# THE BOLTZMANN EQUATION ON A TWO-DIMENSIONAL LATTICE THEORETICAL AND NUMERICAL RESULTS

LAURA FAINSILBER[1], PÄR KURLBERG[2], AND BERNT WENNBERG[1]

ABSTRACT. The construction of discrete velocity models or numerical methods for the Boltzmann equation, may lead to the necessity of computing the collision operator as a sum over lattice points. The collision operator involves an integral over a sphere, which corresponds to the conservation of energy and momentum. In dimension two there are difficulties even in proving the convergence of such an approximation since many circles contain very few lattice points, and some circles contain many badly distributed lattice points. This paper contains a brief description of the proof that was recently presented elsewhere ([L. Fainsilber, P. Kurlberg, B. Wennberg, preprint 2004]). It also presents the results of numerical experiments.

## 1. INTRODUCTION

The Boltzmann equation is

$$\partial_t f(x,v,t) + v \cdot \nabla_x f(x,v,t) \;=\; Q(f,f)(x,v,t)\,. \tag{1}$$

We consider this equation in two spatial dimensions, so $x \in \mathbb{R}^2$, and $v \in \mathbb{R}^2$. The collision operator in the right hand side acts only in velocity space, and is defined as

$$Q(f,f)(v) \;=\; \int_{\mathbb{R}^2}\int_{S^1} \left(f(v')f(v'_*) - f(v)f(v_*)\right) q(|w|,\cos\theta)\,\frac{d\theta}{2\pi}dv_*\,. \tag{2}$$

The velocities "before and after a collision" are related by

$$\begin{aligned} v' &= \tfrac{1}{2}(v+v_*) + |w|\,u \\ v'_* &= \tfrac{1}{2}(v+v_*) - |w|\,u\,. \end{aligned} \tag{3}$$

Here $w = (v_* - v)/2$, and the unit vector $u \in \mathbb{R}^2$ is defined as a rotation by the angle $\theta$ of $w/|w|$:

$$u \;=\; R_\theta \frac{w}{|w|}\,.$$

The two velocities $v$ and $v_*$ are antipodal points on a well defined circle, and (3) implies that after a collision, the two new velocities are different antipodal points on the same circle. We parametrise this circle by $\theta$, and $d\theta/2\pi$ is simply the unit measure. Finally, $q(|w|,\cos\theta)$ is the differential crossection.

In a discrete velocity model (DVM), the velocities are concentrated on a (usually finite) set of points $v_j \in \mathbb{R}^d$ in the velocity space:

$$f(x,v,t) \;=\; \sum_j f_j(x,t)\delta_{v=v_j}\,.$$

The Boltzmann equation (1) is then changed into a nonlinear system of ordinary differential equations, or, when also the spatial dimension is taken into account, a system of conservation laws:

$$\partial_t f_j + v_j \cdot \nabla_x f_j = \sum_{k,k',j'} \Gamma_{j,k}^{j',k'} \left(f_{j'}f_{k'} - f_j f_k\right)\,. \tag{4}$$

The constants $\Gamma_{j,k}^{j',k'} \geq 0$ must be chosen so that (4) makes sense from a physical point of view. In particular we require that $(v_j, v_k)$ and $(v_{j'}, v_{k'})$ are two pairs of antipodal points on the same circle, just as for the usual Boltzmann equation.

The first example of a discrete velocity model is that of Carleman ([5]), which has two velocities in $\mathbb{R}$, but there are many other models with different number of velocities.

There are at least two reasons for studying such models. First, from certain points of view, they are mathematically more tractable than the continuous Boltzmann equation (though certainly not in all respects), and results pertaining to the discrete models could also say something about the full equation. Also they provide a means of doing numerical calculations for gases far away from equilibrium.

When used for real gas simulations, it is essential that the model is physically realistic (i.e., that satisfy the right conservation laws and an entropy principle), and this problem has recently been addressed e.g. by [4, 22, 23].

