

[**116**] D. I. PULLIN, *Direct simulation methods for compressible inviscid ideal gas flow*, J. Comput. Phys. **34** (1980), 231-244.

[**117**] D. I. PULLIN, *Generation of normal variates with given sample*, J. Statist. Comput. Simulation **9** (1979), 303-309.

[**118**] S. RJASANOW and W. WAGNER, *A stochastic weighted particle method for the Boltzmann equation*, J. Comput. Phys. **124** (1996), 243-253.

[**119**] S. RJASANOW and W. WAGNER, *Simulation of rare events by the stochastic weighted particle method for the Boltzmann equation*, Math. Comput. Modelling **33** (2001), 907-926.

[**120**] P. L. ROE and D. SIDILKOVER, *Optimum positive linear schemes for advection in two and three dimensions*, SIAM J. Numer. Anal. **29** (1992), 1542-1568.

[**121**] F. ROGIER and J. SCHNEIDER, *A direct method for solving the Boltzmann equation*, Transport Theory Statist. Phys. **23** (1994), 313-338.

[**122**] G. A. SOD, *A numerical solution of Boltzmann's equation*, Comm. Pure Appl. Math. **30** (1977), 391-419.

[**123**] Y. SONE, K. AOKI, S. TAKATA, H. SUGIMOTO and A.V. BOBYLEV, *Inappropriateness of the heat-conduction equation for description of a temperature field of a stationary gas in the continuum limit: examination by asymptotic analysis and numerical computation of the Boltzmann equation*, Phys. Fluids 8 (1996), 628-638.

[**124**] Y. SONE, T. OHWADA and K. AOKI, *Temperature jump and Knudsen layer in a rarefied gas over a plane wall: Numerical analysis of the linearized Boltzmann equation for hard-sphere molecules*, Phys. Fluids **1** (1989), 363-370.

[**125**] G. STRANG, *On the construction and the comparison of difference schemes*, SIAM J. Numer. Anal. **5** (1968), 506-517.

[**126**] Z. TAN and P. VARGHESE, *The Δ-ε method for the Boltzmann equation*, J. Comput. Phys. **110** (1994), 327-340.

[**127**] S. TIWARI and A. KLAR, *An adaptive domain decomposition procedure for Boltzmann and Euler equations*, J. Comput. Appl. Math. **90** (1998), 223-237.

[**128**] C. TRUESDELL and R.G. MUNCASTER, *Fundamentals of Maxwell's kinetic theory of a simple monatomic gas*, Academic Press, New York 1980.

[**129**] C. VILLANI, *A review of mathematical topics in kinetic theory*, Handbook of fluid mechanics, S. Friedlander and D. Serre, Eds. Elsevier Publ. Amsterdam 2002.

[**130**] W. WAGNER, *A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation*, J. Statist. Phys. **66** (1992), 1011-1044.

[**131**] G. B. WHITHAM, *Linear and nonlinear waves*, J. Wiley, New York 1974.

[**132**] E. WILD, *On Boltzmann's equation in the kinetic theory of gases*, Proc. Cambridge Philos. Soc. **47** (1951), 602-609.

[**133**] J. -S. WU, K.-C. TSENG and C.-H. KUO, *The direct simulation Monte Carlo method using unstructured adaptive mesh and its application*, International Journal for Numerical Methods in Fluids **38** (2002), 351-375.

Abstract

The different aspects of the numerical solution of the Boltzmann equation and related kinetic models are analyzed and discussed from a deterministic as well as from a probabilistic viewpoint. In both settings after a review of classical methods like discrete velocity modelling and direct simulation Monte Carlo, the emphasis is addressed to the most recent developments in the field, such as (fast) spectral methods and time relaxed Monte Carlo methods. Besides the algorithmic aspects and the efficiency of the methods, considerations on stability, accuracy and consistency of the various schemes are reported.

* * *