An Introduction to Monte Carlo methods for the Boltzmann equation*

Lorenzo Pareschi† and Giovanni Russo‡

Abstract

The purpose of this note is to give an introductory overview of the different Monte Carlo methods for the numerical solution of the Boltzmann equation. Particular emphasis is given to some recently developed Monte Carlo schemes. Some numerical examples are also included. The note is far from being complete and only some aspects are treated of such a vast theme.

Contents

1 Introduction **2**

2 The Boltzmann equation **2**
2.1 The physical model . . . 2
2.2 Boundary conditions . . . 3
2.3 Fluid dynamic limit . . . 4
2.4 Splitting approach . . . 4

3 Direct Simulation Monte Carlo (DSMC) Methods **5**
3.1 Random sampling . . . 5
3.1.1 Monovariate distribution . . . 5
3.1.2 Discrete sampling . . . 6
3.1.3 Multivariate distributions . . . 7
3.2 Direct simulation schemes . . . 9
3.2.1 Nanbu-Babovsky scheme . . . 10
3.2.2 Bird's scheme . . . 12

4 Time Relaxed Monte Carlo (TRMC) methods **12**
4.1 Motivations . . . 12
4.2 Asymptotic preserving schemes . . . 13
4.3 Analysis of the TR schemes . . . 15
4.4 Time Relaxed Monte Carlo (TRMC) methods . . . 18
4.4.1 TRMC methods . . . 18
4.4.2 Hybrid TRMC methods . . . 20
4.4.3 Recursive TRMC methods . . . 21
4.5 Numerical Examples . . . 21

5 Monte Carlo Methods: space non homogeneous case **22**
5.1 Transport of particles . . . 24
5.2 Implementation of boundary conditions . . . 24
5.3 Shock wave profiles . . . 25

*This work was partially supported by TMR project "Asymptotic Methods in Kinetic Theory", Contract Number ERB FMRX CT97 0157, and partially by MURST project "Numerical methods for evolution problems".

†Universitá di Ferrara, Dipartimento di Matematica, Via Machiavelli 35, 44100 Ferrara, Italy (pareschi@dm.unife.it). Home page: http://www.unife.it/∼prl.

‡Universitá dell'Aquila, Dipartimento di Matematica, Via Vetoio, Loc. Coppito, 67010 L'Aquila, Italy (russo@univaq.it). Home page: http://univaq.it/∼russo.

1 Introduction

When a gas is near thermodynamical equilibrium, a detailed kinetic description is unnecessary, and the evolution of the moments can be described by Euler or Navier-Stokes equations. From a computational point of view, these equations are much less expensive than the full Boltzmann equation.

Far from thermodynamic equilibrium, the equation of gas dynamics do not give a satisfactory description of the physical system, and the full kinetic description is necessary.

The departure from local thermodynamic equilibrium can be measured by the so called Knudsen number, which is the ratio between the collisional mean free path of the molecules and the characteristic length of variation of macroscopic variables. When the Knudsen number is small, then the distribution function is close to the local Maxwellian. Collisions occur at a fast rate, and a kinetic treatment of the system is very expensive, because of the large ratio of time scales between macroscopic and microscopic effects. One may say that the system is numerically *stiff* [7]. Because of the stiffness, standard explicit schemes for the numerical solution of the Boltzmann equation become extremely expensive in this case.

If the Knudsen number is uniformly small in the region one is interested in, then one should use a gas dynamic description of the system. There are several circumstances, however, where the Knudsen number varies over several orders of magnitude. In such cases a standard kinetic treatment would be too expensive, and a pure gas dynamic approach inappropriate. Domain decomposition methods have been proposed for this problem, in which the computational domain is divided into an *aerodynamic region* in which the system is treated by gas dynamic equations (Euler or Navier-Stokes), and a *Boltzmann region*, where a kinetic treatment is used. Suitable boundary conditions make the two regions coupled together [3]. Such schemes can be very effective in treating large scale problems, but they are quite complicated to implement, and they require a great care by the user.

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 [15, 14].

These algorithms are based on a suitable time discretization of the Boltzmann equation, first introduced in [6], which has the following properties: for large Knudsen number it is a time discretization of the Boltzmann equation of a given order, and when the Knudsen number vanishes, the numerical solution tends to the local Maxwellian. We will call this last property *asymptotic preservation*. In addition, the time discretization preserves mass, momentum and energy.

Although the paper deals mainly with the space homogeneous case, the real motivation for the development of these Monte Carlo methods is their use for spatially dependent problems. In that case, standard splitting can be used, and the evolution of the system is obtained by alternating a collision step, which is space homogeneous in each space cell, and a convection step, which describes the free flow of the molecules. In the limit of very small Knudsen number, the collision step replaces the distribution function by a local Maxwellian with the same moments. A Monte Carlo method based on such time discretization will behave as a sort of stochastic kinetic scheme for the underlying Euler equations of gas dynamics. Such limit scheme was introduced by Pullin [17].

The plan of the paper is the following. In the next section we give a short introduction on the Boltzmann equation and some of its relevant properties.

2 The Boltzmann equation

2.1 The physical model

We consider the Boltzmann equation [4]

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

with the initial condition

$$f(x, v, t = 0) = f_0(x, v), \tag{2}$$

where $f = f(x, v, t)$ is a non negative function describing the time evolution of the distribution of particles which move with velocity v in the position x at time $t > 0$. The parameter $\varepsilon > 0$ is the Knudsen number and is proportional to the mean free path between collisions. The bilinear collision operator $Q(f, f)$ describes the binary collisions of the particles and is given by

$$Q(f,f)(v) = \int_{I\!R^3} \int_{S^2} \sigma(|v - v_1|, \omega)[f(v')f(v'_*) - f(v)f(v_*)]\, d\omega\, dv_*. \tag{3}$$

In the above expression, ω is a unit vector of the sphere S^2, so that $d\omega$ is an element of area of the surface of the unit sphere S^2 in $I\!R^3$. Moreover (v', v'_*) represent the post-collisional velocities associated with the pre-collisional velocities (v, v_*) and the collision parameter ω; i.e.,

$$v' = \frac{1}{2}(v + v_* + |v - v_*|\omega), \quad v'_* = \frac{1}{2}(v + v_* - |v - v_*|\omega). \tag{4}$$

The kernel σ is a nonnegative function which characterizes the details of the binary interactions. In the case of inverse k-th power forces between particles the kernel has the form

$$\sigma(|v - v_*|, \theta) = b_\alpha(\theta)|v - v_*|^\alpha, \tag{5}$$

where $\alpha = (k - 5)/(k - 1)$. For numerical purposes, a widely used model is the Variable Hard Sphere(VHS) model [2], corresponding to take $b_\alpha(\theta) = C_\alpha$ where C_α is a positive constant. The case $\alpha = 0$ is referred to as the Maxwellian gas, whereas the case $\alpha = 1$ yields the Hard Sphere gas.

2.2 Boundary conditions

Equation (1-2) usually is complemented with the boundary conditions for $v \cdot n \geq 0$ and $x \in \partial\Omega$

$$|v \cdot n| f(x, v, t) = \int_{v_* \cdot n < 0} |v_* \cdot n(x)| K(v_* \to v, x, t) f(x, v_*, t)\, dv_*. \tag{6}$$

The boundary condition (6) is the so-called reflective condition on $\partial\Omega$. The ingoing flux is defined in terms of the outgoing flux modified by a given boundary kernel K according to the integral in (6). This boundary kernel is such that positivity and mass conservation at the boundaries are guaranteed

$$K(v_* \to v, x, t) \geq 0, \qquad \int_{v \cdot n(x) \geq 0} K(v_* \to v, x, t)\, dv = 1. \tag{7}$$

From a physical point of view, we assume that at the solid boundary a fraction α 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 solid wall, while the remaining portion $(1 - \alpha)$ is perfectly reflected. This is equivalent to impose for the ingoing velocities

$$f(x, v, t) = (1 - \alpha) Rf(x, v, t) + \alpha M f(x, v, t), \tag{8}$$

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

$$Rf(x, v, t) = f(x, v - 2n(n \cdot v), t), \tag{9}$$

$$Mf(x, v, t) = \mu(x, t) f_s(v). \tag{10}$$

In (10), if we denote by T_s the temperature of the solid boundary, f_s is given by

$$f_s(v) = \exp(-\frac{v^2}{2T_s}),$$

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

$$\mu(x, t) \int_{v \cdot n \geq 0} f_s(v)|v \cdot n| dv = \int_{v \cdot n < 0} f(x, v, t)|v \cdot n| dv. \tag{11}$$

We note that according to (8), for $\alpha = 0$ (specular reflection) the re-emitted molecules have the same flow of mass, temperature and tangential momentum of the incoming molecules, while for $\alpha = 1$ (full accommodation) the re-emitted molecules have completely lost memory of the incoming molecules, except for conservation of the number of molecules.

2.3 Fluid dynamic limit

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

$$\int_{I\!R^3} Q(f,f)\phi(v)\,dv = 0, \quad \phi(v) = 1, v, v^2, \tag{12}$$

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

$$\int_{I\!R^3} Q(f,f)\log(f)dv \leq 0. \tag{13}$$

From a physical point of view, 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

$$M(\rho,u,T)(v) = \frac{\rho}{(2\pi T)^{3/2}}\exp\left(-\frac{|u-v|^2}{2T}\right), \tag{14}$$

where ρ, u, T are the density, mean velocity and temperature of the gas defined by

$$\rho = \int_{\mathbf{R}^3} f\,dv, \qquad u = \frac{1}{\rho}\int_{\mathbf{R}^3} vf\,dv, \qquad T = \frac{1}{3\rho}\int_{\mathbf{R}^3}[v-u]^2 f\,dv. \tag{15}$$

As $\varepsilon \to 0$ the distribution function approaches the local Maxwellian (14). In this case the higher order moments of the distribution f can be computed as function of ρ, u, and T, by using (14) and we obtain to the leading order the closed system of compressible Euler equations of gas dynamics

$$\begin{aligned}
&\frac{\partial\rho}{\partial t} + \nabla_x\cdot(\rho u) = 0\\
&\frac{\partial\rho u}{\partial t} + \nabla_x\cdot(\rho u\otimes u) + \nabla_x p = 0 \qquad (16)\\
&\frac{\partial E}{\partial t} + \nabla_x\cdot(Eu + pu) = 0\\
&p = \rho T, \quad E = \frac{3}{2}\rho T + \frac{1}{2}\rho u^2. \qquad (17)
\end{aligned}$$

2.4 Splitting approach

The most common approach to solve the full Boltzmann equation is based on an operator splitting. Consider the following initial value problem

$$\begin{aligned}
\frac{\partial f}{\partial t} + v\cdot\frac{\partial f}{\partial x} &= Q(f,f)\\
f(x,v,0) &= f_0(x,v)
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial\tilde{f}}{\partial t} &= Q(\tilde{f},\tilde{f})\\
\tilde{f}(x,v,0) &= f_0(x,v)
\end{aligned}$$

for a time step Δt, and then the transport equation (convection step), using the output of the previous step as initial condition,

$$\begin{aligned}
\frac{\partial f}{\partial t} + v\cdot\frac{\partial f}{\partial x} &= 0\\
f(x,v,0) &= \tilde{f}(x,v,\Delta t)
\end{aligned}$$

After obtaining an approximation of the solution at time Δt, the process may be iterated to obtain the numerical solution at later times.