The family of models considered here can be seen as coming from a rather straightforward discretization of the collision integral (2), where the integrand is evaluated only on lattice points, $v \in h\mathbb{Z}$. Integrating over $w = (v_* - v)/2$ rather than over $v_*$ gives

$$\begin{aligned} v' &= v + w + |w|u\,, \\ v'_* &= v + w - |w|u\,. \end{aligned}$$

Also, $v_* = v + 2w$, and writing

$$g_v(w,u) = \left(f(v')f(v'_*) - f(v)f(v_*)\right) q(|w|, \cos\theta)\,, \tag{5}$$

we find

$$Q(f,f)(v) = 4\int_{\mathbb{R}^2} \left( \int_{S^1} g_v(w,u)\, d\frac{\theta}{2\pi} \right) dw. \tag{6}$$

If $g$ is sufficiently regular (continuous), and decays sufficiently rapidly for large $w$, then the Riemann sum for the outer integral converges:

$$\begin{aligned} &(2h)^2 \sum_{\zeta\in\mathbb{Z}^2} \int_{S^1} g_v(h\zeta, u)\, d\theta \\ &\qquad \longrightarrow \quad 4\int_{\mathbb{R}^2} \left( \int_{S^1} g_v(w,u)\, \frac{d\theta}{2\pi} \right) dw \end{aligned} \tag{7}$$

when $h \to 0$. In order to construct a consistent DVM, it is then sufficient to evaluate the inner integral in terms of the values of $g$ on the lattice points $h\mathbb{Z}^2$, in such a way that the result converges to $\int_{S^1} g(w,u)\, \frac{d\theta}{2\pi}$.

While with the formula (3), the collision integral should be taken over all $u \in S^1$, we have here only access to those $u$ for which $v'$ and $v'_*$ belong to $h\mathbb{Z}^2$. But this is automatically achieved if $\zeta \in \mathbb{Z}^2$, and if $u = \zeta'/|\zeta'|$, where $\zeta' \in \mathbb{Z}^2$ and $|\zeta'| = |\zeta|$; then for all $v \in h\mathbb{Z}^2$,

$$v + h\zeta \pm h|\zeta|u \in h\mathbb{Z}^2\,.$$

However, note that with this construction, the center of the sphere is restricted to lie on a lattice point, and so it excludes cases like $v = (0,0)$, $v_* = (h,h)$.

Giving all points on the circle equal weight, one arrives at the following expression for the full collision operator, valid for all $v \in h\mathbb{Z}^2$:

$$Q^h(f,f)(v) = \\ (2h)^2 \sum_{\zeta\in\mathbb{Z}^2} \frac{1}{r(|\zeta|^2)} \sum_{\substack{\zeta'\in\mathbb{Z}^2 \\ |\zeta'|=|\zeta|}} \left(f(v')f(v'_*) - f(v)f(v_*)\right) q(|h\zeta|, \cos\theta)\,. \tag{8}$$

The function $r(n)$ denotes the number of points with integer coordinates on a sphere in $\mathbb{R}^2$ with center at the origin and radius $\sqrt{n}$, i.e. the number of integer solutions to $x_1^2 + x_2^2 = n$. The angle $\theta$ is the angle between $\zeta$ and $\zeta'$. To obtain the discrete velocity model (4) one can then take $f_\xi(x,t) = f(h\xi, x, t)$. This would then be a model with countably many velocities, but it is natural to restrict velocities to belong to a bounded subset of $\mathbb{Z}^2$.

We are interested in proving that

$$Q^h(f,f) \to Q(f,f), \tag{9}$$

when $h \to 0$, at least for sufficiently regular functions $f$. If this convergence holds, we say that the model is *consistent*, which together with *stability* is a main ingredient when proving that a numerical method converges.

Indeed, (9) holds. For dimensions strictly larger than 2, this result was established by Palczewski, Schneider and Bobylev([3]). The same result was proven for $d = 2$ in [12]. In this paper we give a short description of the proof, and present some numerical calculations, which have not been presented elsewhere.

Although we do not pretend to construct valid and useful methods for solving the Boltzmann equation, it is interesting to test whether the model is admissible from a physical point of view. For the particular model given in equation (8), we know that it is admissible, because it is an example of a general method of constructing discrete velocity models that is presented in [4]. We discuss that general method in Section 4, and present some results from a computer implementation of the method.

We also show the results of some calculations for a spatially homogeneous relaxation to equilibrium.

Discretizations of the Boltzmann equation have been discussed by several authors. The most relevant papers in connection with the present one are [3], [2], and also [19]. A different method based on Farey series was presented in [21]. The collision operator in the two-dimensional Boltzmann equation is a three-fold integral, which is evaluated as an iterated integral. A different discretization based on the so-called Carleman representation of the collision integral was presented in [15].