The discretization of these equations can be performed in a variety of ways (particle methods, finite volume, 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.

In this paper we illustrate the use of particle methods, because we assume that we use a Monte Carlo method for the integration of the space homogeneous equation.

The splitting scheme (simple splitting) described above is first order accurate in space and time. Note that the order of accuracy does not improve even if we solve with great accuracy both collision and convection steps.

For the Monte Carlo method we shall use only the simple splitting. The motivation is that usually it is difficult to achieve a great accuracy with Monte Carlo method, and therefore it is not clear whether a better splitting strategy would improve the result. Another motivation is that the space discretization used in the usual splitting approach is not second order accurate, and it is not clear, at this stage, how to make it second order in space. This subject is a presently under investigation by the authors.

3 Direct Simulation Monte Carlo (DSMC) Methods

3.1 Random sampling

In this section we revise some basic features common to many applications of the Monte Carlo method. We assume that our computer is able to generate a uniformly distributed random number between 0 and 1.

3.1.1 Monovariate distribution

Let x be a random variable with density $p_x(x)$, $x \in I\!R$, i.e. $p_x(x) \geq 0$, $\int_\Omega p_x(x)\,dx = 1$, and let ξ be a uniformly distributed random number in $[0,1]$. Then the relation between x and ξ can be found by equating the infinitesimal probabilities

$$p_x(x)\,dx = d\xi$$

By integration one has

$$P_x(x) = \xi \tag{18}$$

where $P_x(x) = \int_{-\infty}^{x} p_x(y)\,dy$ 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 Eq. (18) for x,

$$x = P_x^{-1}(\xi) \tag{19}$$

Example 3.1 *Let* $p_x(x) = \exp(-x)$, $x \geq 0$. *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 a uniformly distrubuted* $[0,1]$ *random number.*

Remark 3.1 *We consider density functions defined on the whole real line. If the support of the density is strictly contained in* $I\!R$, *as in the previous example, the density function can be defined by using the Heaviside function. In the example above one would could define* $p_x(x) = \exp(-x)\Theta(x), x \in I\!R$.

Sometimes it may be expensive to the compute the inverse funtion, since in general a nonlinear equation has to be solved. In this case a different technique, called *acceptance-rejection*, can be used to sample a random variables with a given density.

Let x be a random variable with density $p_x(x)$, $x \in I\!R$. Then we look for a function $w(x) \geq p_x(x)\,\forall x \in I\!R$ whose primitive $W(x)$ is easily invertible. Let $A = \int_{-\infty}^{\infty} w(x)\,dx$. Then the algorithm works as follows

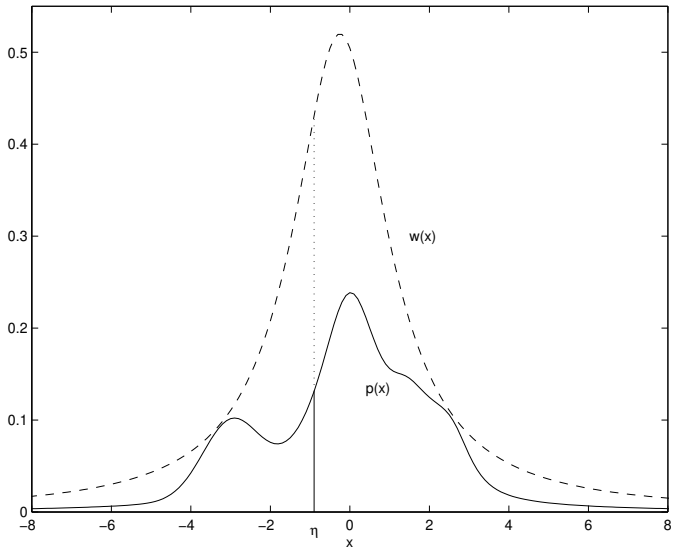

Figure 1: Acceptance-rejection procedure. η has been sampled from $w(x)$. It will be accepted with probability equal to the ratio $p(\eta)/w(\eta)$.

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.

Here ξ_1 and ξ_2 represent, as usual, uniformly $[0, 1]$ random numbers. It is clear that 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 $1/A$ (see Fig. 1).

Sometimes a densify 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

1. select an integer $i \in [1, \ldots, M]$ with probability w_i
2. sample x from a random variable with density $p_i(x)$.

Because of the relevance in several applications, point 1 deserves an extended discussion.

3.1.2 Discrete sampling

Let us suppose that $k \in [1, \ldots, M]$ 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, and extract a uniform random number ξ, and detect the interval k in which ξ belongs.

This can be done in in the following way.

1. Compute $W_k \equiv \sum_{i=1}^{k} w_k$, $W_0 \equiv 1$
2. find the integer k such that $W_{k-1} \leq \xi < W_k$.

For an arbitrary set of probabilities $\{w_i, \}$ the integer k can be computed by using a binary search, in $O(\ln M)$ operations.

This approach is efficient if the numbers W_k can be computed once, and then used several times, and if M is not too large. In several circumstances, however, the number M may be very

large, and the numbers w_k may change, or, in other words, one is interested in sampling k with probability w_k without explicitely computing all the probabilities w_k. This can be done by using the acceptance-rejection technique, if at least an estimate $\overline{w}$ such that $\overline{w} \geq w_i,\ i = 1, \ldots, M$ is known. The procesure is used as follows.

1. Select an integer random number uniformly in $[1, \ldots, M]$,

$$k = \lfloor M\xi_1 \rfloor + 1$$

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

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.

Sampling without repetition Sometimes it is useful to extract n numbers 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 3.1

```
for i = 1 to N
    ind(i) = i
end
M = N
for i = 1 to n
    j = ⌊Mξ⌋ + 1
    seq(i) = ind(j)
    ind(j) = ind(M)
    M = M − 1
end
```

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

3.1.3 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, i.e. if

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

then the n scalar random variables $x_1, \ldots x_n$ are independent, and they can be sampled separately, i.e. the problem is equivalent to sampling n monovariate random variables.

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

$$p_x(x_1, \ldots x_n)dx_1 dx_2 \ldots dx_n = p_{\eta_1}(\eta_1)p_{\eta_2}(\eta_2) \cdots p_{\eta_n}(\eta_n)d\eta_1 d\eta_2 \ldots d\eta_n, \tag{20}$$

sample the variables $\eta_1, \ldots \eta_n$, and then 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 $\xi = (\xi_1, \ldots \xi_n)$ 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 $p_x(x_1, \ldots x_n)$. An explicit transformation is constructed as follows:

$$T_1(x_1) \quad = \quad \int_{-\infty}^{x_1} d\eta \int_{I\!R^{n-1}} dx_2 \ldots dx_n\, p_x(\eta, \ldots x_n)$$

$$
\begin{aligned}
T_2(x_1, x_2) &= \frac{\int_{-\infty}^{x_2} d\eta \int_{I\!R^{n-2}} dx_3 \dots dx_n \, p_x(x_1, \eta, \dots x_n)}{\int_{-\infty}^{\infty} d\eta \int_{I\!R^{n-2}} dx_3 \dots dx_n \, p_x(x_1, \eta, \dots x_n)} \\
T_3(x_1, x_2, x_3) &= \frac{\int_{-\infty}^{x_3} d\eta \int_{I\!R^{n-3}} dx_4 \dots dx_n \, p_x(x_1, x_2, \eta, \dots x_n)}{\int_{-\infty}^{\infty} d\eta \int_{I\!R^{n-3}} dx_4 \dots dx_n \, p_x(x_1, x_2, \eta, \dots x_n)} \\
&\vdots \\
T_n(x_1, \dots, x_n) &= \frac{\int_{-\infty}^{x_n} d\eta \, p_x(x_1, \dots, \eta)}{\int_{-\infty}^{\infty} d\eta \, p_x(x_1, \dots, \eta)}
\end{aligned}
\tag{21}
$$

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 system (21), 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 3.2 *The transformation (21) 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 choosen 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 [11]*

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

1. sample from x from $w(x)/A$ by any known method,
2. if $w(x)\xi_2 < p_x(x)$ then accept the sample, else reject it and repeat step 1.

Example 3.2 *As an example we show how to sample from a Gaussian distribution. 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 distrubution function $P(x) = (1 + \mathrm{erf}(x/\sqrt{2})/2$, where $\mathrm{erf}(x)$ denotes the error funtion. However the inversion of the error function may be expensive. An alternative procedure is obtained by the following method. 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

$$
x = \rho \cos\theta, \;\; y = \rho \sin\theta \tag{22}
$$

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*

$$\begin{aligned} p_\rho &= \exp\left(-\frac{\rho^2}{2}\right)\rho, \quad \rho \geq 0 \\ p_\theta &= \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi \end{aligned}$$

The random variables ρ *and* θ *are readily sampled by inverting* $P\rho$ *and* $P\theta$, *i.e.*

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

and, finally, x *and* y *are computed from Eq.(22).*

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

$$x = \mu_x + \sigma_x\rho\cos\theta,\ y = \mu_y + \sigma_y\rho\sin\theta$$

Example 3.3 *Here we show how to sample a point uniformly from the surface of a 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

$$\begin{aligned} \frac{d\varphi}{2\pi} &= d\xi_1 \\ \frac{\sin\theta\,d\theta}{2} &= d\xi_2 \end{aligned}$$

Integrating the above expresisons we have

$$\begin{aligned} \varphi &= 2\pi\xi_1 \\ \theta &= \arccos(1-2\xi_2) \end{aligned}$$

3.2 Direct simulation schemes

In this section we consider the spatially homogeneous Boltzmann equation (18).

Here we will describe the classical DSMC methods [12, 2] in the case of the patially homogeneous Boltzmann equation (18).

The kinetic equations we are considering can be written in the form

$$\frac{\partial f}{\partial t} = \frac{1}{\epsilon}[P(f,f) - \mu f]. \tag{23}$$

In particular, for both Kac equation and Boltzmann equation for Maxwellian molecules we have $P(f,f) = Q^+(f,f)$ (see Remark 3.3), whereas for general kernels with cut-off $P(f,f)$ is given by (33).

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

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

3.2.1 Nanbu-Babovsky scheme

Let us discretize time, and denote by $f^n(v)$ an approximation of $f(v, n\Delta t)$. The forward Euler scheme applied to Eq. (23) writes

$$f^{n+1} = (1 - \frac{\mu\Delta t}{\epsilon})f^n + \frac{\mu\Delta t}{\epsilon}\frac{P(f^n, f^n)}{\mu}. \tag{24}$$

This equation has the following probabilistic interpretation: a particle with velocity v_i will not collide with probability $(1 - \mu\Delta t/\epsilon)$, and it will collide with probability $\mu\Delta t/\epsilon$, according to the collision law described by $P(f^n, f^n)(v)$.

For clarity of exposition, 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 [12] and can be described as follows

Algorithm 3.2

- 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 $n_{\rm TOT}$
  - for $i = 1$ to N
    - with probability $1 - \mu\Delta t/\epsilon$
      - set $v_i^{n+1} = v_i^n$
    - with probability $\mu\Delta t/\epsilon$
      - 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

In its first version, Nanbu's algorithm was not conservative, i.e. energy was conserved only in the mean, but not at each collision. A conservative version of the algorithm was introduced by Babovsky [1]. Instead of selecting single particles, independed particle pairs are selected, and conservation is maintained at each collision. The expected number of particle that collide in a small time step Δt is $N\mu\Delta t/\epsilon$, and the expected number of collision pairs is $N\mu\Delta t/2\epsilon$. The algorithm is the following

Algorithm 3.3

- 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 $n_{\rm TOT}$
  - given $\{v_i^n, i = 1, \ldots, N\}$, compute $\{v_i^{n+1}\}$ as follows:
    - set $N_c = \mathrm{Iround}(\mu N\Delta t/(2\epsilon))$
    - 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 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 been selected
  - end for

During each step, all the other $N - 2N_c$ particle velocity remain unchanged. Here by $\mathrm{Iround}(x)$ we denote a suitable integer rounding of a positive real number x. In our algorithm we choose

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

where $[x]$ denotes the integer part of x.

Remark 3.3 As we said before, the algorithm just described 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.

The Kac equation in fact reads [9]

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon}\int_{-\infty}^{+\infty} \frac{1}{2\pi}[f(v')f(v'_*) - f(v)f(v_*)]\, dv_*\, d\theta \tag{25}$$

where

$$v' = v\cos\theta - v_*\sin\theta, \qquad v'_* = v\sin\theta + v_*\cos\theta. \tag{26}$$

Hence, the previous scheme can be applied directely taking

$$P(f,f) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} f(v')f(v'_*)\, dv_*\, d\theta, \tag{27}$$

and $\mu = \rho$.

Since, the only conserved quantities are the total mass and the energy, i.e. the collision invariants are $\phi(v) = 1, v^2$. When the above scheme is applied, the new velocities v'_i and v'_j are computed as

$$v'_i = v_i\cos\theta - v_j\sin\theta, \quad v'_j = v_i\sin\theta + v_j\cos\theta, \tag{28}$$

where $\theta = 2\pi\,\mathrm{Rand}$ and Rand denotes a random number, uniformly distributed in $[0,1]$.

For Maxwell molecules one has

$$v'_i = \frac{v_i + v_j}{2} + \frac{|v_i - v_j|}{2}\omega, \quad v'_j = \frac{v_i + v_j}{2} - \frac{|v_i - v_j|}{2}\omega, \tag{29}$$

where ω is choosen uniformly in the unit sphere, according to:

2D:

$$\omega = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \theta = 2\pi\,\mathrm{Rand}, \tag{30}$$

3D:

$$\omega = \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix}, \quad \theta = \arccos(2\xi_1 - 1), \quad \phi = 2\pi\xi_2, \tag{31}$$

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

The above algorithm has to be modified when the scattering cross section is not constant. To this aim we will assume that the collision kernel satisfies numerically essential cut-off hypothesis.

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

$$\sigma_\Sigma(|v - v_*|, \omega) = \min\{\sigma(|v - v_*|, \omega), \Sigma\}, \quad \Sigma > 0.$$

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

$$\frac{\partial f}{\partial t} = \frac{1}{\varepsilon}Q_\Sigma(f,f). \tag{32}$$

Problem (32) can be written in the form (23) taking

$$P(f,f) = Q^+_\Sigma(f,f) + f(v)\int_{I\!R^3}\int_{S^2}[\Sigma - \sigma_\Sigma(|v - v_*|, \omega)]f(v_*)\, d\omega\, dv_*, \tag{33}$$

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

$$Q^+_\Sigma(f,f,) = \int_{I\!R^3}\int_{S^2}\sigma_\Sigma(|v - v_*|, \omega)f(v')f(v'_*)\, d\omega\, dv_*. \tag{34}$$

It is a simple exercise to verify that the functions $\phi(v) = 1, v, v^2$ are the collision invariants.

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

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

Algorithm 3.4

- compute the initial velocity of the particles, $\{v_i^0, i = 1, \dots, N\}$, by sampling them from the initial density $f_0(v)$
- for $n = 1$ to $n_{\rm TOT}$
  given $\{v_i^n, i = 1, \dots, N\}$, compute an upper bound Σ of the cross section
  - set $N_c = {\rm Iround}(N\rho\Sigma\Delta t/(2\epsilon))$
  - select N_c dummy collision paires (i, j) uniformly among all possible pairs, and for those
    - compute the relative cross section $\sigma_{ij} = \sigma(|v_i - v_j|)$
    - if $\Sigma\,{\rm Rand} < \sigma_{ij}$
      - 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

The upper bound Σ should be chosen as small as possible, to avoid inefficient rejection, and it should be computed fast. It would be too expensive to compute Σ as

$$\sigma_{\max} = \max_{ij} \sigma(v_i - v_j)$$

since this computation would require an $O(N^2)$ operations. A lower bound of $\sigma_{\max}$ is obtained by taking $\Sigma = \sigma(2\Delta v)$, where $\Delta v = \max_i |v_i - \bar v|$, and $\bar v := \sum_i v_i/N$.

The new velocities v_i' and v_j' are computed using Eqs. (29–31).

For general collisional kernel, the algorithm is slightly modified by introducing the angular dependence. In this case the collision is always performed, and the new velocities are extracted according to the differential cross section. Then a rejection is used to decide whether the collision is accepted.

We remark here that the probabilistic interpretation of Eq. (24) breaks down if $\Delta t/\epsilon$ 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 the classical DSMC method becomes unusable near the fluid regime.

3.2.2 Bird's scheme

4 Time Relaxed Monte Carlo (TRMC) methods

4.1 Motivations

When a gas is near thermodynamical equilibrium, a detailed kinetic description is unnecessary, and the evolution of the moments can be described by Euler or Navier-Stokes equations. From a computational point of view, these equations are much less expensive than the full Boltzmann equation.

Far from thermodynamic equilibrium, the equation of gas dynamics do not give a satisfactory description of the physical system, and the full kinetic description is necessary.

The departure from local thermodynamic equilibrium can be measured by the so called Knudsen number, which is the ratio between the collisional mean free path of the molecules and the characteristic length of variation of macroscopic variables. When the Knudsen number is small, then the distribution function is close to the local Maxwellian. Collisions occur at a fast rate, and a kinetic treatment of the system is very expensive, because of the large ratio of time scales between macroscopic and microscopic effects. One may say that the system is numerically *stiff* [7]. Because of the stiffness, standard explicit schemes for the numerical solution of the Boltzmann equation become extremely expensive in this case.

If the Knudsen number is uniformly small in the region one is interested in, then one should use a gas dynamic description of the system. There are several circumstances, however, where the Knudsen number varies over several orders of magnitude. In such cases a standard kinetic treatment would be too expensive, and a pure gas dynamic approach inappropriate. Domain

decomposition methods have been proposed for this problem, in which the computational domain is divided into an *aerodynamic region* in which the system is treated by gas dynamic equations (Euler or Navier-Stokes), and a *Boltzmann region*, where a kinetic treatment is used. Suitably boundary conditions make the two regions coupled together [3]. Such schemes can be very effective in treating large scale problems, but they are quite complicated to implement, and they require a great care by the user.

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 [15, 16, 14].

These algorithms are based on a suitable time discretization of the Boltzmann equation, first introduced in [6], which has the following properties: for large Knudsen number it is a time discretization of the Boltzmann equation of a given order, and when the Knudsen number vanishes, the numerical solution tends to the local Maxwellian. We will call this last property *asymptotic preservation* property. In addition the time discretization preserves mass, momentum and energy. The stability property is similar to the L-stability of numerical schemes for systems of ode's [7].

Although this chapter deals with the space homogeneous case, the real motivation for the development of these Monte Carlo methods is their use for spatially dependent problems. Because the distribution function approaches the local Maxwellian for small Knudsen numbers, a Monte Carlo method based on these time discretizations will naturally behave as a sort of kinetic schemes for the Euler equations in the fluid dynamic limit.

4.2 Asymptotic preserving schemes

Time relaxed (TR) schemes

As proposed in [6], a general idea for deriving unconditionally stable and asymptotic preserving numerical schemes for a nonlinear equation like (18) is to replace high order terms of a suitable well-posed power series expansion by the local equilibrium. This class of schemes has been called *Time Relaxed* (TR) schemes.

Here we will briefly recall the schemes presented in [6] and some generalization used in [16].

Derivation

Let us consider a differential system of the type

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

with the initial condition

$$f(0) = f_0 \tag{2}$$

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

Then the function f can be expressed by the following formal expansion (Wild sum) [21, 6]

$$f(t) = e^{-\mu t/\varepsilon} \sum_{k=0}^{\infty} \left(1 - e^{-\mu t/\varepsilon}\right)^k f_k, \tag{3}$$

where the functions f_k are given by the recurrent formula

$$f_{k+1} = \frac{1}{k+1} \sum_{h=0}^{k} \frac{1}{\mu} P(f_h, f_{k-h}), \quad k = 0, 1, \ldots \tag{4}$$

From this representation, a class of numerical schemes can be naturally derived. Let us introduce the scaled time $\tau = 1 - \exp(-\mu t/\varepsilon)$.

In [6] the following class of numerical schemes, based on a suitable truncation for $m \geq 1$ of (3), has been constructed

$$f^{n+1}(v) = e^{-\mu \Delta t/\varepsilon} \sum_{k=0}^{m} \left(1 - e^{-\mu \Delta t/\varepsilon}\right)^k f_k^n(v) + \left(1 - e^{-\mu \Delta t/\varepsilon}\right)^{m+1} M(v), \tag{5}$$

where $f^n = f(n\Delta t)$ and Δt is a small time interval. The quantity M (referred as the local Maxwellian associated to f) is the asymptotic stationary solution of the equation.

It can be shown that the schemes obtained in this way are of order m in time. Furthermore, they satisfy the following properties:

i) *conservation:*

If $P(f,g)$ is a non negative bilinear operator such that there exist some functions $\phi(v)$ with the following property

$$\int_{R^3} P(f,f)\phi(v)\,dv = \mu \int_{R^3} f\phi(v)\,dv, \tag{6}$$

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

$$\int_{R^3} f^{n+1}\phi(v)\,dv = \int_{R^3} f^n\phi(v)\,dv; \tag{7}$$

ii) *asymptotic preservation:*

For any $m \geq 1$, we have

$$\lim_{\mu\Delta t/\varepsilon \to \infty} f^{n+1} = M(v). \tag{8}$$

When applied to the Boltzmann equation, the first property guarantees exact conservation of mass, momentum, and energy, for the semi-discrete scheme, while the second property guarantees the correct fluid dynamic limit.

Generalization

The approach we have seen in the previous sections can be generalized using different weight functions to combine the influence of the high order coefficients appearing in the Wild sum (3). In general such schemes can be written as

$$f^{n+1} = \sum_{k=0}^{m} A_k f_k + A_{m+1} M, \tag{9}$$

where the coefficients f_k are given by Eq. (4). The weights $A_k = A_k(\tau)$ are nonnegative function that satisfy the following properties

i) *consistency:*

$$\lim_{\tau\to 0} A_1(\tau)/\tau \quad = \quad 1, \quad \lim_{\tau\to 0} A_k(\tau)/\tau = 0, \quad k = 2,\ldots,m+1 \tag{10}$$

ii) *conservation:*

$$\sum_{k=0}^{m+1} A_k = 1 \quad \tau \in [0,1], \tag{11}$$

iii) *asymptotic preserving:*

$$\lim_{\tau\to 1} A_k(\tau) = 0, \quad k = 0,\ldots,m \tag{12}$$

Such properties are a generalization of those defined in [16] for first and second order schemes. It is possible to show that a scheme that satisfies these properties will satisfy properties (7) and (8).