## 2. Main result and ideas of the proof

The purpose of this section is to properly state the convergence result (7), and to discuss its proof. All details of the proof can be found in [12].

In addition to the notation in Section 1, we write

$$G_v(w) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} g_v(w,\theta)\,d\theta\,,$$

in the continuous case, and for the discrete case (then we assume, of course, that $v \in h\mathbb{Z}^2$)

$$G_v^h(h\zeta) \;=\; \frac{1}{r(|\zeta|^2)}\sum_{\substack{\zeta'\in\mathbb{Z}^d\\ |\zeta'|=|\zeta|}} g_v(h\zeta,\theta)\,,$$

where $\theta$ is the angle between $\zeta'$ and $\zeta$. As before, $r(|\zeta|^2)$ denotes the number of integer points on a sphere with radius $|\zeta|$.

We also write

$$Z_{h,R} = \{z \in \mathbb{Z}^2 \;\text{ s.t. }\; |z| \le R/h\} \tag{10}$$

for some $R > 0$ (this is the most straight forward way of restricting to a finite set of velocities, but other choices might be more efficient, as we shall see later).

The convergence result can now be expressed as

$$Q(f,f)(v) - (2h)^2 \sum_{\zeta\in Z_{h,R}} G_v^h(h\zeta) \to 0 \tag{11}$$

when $h \to 0$.

**Theorem 1.** *Suppose that $g_v(w,\theta)$ in (5) satisfies*

1. *$g_v(w,\theta)$ is a $C^1$-function w.r.t. $w$*
2. *$g_v(w,\theta)$ is a $C^2$-function w.r.t. $\theta$*
3. *$\|g_v(\cdot,\theta)(1+|\cdot|^2)\|_{L^1(dw)} \le C$*

*(This holds e.g. if the function $f$ and the crossection $q$ are $C^2$.) For given $R > 0$ and $h > 0$, let $Z_{h,R}$ be as in (10). Then given $\varepsilon > 0$ there are reals $R > 0$ and $h > 0$ such that*

$$\left| Q(f,f)(v) - (2h)^2 \sum_{\zeta\in Z_{h,R}} G_v^h(h\zeta) \right| \le \varepsilon\,.$$

*For a given $\varepsilon$, one can take $h$*

$$h = o\left(\exp(-2\,(\log\varepsilon)^2\,\varepsilon^{-2/(1-\frac{2}{\pi})})\right)\,, \tag{12}$$

*which corresponds to a rate of convergence no better than $O((\log(1/h))^{-p})$, where $p < (1-2/\pi)/2$.*

*Proof.* We still consider $Q(f,f)$ as an iterated integral, and write (for $v \in h\mathbb{Z}^2$)

$$\begin{aligned} Q(f,f)(v) - (2h)^2 \sum_{\zeta \in Z_{h,R}} G_v^h(h\zeta) \\ = \int_{\mathbb{R}^2} G_v(w)\, dw - (2h)^2 \sum_{\zeta \in Z_{h,R}} G_v(h\zeta) \\ +(2h)^2 \sum_{\zeta \in Z_{h,R}} \left( G_v(h\zeta) - G_v^h(h\zeta) \right) . \end{aligned} \tag{13}$$

The difference between the integral in the right hand side and the first sum can be estimated easily by truncating the integral for large velocities and using that the sum is a Riemann sum for the remaining part of the integral. So the difference is bounded by

$$\frac{C_1}{R^2} + C_2 R^2 h \,,$$

where the constants depend on the $C^1$-bounds of $g$.

Next we turn to the difference $G_v(h\zeta) - G_v^h(h\zeta)$, i.e. of

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g_v(h\zeta, \theta)\, d\theta - \frac{1}{r(|\zeta|^2)} \sum_{\substack{\zeta' \in \mathbb{Z}^2 \\ |\zeta'|=|\zeta|}} g_v(h\zeta, \theta) \,, \tag{14}$$

(recall that in the second term, $\theta$ is the angle between $\zeta'$ and $\zeta$). We first write the periodic function $g_v(h\zeta, \theta)$ as a Fourier series,

$$g_v(h\zeta, \theta) = \sum_{k \in \mathbb{Z}} \hat{g}_v(\zeta, k) e^{ik\theta} \,,$$

where

$$\hat{g}_v(\zeta, k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_v(h\zeta, \theta) e^{-ik\theta}\, d\theta \,.$$