A choice of functions which satisfies the previous requirements is given by

$$A_k = (1-\tau)\tau^k,\ k = 0,\ldots,m, \quad A_{k+1} = \tau^{m+1}, \tag{13}$$

which correspond to the scheme (5).

However, other choices are possible and it is an open problem the determination of the *optimal* set of functions A_k that satisfies the previous requirements and guarantees the most accurate approximation. Some preliminary results in this direction will be shown in the next section.

4.3 Analysis of the TR schemes

Linear stability

In this section we perform a stability analysis of the TR schemes. In particular, we show that the first and second order schemes used in [6] are L-stable.

Let us apply the TR scheme to the test equation

$$\left\{ \begin{array}{rcl} y' & = & \lambda y \\ y(0) & = & 1 \end{array} \right. \tag{14}$$

with $\lambda \in C^-$. The equation can be written in the form

$$y' = P(y, y) - \mu y$$

with $P(f, g) = \frac{1}{2}(\mu + \lambda)(f + g)$, and μ is a positive constant such that $|\lambda| < \mu$. The numerical solution after one time step is given by

$$y^1 = \sum_{k=0}^{m} A_k y_k \tag{15}$$

The $m+1$ term vanishes because the equilibrium solution is zero. The term y_k can be explicitly computed by applying the formula (4). A simple calculation yields

$$y_k = \frac{1}{k!} \prod_{j=1}^{k} (j + a) = \frac{\Gamma(k + 1 + a)}{\Gamma(k + 1)\, \Gamma(1 + a)}, \quad k = 0, \ldots, m \tag{16}$$

where $a = \lambda/\mu$. The numerical solution after one time step, called the *function of absolute stability*, is therefore a function of a real parameter τ and a complex parameter a

$$R_m(\tau, a) = \sum_{k=0}^{m} A_k(\tau) y_k(a),$$

where $\tau = 1 - \exp(-\mu \Delta t)$ is the relaxed time, and a is a complex number $a \in C^-, |a| \leq 1$. Here we give definitions of A- and L-stability which are the natural generalization of the standard definition [7] to the case where the function of absolute stability depends on an additional real parameter.

Definition 4.1 (A-stability) *We say that the scheme is A-stable if*

$$|R(\tau, a)| \leq 1 \quad \forall \tau \in [0, 1],\ a \in C^-, |a| \leq 1$$

Definition 4.2 (L-Stability) *We say that the scheme is L-stable if it is A-stable and*

$$\lim_{\tau \to 1} R(\tau, a) = 0, \quad \forall a \in C^-, |a| \leq 1$$

Definition 4.3 ($A\alpha$-stability) *We say that the scheme is $A\alpha$-stable if*

$$|R(\tau, a)| \leq 1 \quad \forall \tau \in [0, 1],\ a = -\rho \exp(i\theta),\ \rho \leq 1, \theta \in [-\alpha, \alpha]$$

Definition 4.4 ($L\alpha$-Stability) *We say that the scheme is $L\alpha$-stable if it is $A\alpha$-stable and*

$$\lim_{\tau \to 1} R(\tau, a) = 0 \quad \forall a = -\rho \exp(i\theta),\ \rho \leq 1, \theta \in [-\alpha, \alpha]$$

It is clear that, because of the asymptotic preserving properties, if a TR scheme is A-stable it is also L-stable.

Here we prove that the first and second order schemes (5) are in fact A-stable (and therefore L-stable).

We conjecture that the A-stability property is shared by schemes (5) of all order, but we have not been able to prove it.

First order schemes Let us consider the first order scheme given by (13, 15) for $m = 1$

$$y^1 = (1-\tau)y_0 + \tau(1-\tau)y_1. \tag{17}$$

The function of absolute stability in this case is given by

$$R_1(\tau, a) = (1-\tau)[1 + \tau(1+a)].$$

We prove the following

Theorem 4.1 *The scheme (17) is A-stable according to Def. 4.1, i.e.*

$$|R_1(\tau, a)| \leq 1, \forall \tau \in [0,1],\ a \in C^-, |a| \leq 1.$$

Proof

Let $a = -\rho\exp(i\theta),\ \theta \in [-\pi/2, \pi/2],\ \rho \in [0,1],$ *and let*

$$F_1(\tau, \rho, \theta) = 1 - |R_1(\tau, -\rho e^{i\theta})|^2.$$

Direct computation shows that

$$F_1 = A + B\cos\theta$$

with $A = \tau^2[(1-\tau^2) + 1 - \rho^2(1-\tau)^2]$, $B = 2\rho\tau(1+\tau)(1-\tau)^2$

Because of the range of the variables, it is $F_1 \geq A \geq 0$, *and therefore* $|R_1(\tau, a)| \leq 1$.

■

Second order schemes Now we consider the second order scheme given by (13, 15) with $m = 2$,

$$y^1 = (1-\tau)[y_0 + \tau y_1 + \tau^2 y_2]. \tag{18}$$

The function of absolute stability in this case is given by

$$R_2(\tau, a) = (1-\tau)[1 + \tau(1+a) + \frac{1}{2}\tau^2(1+a)(2+a)].$$

Let $\Omega_a \equiv \{z \in C : \Re z \leq 0, |z| \leq 1\}$. We observe that $R_2(\tau, a)$ is analytic in a, and therefore, for any fixed value of the parameter $\tau \in [0,1]$, the function takes the maximum absolute value on the border Γ_a of Ω_a. Using this property, we can prove the following

Theorem 4.2 *The scheme (18) is A-stable according to Def. 4.1, i.e.*

$$|R_2(\tau, a)| \leq 1, \forall \tau \in [0,1], a \in \Omega_a \equiv \{z \in C : \Re z \leq 0, |z| \leq 1\}.$$

Proof

Let $a = -\rho\exp(i\theta)$. *Then* $\Gamma_a = \partial\Omega_a = \Gamma_1 \cup \Gamma_2$, *where* $\Gamma_1 = \{z \in C : z = iy, y \in [-1,1]\}$, *and* $\Gamma_2 = \{z \in C : z = -e^{i\theta}, \theta \in [-\pi/2, \pi/2]\}$.

Direct calculation shows that

$$1 - |R_2(\tau, a)|^2 = \frac{1}{4}\tau F_2,$$

with

$$\begin{aligned} F_2(\tau, -\rho e^{i\theta}) &= -8\tau\rho^2(1+\tau)(1-\tau^3)\cos^2\theta \\ &+ 2\rho(3\tau+2)(1-\tau)^2(\rho^2\tau^2 + 2\tau^2 + 2\tau + 2)\cos\theta \\ &+ \tau^2[8 - 4\tau^3 - (5\tau+8)(1-\tau)^2\rho^2 - \tau(1-\tau)^2\rho^4] \end{aligned}$$

The restriction of F_2 *to* Γ_1 *is*

$$F_2(\tau, iy) = (A\xi^2 + B\xi + C),$$

where

$$\xi \equiv y^2,\ A = -\tau^3(1-\tau)^2,\ B = -\tau^2(5\tau+8)(1-\tau)^2,\ C = (8-4\tau^3)\tau^2.$$

It is $F_2(\tau, 0) = \tau^2(8 - 4\tau^3) \geq 0$, $F_2(\tau, i) = \tau^3(5 + 2\tau - 5\tau^2) \geq 0$,*and* $A \leq 0$, *therefore* $F_2(\tau, iy) \geq 0, \forall y \in [-1,1]$.

The restriction of F_2 *to* Γ_2 *is*

$$F_2(\tau, e^{i\theta}) = \tilde{F}(\tau, x) \equiv \tilde{A}x^2 + \tilde{B}x + \tilde{C},$$

where

$$x = \cos\theta, \quad \tilde{A} = -8\tau(1+\tau)(1-\tau)^3,$$
$$\tilde{B} = 2(3\tau+2)(1-\tau^2)(3\tau^2+2\tau+2), \quad \tilde{C} = 2\tau^3(5+2\tau-5\tau^2).$$

It is $\tilde{F}(\tau, 0) \geq 0$, $\tilde{F}(\tau, 1) = 4 - 2\tau > 0$, *and* $A \leq 0$, *therefore* $\tilde{F}(\tau, x) \geq 0, \forall \tau \in [0, 1]$, $\theta \in [-\pi/2, \pi/2]$.

■

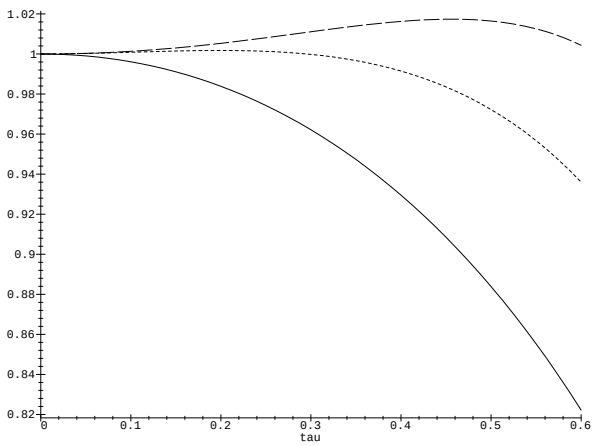

Figure 2: $R(\tau, -1/2)/\exp(-\mu\Delta t/2)$ vs τ. First order schemes. Continuous line: $r = 0$, dotted line: $r = 1$, dashed line: $r = 2$

Accuracy consideration

As we said in the previous section, the choice of the weights (5) is not necessarily the best for numerical computations. In fact, numerical experiments [16] suggest that other choices give better results. Let us consider the family of schemes defined by the following coefficients

$$A_k = (1-\tau)\tau^k, \; k = 0, m-1, \quad A_m = (1-\tau^{r+1})\tau^m, \quad A_{m+1} = \tau^{m+r+1}, \tag{19}$$

where $r = 0, 1, \ldots$. Schemes (5) are a particular member of this family, and correspond to the choice $r = 0$. Schemes with $r \geq 1$ are more accurate than the schemes corresponding to $r = 0$ for small time steps. This can be seen by computing the leading error term in the Taylor expansion of the difference between the numerical solution and the exact solution. The constant in front the term Δt^{m+1} is in fact smaller for $r \geq 1$ than for $r = 0$.

A good choice seems $r = 1$. This is suggested by the comparison between the exact solution and the function $R_m(\tau, a)$ for real values of $a \in [-1, 0]$.

In Fig. (2) we show the ratio $R_1(\tau, -1/2)/e^{-\mu t/2}$, as a function of the scaled time τ, for the first order schemes, corresponding to $r = 0, 1, 2$. In Fig. (3) the same ratio are shown, for second order schemes. It is evident that for $r = 1$ the ratio is very close to 1, even for considerable values of the parameter τ.

We remark here that for this more general schemes, A-stability is not guaranteed, as can be explicitly checked for the first order scheme with $r = 1$. It can be shown, however, that this scheme is A-α stable, with $\alpha \approx \pi/2 - 0.05$.

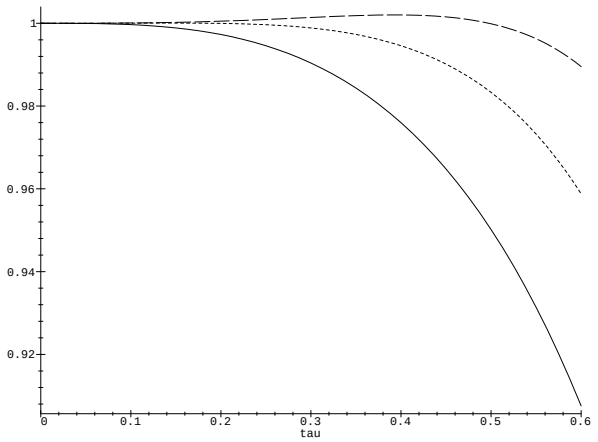

Figure 3: $R(\tau, -1/2)/\exp(-\mu\Delta t/2)$ vs τ. Second order schemes. Continuous line: $r = 0$, dotted line: $r = 1$, dashed line: $r = 2$

4.4 Time Relaxed Monte Carlo (TRMC) methods

In this section we describe a new class of simulation schemes for the Boltzmann equation, the *Time Relaxed Monte Carlo* (TRMC) methods for the evolution of the density function f.

Monte Carlo methods are naturally derived from the probabilistic interpretation of time discrete schemes. Since we deal only with the space homogeneous case, we shall make the simplifying assumption that the function $f^n(v)$ is a probability distribution, i.e.

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

We recall that the probability interpretation of the DSMC scheme we have seen in the previous chapter is valid as long as $\mu\Delta t \leq 1$. This corresponds to the stability restriction of the explicit Euler scheme.

Before we describe TR schemes, a remark is in order. We observe that in actual implementations, one choses a given number of *dummy collision pairs*, equal to $\mu N\Delta t/2$. Such pairs are sampled uniformly among all possible pairs. The parameter μ is obtained by an estimate Σ of the maximum cross section between all particle pairs. Then dummy collisions are performed, and each of them is accepted with a probability which depends on the ratio $\sigma(v_i, v_j)/\Sigma$. If we do not had to compute Σ, then the computational cost of such Monte Carlo method would be independend on the time step, since its cost would be proportional to the number of collisions, which is basically independent on the time step, for small time steps. This is the case, for example, of Maxwell molecules. However, if we compute Σ at every time step, then the cost of the computation increases when we decrease the time step, since the estimate of Σ has a cost $\mathcal{O}(N)$ and not $\mathcal{O}(N\Delta t)$.

4.4.1 TRMC methods

First order TR schemes are written as

$$f^{n+1} = A_0 f^n + A_1 f_1 + A_2 \tag{20}$$

The probabilistic interpretation of the above equation is the following.

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 replaced by a particle sampled from a Maxwellian with probability A_2.

In this formulation the probabilistic interpretation holds uniformly in $\mu\Delta t$, at variance with standard DSMC, which requires $\mu\Delta t < 1$. Furthermore, as $\mu\Delta t \to \infty$, the distribution at time $n+1$ is sampled from a Maxwellian. In this limit, the density f^{n+1} relaxes immediately to its equilibrium distribution. In a space non homogeneous case, this would be equivalent to the particle method for Euler equations proposed by Pullin [17].

The above procedure can be formalized in the following algorithm

Algorithm 4.1 (TRMC1)

- 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 $n_{\rm TOT}$
  given $\{v_i^n, i = 1, \ldots, N\}$, compute an upper bound Σ of σ_{ij} (as in DSMC)
  - compute $\tau = \exp(-\Delta t \Sigma \rho/\epsilon)$ and the corresponding quantities A_i, $i = 0, 1, 2$
  - set $N_c = \text{Iround}((A_1 + A_2)N/2)$, $N_1 = \text{Iround}(A_1 N/2)$, $N_2 = 2(N_c - N_{c1})$
  - perform N_1 dummy collisions, i.e.
    repeat N_1 times {
    - extract (i, j) without repetition
    - compute $\sigma_{ij} = \sigma(|v_i - v_j|)$
    - if Σ Rand $< \sigma_{ij}$
      perform the collision between v_i and v_j and compute v_i' and v_j' according to the collisional law
      set $v_i^{n+1} = v_i'$, $v_j^{n+1} = v_j'$
    - end if
    }
  - replace N_2 particles by a particle sampled from a Maxwellian, i.e.
    repeat N_2 times {
    - extract i without repetition and store index i
    - compute mean velocity U and temperature T of selected particles
    }
    - sample N_2 particles w_i from a Maxwellian with mean velocity U and temperature T
    repeat N_2 times {
    - for the stored index i set $v_i^{n+1} = w_i$
    }
  - set $v_i^{n+1} = v_i^n$ for all the particles that have not collided

end for

A second order Monte Carlo scheme is obtained by the TR scheme

$$f^{n+1} = A_0 f^n + A_1 f_1 + A_2 f_2 + A_3 M, \tag{21}$$

with

$$f_1 = \frac{P(f^n, f^n)}{\mu}, \quad f_2 = \frac{P(f^n, f_1)}{\mu}.$$

The probabilistic interpretation of scheme (21) is the following. Given N particles distributed according to f^n:

- NA_0 particles do not collide,
- NA_1 are sampled from f_1, as in the first order scheme,
- NA_2 are sampled from f_2, i.e. $NA_2/2$ particles sampled from f^n will undergo dummy collisions with $NA_2/2$ particles sampled from f_1,
- NA_3 particles are sampled from a Maxwellian.

The Monte Carlo schemes described above are conservative in the mean. It is possible to make them exactly conservative [16]. This goal is achieved by

- selecting collision pairs uniformly, rather than individual particles,
- using a suitable algorithm for sampling a set of particles with prescribed momentum and energy from a Maxwellian [18].

We remark here that the TR scheme is a direct consequence of the time discretization of the Boltzmann equation. In this respect it is not crucial to select a particular Monte Carlo scheme in order to perform the comparison. Here we considered a commonly used Monte Carlo scheme. If a different technique is used to model the collision, then the same technique could be used in conjunction with TR disretization.

4.4.2 Hybrid TRMC methods

The variance of the Monte Carlo solution can be considerably reduced if one represents the equilibrium part of the distribution function analytically by a Maxwellian. Such idea has been recently presented in the paper [14]. One makes the *ansatz* that the distribution function is the sum of a continuous and a discrete part,

$$f^n(v) = (1-\beta^n)g^n(v) + \beta^n M(v).$$

By substituting this expression in the first order scheme (20), one obtains the evolution equations for β^n and g^n

$$\beta^{n+1} = A_0\beta^n + A_1(\beta^n)^2 + A_2 \tag{22}$$

$$g^{n+1} = p_1 g^n + p_2\left[q_1\frac{P(g^n,g^n)}{\mu} + q_2\frac{P(g^n,M)}{\mu}\right], \tag{23}$$

where

$$p_1 = \frac{A_0}{A_0 + A_1(1+\beta^n)}, \quad p_2 = \frac{A_1(1+\beta_n)}{A_0 + A_1(1+\beta_n)}, \tag{24}$$

$$q_1 = \frac{1-\beta^n}{1+\beta^n}, \quad q_2 = \frac{2\beta^n}{1+\beta^n}. \tag{25}$$

Note that $p_1, p_2 \geq 0$, $p_1 + p_2 = 1$, $q_1, q_2 \geq 0$, $q_1 + q_2 = 1$.

The probabilistic interpretation of the evolution equation for g^n is similar to the one for the purely particle scheme.

An acceptance-rejection technique, similar to the one used for DSMC, can be adopted. The conservative algorithm to update β^n and g^n can be written as

Algorithm 4.2 (TRMCH) :

1. *set* $\beta = 0, N_c = 0$
2. *compute an upper bound* Σ *of* σ_{ij} *(as in DSMC)*
3. *compute* $\tau = \exp(-\rho\Sigma\Delta t/\epsilon)$ *and the corresponding quantities* $A, B, C, p_1, p_2, q_1, q_2$
4. *compute the number of dummy collision pairs:*
   - $N_{gM} = p_2 q_2 N_n/2$
   - $N_{gg} = p_2 q_1 N_n/2$
5. *perform* N_{gg} *dummy collisions between g-particles, i.e.*
   - *extract* (i,j) *without repetition*
   - *if* $\Sigma\,\mathrm{Rand} < \sigma_{ij}$ *then perform the collision between* v_i *and* v_j *(as in standard DSMC)*
     - $\beta \leftarrow \beta + 2\beta_{ij}$
     - $N_c \leftarrow N_c + 2$
6. *perform* $2N_{gM}$ *dummy collisions between the g-particles and the Maxwellian,*
   - *extract* i *without repetition*
   - *sample one particle,* m*, from the Maxwellian*
   - *if* $\Sigma\,\mathrm{Rand} < \sigma_{ij}$ *then perform the collision between* v_i *and the Maxwellian*
     - $\beta \leftarrow \beta + \beta_{im}$
     - $N_c \leftarrow N_c + 1$

7. $\beta^{n+1} = \beta/N_c$
8. *update* N: $N_{n+1} = \text{Round}(N_0(1-\beta^{n+1}))$
9. *correct* β^{n+1} *in order to preserve mass*
10. *correct* u_M *and* T_M *to maintain conservation of momentum and energy.*

The above scheme conserves momentum and energy only on the average, but not exactly. This is because the collisions with the Maxwellian M, if performed independently from each other, do not maintain exact conservation of momentum and energy. By taking this into account, a conservative algorithm can be constructed by changing momentum u_M and energy E_M of the Maxwellian fraction after each collision (*local conservation*), according to

$$u'_M = u_M - \frac{v'_j - v_j}{N^0\beta^n}, \quad E'_M = E_M - \frac{(v'_j)^2 - (v_j)^2}{2N^0\beta^n}, \quad n > 0, \tag{26}$$

where

$$E_M = \frac{1}{2}(dT_M + u_M^2),$$

d is the dimension of the velocity space and T_M is the temperature of the Maxwellian.

Alternatively, conservation can be restored by modifying the moments associated to the Maxwellian fraction at the end of each time step (*global conservation*).

This is obtained by imposing

$$(1-\beta^{n+1})E_p + \beta^{n+1}E_M = E^0, \quad (1-\beta^{n+1})u_p + \beta^{n+1}u_M = u^0, \quad n > 0,$$

where u_p and E_p are the mean velocity and energy of the particles.

The two approaches give very similar results and can be used if β^n is not too small.

If the distribution is very far from equilibrium, i.e. if $\beta_n \ll 1$, then because of fluctuations, it may happen that the energy decreases too much, and it is impossible to change the parameters of the Maxwellian to impose conservation. On the other hand, in this case, only a very small fraction of collisions will be non conservative, and therefore the lack of exact conservation will not affect the quality of the result.

Note that with this approach the number of particles is a function of time (it is $N^n \approx N^0(1-\beta^n)$).

The advantage of this approach with respect to the purely particle method is twofold: the variance of the distribution is reduced because a fraction of the distribution function is represented analytically, and the scheme is more efficient, because the number of particles is reduced.