The assumptions on $g$ imply the existence of a constant $C_3$ so that

$$|\hat{g}_v(\zeta, k)| \le \frac{C_3}{1+k^2} \,. \tag{15}$$

Then (14) becomes

$$\hat{g}_v(\zeta, 0) - \frac{1}{r(|\zeta|^2)} \sum_{\substack{\zeta' \in \mathbb{Z}^2 \\ |\zeta'|=|\zeta|}} \hat{g}_v(\zeta, 0) + \frac{1}{r(|\zeta|^2)} \sum_{\substack{\zeta' \in \mathbb{Z}^2 \\ |\zeta'|=|\zeta|}} \sum_{k \neq 0} \hat{g}_v(\zeta, k) e^{ik\theta} \,,$$

where the first terms cancel out, and only last sum remains. We next split that sum into a part with $|k| \le M$, and a remainder. The estimate (15) implies that the remainder is smaller than

$$\frac{R^2 C_4}{M} \,. \tag{16}$$

The terms that remain after these truncations add up to the main contribution. This is the the most difficult part to estimate. Using (15) again, we find a bound of the form

$$\left| \sum_{0<|k|<M} \frac{C_3}{1+k^2} \frac{1}{r(|\zeta|^2)} \sum_{\substack{\zeta' \in \mathbb{Z}^2 \\ |\zeta'|=|\zeta|}} e^{ik\theta} \right| \leq \max_{0<|k|<M} \left| \frac{S(|\zeta|^2,k)}{r(|\zeta|^2)} \right| \cdot \sum_{0<|k|<M} \frac{C_3}{1+k^2}$$

(17)

Here we have introduced the notation

$$S(n,k) = \sum_{u \in \mathbb{Z}^2 : |u|^2 = n} e^{ik\theta_u} \tag{18}$$

where $\theta_u$ is defined by $u = |u| \cdot (\sin\theta_u, \cos\theta_u)$. From this it is straightforward to derive (we refer to [12] for the details)

$$\begin{aligned} &|Q(f,f)(v) - Q^h(f^h,f^h)(v)| \\ &\leq \frac{C_1}{R^2} + C_2 R^2 h + \frac{R^2 C_4}{M} + C_3 (2h)^2 \max_{0<|k|<M} \sum_{n<(R/h)^2} \left| S(n,k) \right|, \end{aligned} \tag{19}$$

Proposition 3, which is stated in the next section, gives an estimate of exponential sums of this kind, and using it, we obtain

$$\sum_{n<(R/h)^2} \left| S(n,k) \right| \leq C_5 \left( \frac{R}{h} \right)^2 \exp\left( -\big(1 - \frac{2}{\pi}\big) \log \frac{\log\left((R/h)^2\right)}{\left(\log M\right)^2} \right),$$

where $C_5$ is a positive constant. We conclude the proof by choosing

(1) $R = \sqrt{4C_1/\varepsilon}$,
(2) $h < \varepsilon/(4R^2C_2) = \varepsilon^2/(4C_1C_2)$,
(3) $M = 4R^2C_4/\varepsilon = 64C_1C_4/\varepsilon^2$.

With these choices of $R$ and $M$, the last term in (19) can then be bounded by

$$4C_3C_5 \frac{4C_1}{\varepsilon} \exp\left( -\big(1 - \frac{2}{\pi}\big) \log \frac{\log(4C_1/(\varepsilon h^2))}{\left(\log(64C_1C_4/\varepsilon^2)\right)^2} \right), \tag{20}$$

which converges to zero when $h \to 0$, and so there is an $h$ so small that also the last term in (19) is smaller than $\varepsilon/4$. Solving for $h$ in (20) gives (12). □

## 3. Number theoretic background

In order to explain the origin of Proposition 3, and also to explain the numerical algorithm used to produce the results in Section 4, we need to introduce the concept of Gaussian integers, and give some related results.