4.4.3 Recursive TRMC methods

4.5 Numerical Examples

In this section we test the different TRMC methods presented in the previous section by comparing them with the standard DSMC. Since we are interested in checking the accuracy of the schemes we consider two initial value problems, respectively for the Kac equation and for the Boltzmann equation for Maxwell molecules, for which we have an exact solution. To simplify notations we denote by TRMC1, TRMC2 and TRMCH respectively the first order and second order TRMC methods and the hybrid TRMC method. In our tests we use the family of schemes characterized by (19) with $r = 1.$ and perform a single run, with a number of particles sufficiently large to control the effects of the fluctuations. We express the results as a function of the scaled time variable t/ϵ which we denote again by t.

First we compare the Monte Carlo solutions with the exact solution of the Kac equation [9]

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

where

$$C(t) = \left[3 - 2e^{-\sqrt{\pi}t/16}\right]^{-1}.$$

We reconstruct the density function on a grid, by convolving the particle distribution by a suitable mollifier [14, 16].

All the simulations have been performed for $t \in [0, 6]$ by starting with $N = 5 \times 10^4$ particles.

In Fig. 4 we report the L^2-norm of the error obtained with the different schemes using the same time step $\Delta t = 0.5$.

It is evident that all the TRMC methods provide a more accurate representation of the solution. In particular, by comparing the solution obtained with DSMC and TRMC1 one can see the improvement given by the first order TR scheme with respect to the explicit Euler time discretization. It is remarkable that the TRMCH method, as a consequence of its deterministic evolution of the equilibrium part, yields the most accurate solution.

Fig. 5 shows the L^2-norm of the error for the same test where the time step of the TRMC schemes are respectively two times (TRMC1), three times (TRMC2) and four times (TRMCH) larger then the one used by DSMC. Note that in spite of the different time steps the accuracy of the different schemes is similar.

Next we consider the numerical solution of the two-dimensional homogeneous Boltzmann equation for Maxwell molecules.

An exact solution for this equation is given by

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

where $C(t) = 1 - (1/2)e^{-t/8}$.

The density distribution is obtained by reconstructing the function on a regular grid of spacing $\Delta v = 0.25$ by the "weighted area rule" [8] and the simulations have been performed for $t \in [0, 16]$ by starting with $N = 10^5$ particles.

In Fig. 6, 7 we show the time evolution of the fourth order moment of the solution. The results confirm the gain of accuracy and the reduction of fluctuations of the TRMC methods with respect to the DSMC method for larger time steps.

We conclude this section with some additional considerations on the efficiency of TRMC methods.

First, in spite of the fact that the time step for TRMC methods can be chosen larger than that of DSMC, the number of dummy collision is higher for DSMC. The reason is that this number is proportional to $\mu\Delta t$ for DSMC, and it is proportional to $1 - \exp(-\mu\Delta t)$ for TRMC. This is a further advantage of TRMC methods with respect to DSMC.

Second, we emphasize that TRMC2 takes into account re-collisions of particles during the same time step. Thus, although the direct combination of the previous schemes with a solver for the transport step gives a scheme which is still first order, we expect that the overall accuracy of the solution will be improved also in space non homogeneous problems. Furthermore, it is well known that a suitable combination of collision and convection steps may lead to second order accurate schemes both in space and time steps (an application to deterministic methods for the Boltzmann equation is given for example, in [13]). In a future work, we hope to construct a second order TRMC scheme for non homogeneous Boltzmann equation that works uniformly with respect to the Knudsen number.

A similar effect is more evident in the homogeneous test problems we have presented here, where the accuracy of the virtually first order TRMCH is higher than the accuracy of the second order TRMC2. In fact, the reduction of fluctuation given by the deterministic evolution of the Maxwellian part is stronger than the reduction of fluctuation given by the re-collision of particles.

5 Monte Carlo Methods: space non homogeneous case

In this section we consider the full space non homogeneous Boltzmann equation. We shall illustrate the method with some simple application in one space dimension.

The splitting approach with particle method is particularly simple. The region Ω is divided into a certain number N_c of cells, and for each cell J one counts the number of particles per

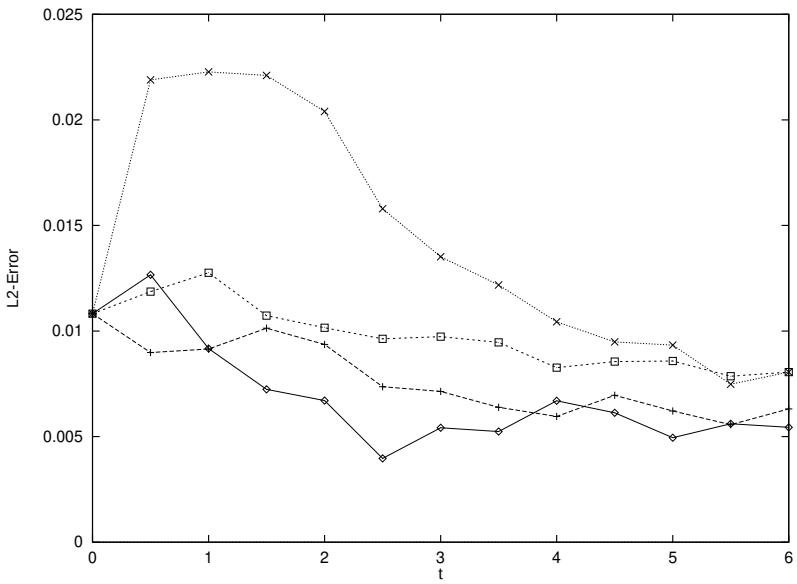

Figure 4: Kac equation : L^2 error vs time with $\Delta t = 0.5$ for DSMC ($\times$), TRMC1 ($\Box$), TRMC2 ($+$) and TRMCH ($\diamond$).

cell, N_J. Then during the collision step, the space homogeneous Boltzmann equation is solved by Monte Carlo in each cell, assuming that the density is constant in each cell. Each particle may collide only with particles within the same cell.

The solution of the collision step in the time interval t_n, t_{n+1} can be performed by one of the techniques illustrated in the previous sections (Nanbu-Babovsky scheme, first or second order TRMC, and so on). In the space homogeneous case we assumed that the density function was a probability density, and therefore its in integral over velocity was one. In the space non homogeneous case, the integral in velocity gives the number density of the the gas, i.e.

$$\int_{R^3} f(x, v, t)\, \mathrm{d}v = \frac{\rho}{m}.$$

The treatment of the previous section, with the modification that the parameter μ in Eq. (23) is in this case given by

$$\mu = 4\pi\Sigma \frac{\rho}{m}$$

The density ρ in each cell, in turn, is proportional to the number of particles in the cell. More presisely, the density is given by

$$\rho_J = N_J m^*$$

where m^* is the "mass" of the simulation particles, which in general is much larger than the effective mass of a particle. The reason of this is that one wants to simulate physical systems with a large number of particles, by a much smaller number of simulation particles. As a result, a simulation particle usually represents several physical particles. Detailed discussions about this can be found, for example, in [2].

During the collision step, mass, momentum and energy in each cell remain constant.

At time $t_n = n\Delta t$, each particle is characterized by a position, x_i^n, and a velocity, v_i^n. Before the collision step, one has to determine the particle belonging to each cell. From the point of view of the implementation, one possibility to achieve this is to perform a sweep on all the particles and to determine in which cell each particle lies. This can be obtained, for example, by creating two vectors, $\mathtt{kp}[N]$ and $\mathtt{cp}[N_c]$, where N is the total number of paritcles, with the following meaning: the particles in cell J are all the partiles with index $\mathtt{kp}[k]$, with $\mathtt{cp}[J-1] < k \leq \mathtt{cp}[J]$. Such ordering of the particles can be constructed very quickly in $O(N)$ operations.

After the collision step, the velocity of the paticles is updated, and each particle has been assigned a new velocity, let us call it $\tilde{v}_i$, $i = 1, \ldots, N$.

5.1 Transport of particles

The convection step is performed very easily, with a single loop over all the particles, namely

$$\tilde{x}_i = x_i^n + \tilde{v}_i \Delta t, \quad i = 1, \ldots, N$$

If there are no boundaries, then the new position and velocities of the particles at the new time t_{n+1} will simply be

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

The presence of boundaries will be considered in the next section.

As we shall see, the time step Δt is mainly determined by the requirement that the particles do not travel more than a few cells in one time step. This requirement is related to a stability requirement of upwind schemes for the numerical solution of the single scalar equation.

The time step used in the numerical solution of the space homogeneous Boltzmann equation, on the other hand, is determined by different requirements. For example, in the case of classical DSMC, one has to require positivity of the probability, and therefore the collisional time step must satisfy the restriction

$$\frac{\rho}{m} \Sigma \Delta t_c \leq 1 \tag{27}$$

This condition can be guaranteed by dividing the convection time step, Δt, into an integer number of smaller collision time steps, Δt_c, such that condition (27) is satisfied.

TR schemes do not have such restriction, and therefore can generate more efficient codes. An application to some space non homogeneous problems will be seen later.

5.2 Implementation of boundary conditions

In this section we shall discuss how to deal with boundaries. Let us suppose, for simplicity, that our problem in one dimensional in space. Let $\Omega = [0, L]$ be the computational domain in space. We shall distinguish between different kinds of boundary conditions. Boundary conditions are applyed during the convection step, after the candidate positions $\tilde{x}_i$ have been determined.

Periodic BC In the simple case of periodic boundary conditions, one has to check whether the candidate position is still in Ω, and if not, add or subtract the period,

for $i = 1$ to N
  if $0 < \tilde{x}_i < L$
    $x_i^{n+1} = \tilde{x}_i, \quad v_i^{n+1} = \tilde{v}_i$
  end if
  if $\tilde{x}_i < 0$
    $x_i^{n+1} = \tilde{x}_i + L, \quad v_i^{n+1} = \tilde{v}_i$
  end if
  if $\tilde{x}_i > L$
    $x_i^{n+1} = \tilde{x}_i - L, \quad v_i^{n+1} = \tilde{v}_i$
  end if
end for

Reflecting boundary conditions In this case the particle is simply reflected by the wall if the candidate space coordinate is out of the domain. This case corresponds to a reflection scattering kernel given by $R(v' \to v) = \delta(v' - v + 2\mathbf{n}(\mathbf{n} \cdot v))$. Algorithmically, this can be realized as follows

for $i = 1$ to N
  if $0 < \tilde{x}_i < L$
    $x_i^{n+1} = \tilde{x}_i, \quad v_i^{n+1} = \tilde{v}_i$

end if
if $\tilde{x}_i < 0$ or $\tilde{x}_i > L$
$x_i^{n+1} = b - |\tilde{x}_i - b| \mathrm{sgn} \tilde{v}_{ix}, \quad v_{ix}^{n+1} = -\tilde{v}_{ix}, \quad v_{iy}^{n+1} = \tilde{v}_{iy}, \quad v_{iz}^{n+1} = \tilde{v}_{iz}$
end for

where b denotes either 0 or L. The generalization of the above condition to more dimensions is straightforward.

Maxwell boundary conditions In this case the particle is either specularly reflected, or absorbed by the boundary, and re-emitted with a Maxwellian distribution corresponding to the thermal equilibrium with the wall. The reflection scattering kernel is given by

$$R(v' \to v) = \alpha M_w(v) |v \cdot \mathbf{n}| + (1 - \alpha) \delta(v' - v + 2\mathbf{n}(\mathbf{n} \cdot v)).$$

Here α is the co called *accomodation coefficient*, and $M_w(v)$ is the Maxwellian distribution corresponding to a gas in thermodynamic equilibrium with the wall. Algorithmically, the boundary condition is implemented in this way. Suppose, for example, that $\tilde{x}_i < 0$. Then, with probability $1 - \alpha$ the particle will undergo a specular reflection. With probability α the particle is sampled from the half Maxwellian $M_w(v)$, $v_x > 0$. In this case the new positions and velocities are computed according to

compute $\tilde{\Delta t} = |\tilde{x_i}/\tilde{v}_x|$
sample $\tilde{v}$ from the Maxwellian $M_w(v)$, and assign $v_{ix}^{n+1} = |\tilde{v}_{ix}|$, $v_{iy}^{n+1} = \tilde{v}_{iy}$, $v_{iz}^{n+1} = \tilde{v}_{iz}$
compute $x_i^{n+1} = v_{ix}^{n+1} \tilde{\Delta t}$

An analogue procedure can be used at the wall located at $x = L$.

5.3 Shock wave profiles

The test problem deals with the numerical solution of the space non homogeneous Boltzmann equation for hard sphere molecules (VHS, for $\alpha = 1$) with $C_\alpha = 1$. We present some numerical results for one-dimensional stationary shock profiles. In particular we have computed the structure of the shock for different Knudsen numbers, from the rarefied regime up to the fluid limit. In all our numerical tests the gas is initially at the upstream equilibrium state in the left half-space and in the downstream equilibrium state in the right-half space. The upstream state is determined from the downstream state using the Rankine-Hugoniot relations [20].

In the present calculations, the downstream state is characterized by

$$\rho = 1.0, \quad T = 1.0, \quad M = 3.0,$$

where M is the Mach number of the shock. The downstream mean velocity is then given by

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

with $\gamma = 2$ since we have considered a 2D monatomic gas in velocity space.

The infinite physical space is truncated to the finite region $[-7.5, 7.5]$. A suitable stabilizing technique [2] has been used to keep the number of particles constant during the time evolution. We report the result obtained with the different schemes using 50 space cells and 500 particles in each downstream cell. Since we are computing a stationary solution after a fixed initial time, we can strongly improve the accuracy of the Monte Carlo solution by averaging in time the solution itself. To this aim we started accumulating statistics at time $t = 5$ and we average up to $t = 20$ which corresponds approximatively to 1500 time steps. The reference solution in the rarefied and intermediate regime is obtained using the DSMC method with 200 space cells and 500 particles in each downstream cell and averaging over approximatively 8000 time steps. All the results are reported on the space interval $[-4.5, 4.5]$.

In Fig. 8 we plot the result obtained in the rarefied regime ($\epsilon = 1$) using the different Monte Carlo methods. The computational cost of the methods is comparable since here we are far from the fluid limit. As expected, the results show that essentially the TRMC methods and the DSMC method are equivalent and provide a good description of the rarefied shock.

Next we consider the intermediate regime ($\epsilon = 0.1$). Two different time steps have been used for DSMC, a convection time step Δt and a collision time step $\Delta t_{\rm coll} = \Delta t/10$. The collision time step for TRMC is equal to the convection time step. The result show that, despite the larger time step, the accuracy of the solution obtained with TRMC schemes is essentially the same of DSMC (see Fig. 9).

Finally we give the result of the computations close to the Euler limit ($\epsilon = 10^{-6}$). The profiles obtained with the TRMC methods are reported in Fig. 10. Due to the small Knudsen number the DSMC method is unusable in practice in this test. Note that the TRMC methods are all equivalent to Pullin's method. Thus we have plotted only one approximate solution.

References

[1] H. Babovsky *On a simulation scheme for the Boltzmann equation*, Math. Methods Appl. Sci., **8** (1986), pp. 223-233.

[2] G.A. Bird. *Molecular Gas Dynamics*. Oxford University Press, London, 1976.

[3] J.F Bourgat, P. LeTallec, B. Perthame, and Y. Qiu. Coupling Boltzmann and Euler equations without overlapping. In *Domain Decomposition*. AMS, 1992.

[4] C. Cercignani The Boltzmann equation and its applications, Springer-Verlag, Berlin, (1988).

[5] C. Cercignani, R. Illenr, M. Pulvirenti, The mathematical theory of dilute gases, Springer-Verlag, New York, (1995).

[6] E. Gabetta, L. Pareschi, and G. Toscani. Relaxation schemes for nonlinear kinetic equations. *SIAM Journal on Numerical Analysis*, 34:2168–2194, 1997.

[7] E. Hairer and G. Wanner. *Solving ordinary differential equations, Vol.2 Stiff and differential-algebraic problems*. Springer-Verlag, New York, 1987.

[8] R.W. Hockney and J.W. Eastwood. *Computer simulation using particles*. Mc-Graw Hill International Book Co., 1981.

[9] M. Kac, *Probability and related topics in physical sciences*, Lectures in Applied Mathematics, (1957), Interscience Publishers London-New York.

[10] P. L. Lions, B. Mercier *Splitting algorithms for the sum of two maximal monotone operators*, SIAM J. Numer. Anal., 31, (1979), 964-979.

[11] W. Morokoff and R.E. Caflisch *Quasi-Monte Carlo Integration*, J. Comp. Phys. , (1995), 122, pp. 218-230.

[12] K. Nanbu *Direct simulation scheme derived from the Boltzmann equation. I. Monocomponent Gases*, J. Phys. Soc. Japan, 52, (1983), pp. 2042-2049.

[13] T. Ohwada. Higher order approximation methods for the boltzmann equation. *Journal of Computational Physics*, 139:1–14, 1998.

[14] L. Pareschi and R.E. Caflisch. Implicit Monte Carlo methods for rarefied gas dynamics I: the space homogeneous case. *Journal of Computational Physics*, to appear.

[15] L.Pareschi, G.Russo *Asymptotic preserving Monte Carlo methods for the Boltzmann equation*, TTSP to appear.

[16] L.Pareschi, G.Russo *Time relaxed Monte Carlo methods for the Boltzmann equation*, preprint 1999.

[17] D.I. Pullin. Direct simulation methods for compressible inviscid ideal gas flow. *Journal of Computational Physics*, 1978.

[18] D.I. Pullin. Generation of normal variates with given sample. *J. Statist. Comput. Simul.*, 9:303–309, 1979.

[19] Z.Tan, P. Varghese *The $\Delta - \epsilon$ method for the Boltzmann equation*, J. Comp. Phys., 110, (1994) 327-340.

[20] G.B.Whitham *Linear and nonlinear waves*, Wiley-Interscience, (1974).

[21] E. Wild. On Boltzmann's equation in the kinetic theory of gases. *Proc. Camb. Phil. Soc.*, 47:602–609, 1951.

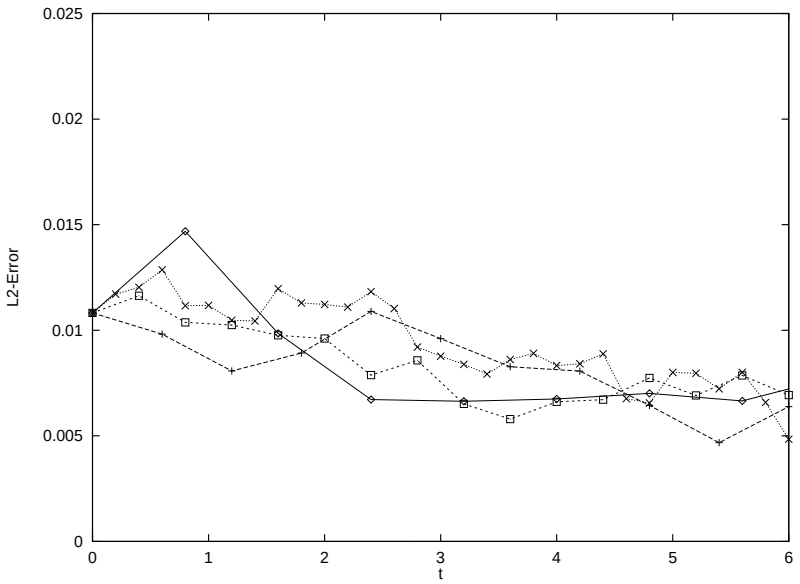

Figure 5: Kac equation : L^2 error vs time with $\Delta t = 0.2$ for DSMC ($\times$), $\Delta t = 0.4$ for TRMC1 ($\square$), $\Delta t = 0.6$ for TRMC2 ($+$) and $\Delta t = 0.8$ for TRMCH ($\diamond$).

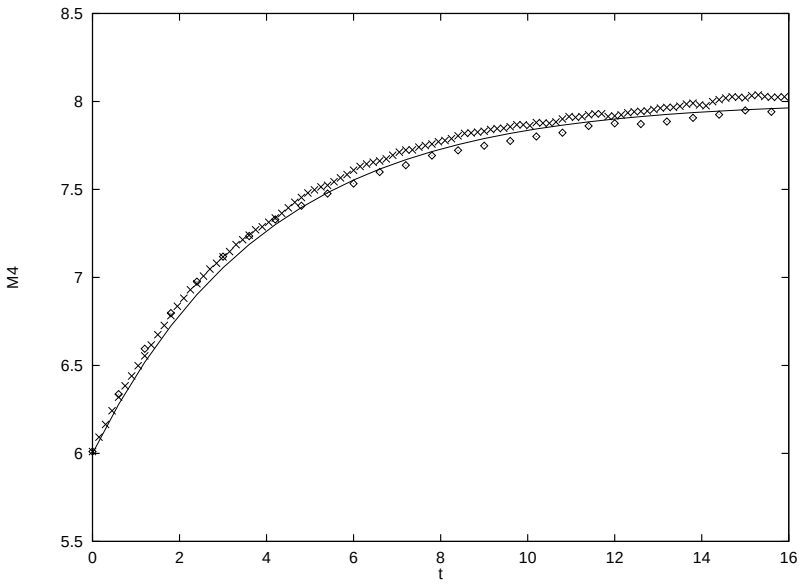

Figure 6: Maxwell molecules : Fourth order moment vs time with $\Delta t = 0.15$ for DSMC ($\times$) and $\Delta t = 0.6$ for TRMCH ($\diamond$). The line represents the exact solution.

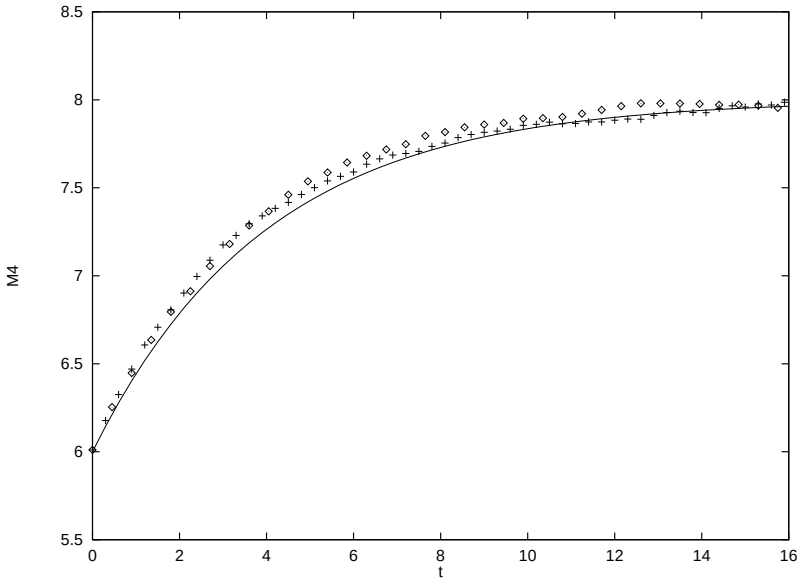

Figure 7: Maxwell molecules : Fourth order moment vs time with $\Delta t = 0.3$ for TRMC1 ($\times$) and $\Delta t = 0.45$ for TRMC2 ($\diamond$). The line represents the exact solution.

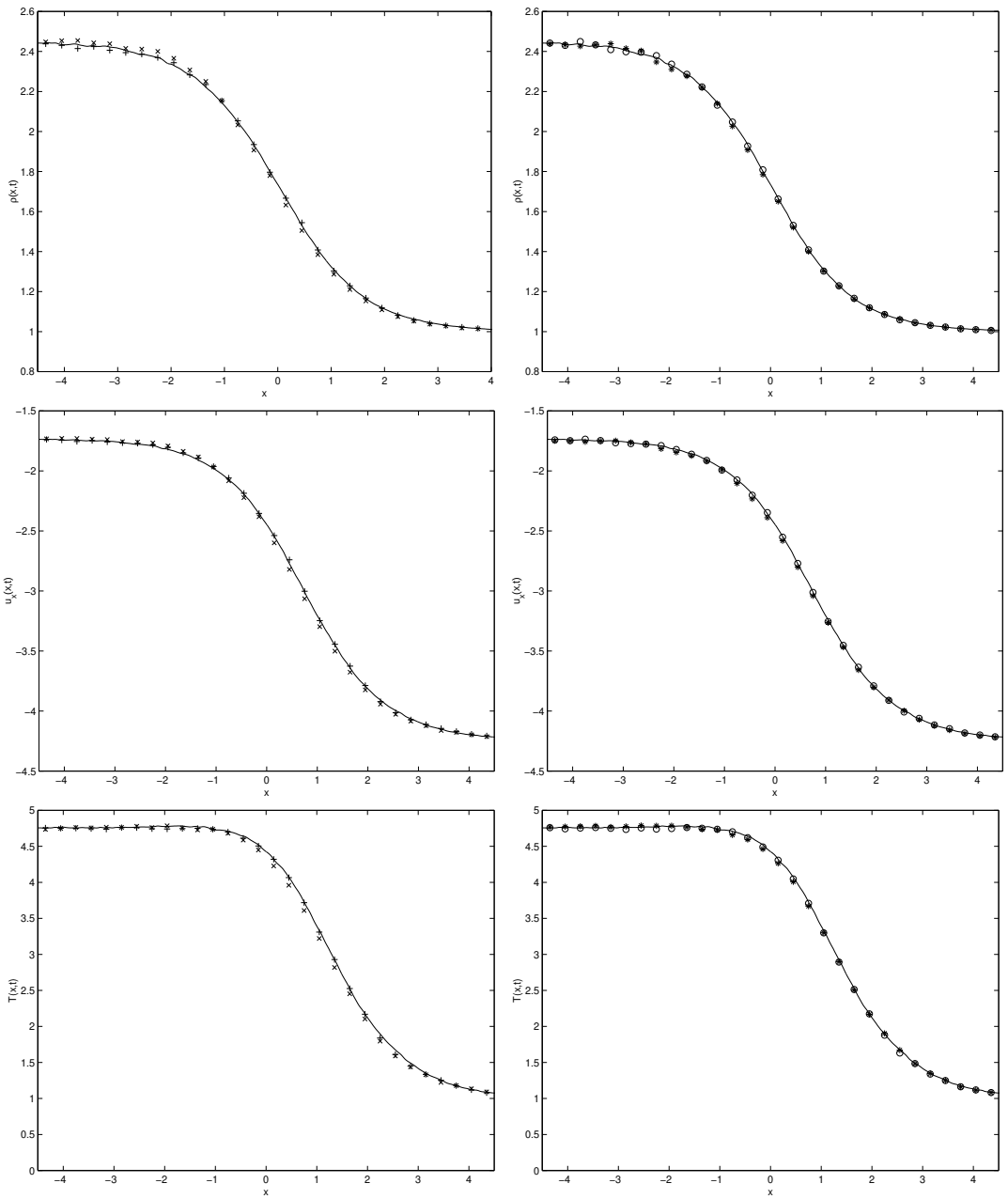

Figure 8: Shock wave profiles (rarefied regime): DSMC(+) and first order TRMC ($\times$) (left column), second order ($*$) and third order ($\circ$) TRMC (right column) for $\epsilon = 1.0$ and $\Delta t = 0.025$. From top to bottom: ρ, u, T. The line is the reference solution.

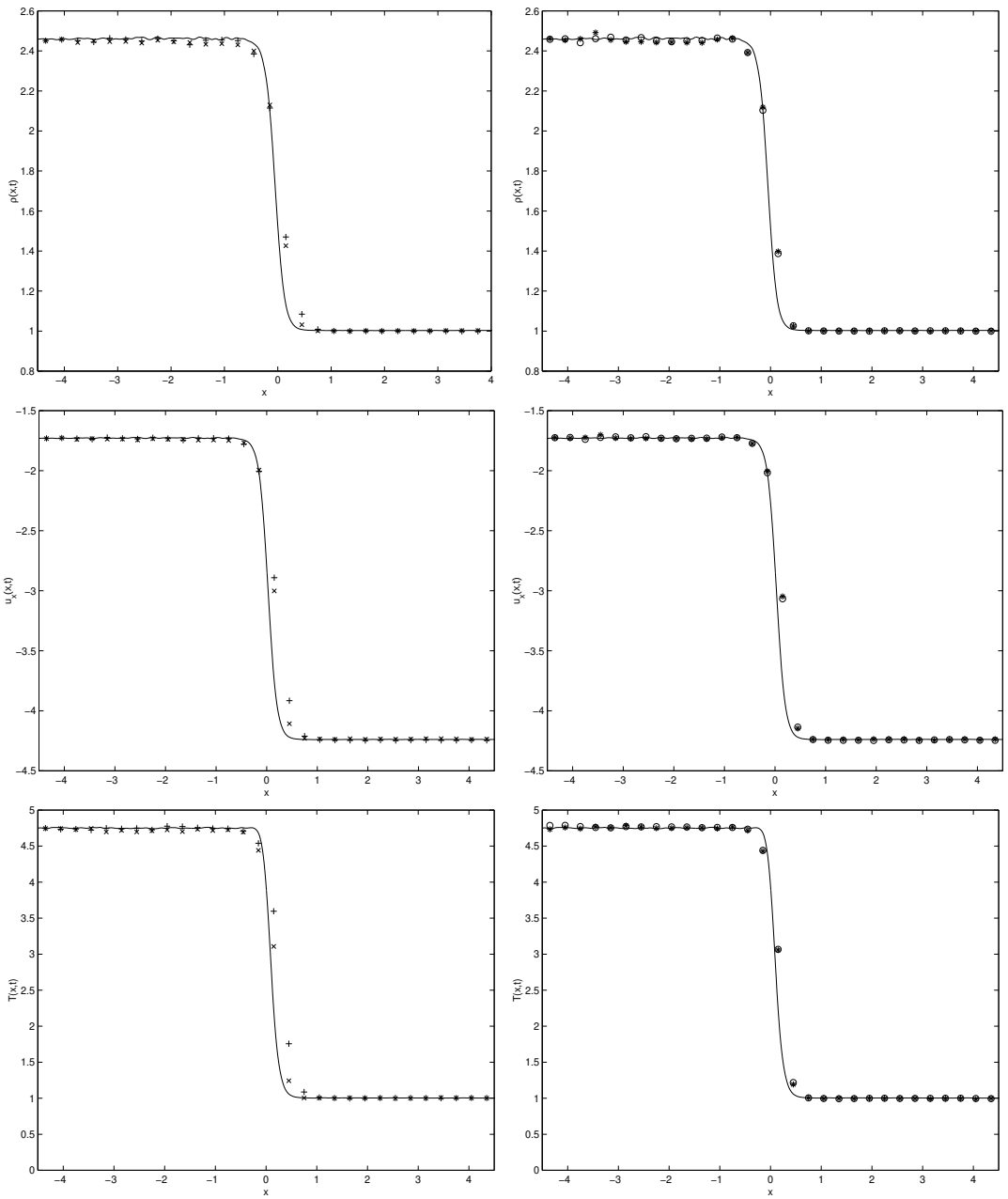

Figure 9: Shock wave profiles (intermediate regime): DSMC(+) and first order TRMC ($\times$) (left column), second order ($*$) and third order ($\circ$) TRMC (right column) for $\epsilon = 1.0$ and $\Delta t = 0.0025$ for DSMC, $\Delta t = 0.025$ for TRMC. From top to bottom: ρ, u, T. The line is the reference solution.

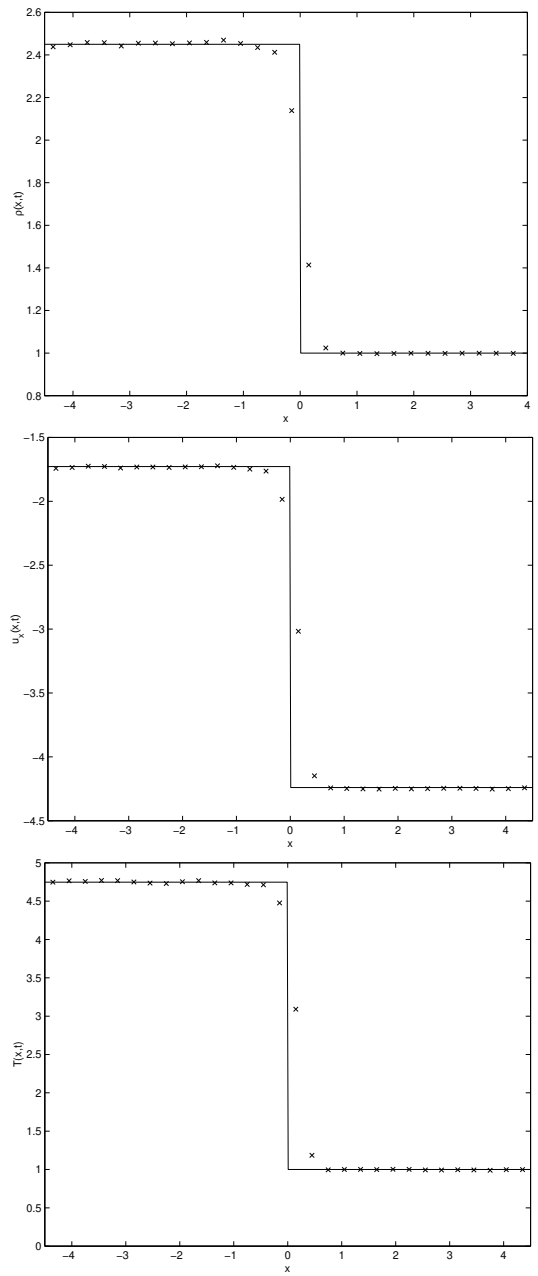

Figure 10: Shock wave profiles (fluid regime): first order TRMC ($\times$) for $\epsilon = 10^{-8}$ and $\Delta t = 0.025$. From top to bottom: ρ, u, T.