To prove that the inner sum of (8) converges to the correct limit when $h \to 0$, one is lead to study the set

$$\{\zeta/|\zeta| : \zeta \in \mathbb{Z}^2, |\zeta|^2 = n\} \tag{21}$$

and to show that there are many points in this set, and also that these points are well distributed on $S^1$ when $n$ is large. This is not true in general. For example, when $n$ is a power of 2, there are exactly four points in the set. But even circles which do have a large number of points may behave poorly, as the following theorem shows:

**Theorem 2.** (Cilleruelo [6]) *For any $\epsilon > 0$ and for any integer $k$, there exists a circle $x^2 + y^2 = n$ with more than $k$ lattice points such that all the lattice points are on the arcs $\sqrt{n}e^{(\pi/2)(t+\theta)i}$ with $|\theta| < \epsilon$, $t \in \{0, 1, 2, 3\}$.*

On the other hand, we may use some other techniques from analytic number theory to show that lattice points on circles are equidistributed *on average*, and this is good enough for our purpose. To do this it is convenient to rephrase the problem in terms of the *Gaussian integers*, i.e., the ring of integers of the field $\mathbb{Q}(i)$,

$$\mathbb{Z}[i] = \{x + iy \in \mathbb{C}, (x, y) \in \mathbb{Z}^2\}.$$

The Gaussian integers behave in many ways like the usual integers, and in particular there is a unique factorization into *Gaussian primes*. We refer to [14], for basic number theoretical results.

The Gaussian primes (i.e. the elements of $\mathbb{Z}[i]$ that cannot be written as a product of Gaussian integers with smaller modulus), are of three types:

- the prime numbers $q \in \mathbb{Z}$ such that $q \equiv 3 \mod 4$ remain prime in $\mathbb{Z}[i]$ (e.g. 3, 7, 11, 19,...);
- for prime numbers $p \in \mathbb{Z}$ such that $p \equiv 1 \mod 4$, there exist $x, y \in \mathbb{Z}$ s.t. $p = x^2 + y^2$. Hence $p$ factors in $\mathbb{Z}[i]$ as a product of two Gaussian primes
$$p = (x + iy)(x - iy)$$
(e.g. 5 factors into $(2 + i)(2 - i)$ in $\mathbb{Z}[i]$)
- last, $1 + i$ is prime (note that $(1 + i)(1 - i) = 2$ and that $1 - i = -i(1 + i)$ is merely "another form of the same prime" just as 3 and $-3$ represent the same prime).

If $n$ is the sum of two squares, then it can be factored in $\mathbb{Z}[i]$:

$$n = X^2 + Y^2 = (X + iY)(X - iY).$$

If $z = x + iy$ is a prime factor of $X + iY$, then $\bar{z} = x - iy$ must be a prime factor of $X - iY$. It follows that prime factors $q \equiv 3 \mod 4$ of $n$ must appear in even powers. In addition, multiplying $n$ by an even power of a prime $q$ that is congruent

with $3 \mod 4$ changes neither the number of solutions to $n = X^2 + Y^2$ nor the distribution of arguments of the solutions.

Suppose now that $n$ contains a factor $p^\alpha$, where $p \equiv 1 \mod 4$. The number $p$ can be factored in $\mathbb{Z}[i]$ as $(x+iy)(x-iy)$, and hence the multiplicity of $x+iy$ as a factor of $n$ is $\alpha$, and the same is true for $x-iy$. It follows that the multiplicity of $x+iy$ in $X+iY$ can be any integer $j$, with $0 \leq j \leq \alpha$, and the multiplicity of $x-iy$ is then $\alpha - j$.

The same calculation can be done for powers of 2; however, the solutions given by different choices of $j$ in that case differ by a multiplication by a power of $i$, and so the power of 2 does not influence the number of solutions.

All solutions to $n = X^2 + Y^2$ can now be expressed as $X + iY = \sqrt{n}\exp(i\theta)$, where all possible values of the argument $\theta$ can be computed as sums of terms deriving from the different factors of $n$ in the following way:

(1) $X + iY$ can be multiplied by any unit, i.e. by $\pm 1$ or $\pm i$. This gives a term $k\pi/2$ in the argument, $k = 0, 1, 2, 3$.
(2) If the multiplicity of 2 in $n$ is odd, then the argument must contain $\pi/4$, the argument of $1+i$; the number of solutions does not change.
(3) For each prime factor $p \equiv 1 \mod 4$ in $n$, let $\alpha_p$ be the multiplicity of $p$ in $n$, let $p = x_p^2 + y_p^2$, and set $\theta_p = \arg(x_p + iy_p)$. For a particular choice of $j$, $0 \leq j \leq \alpha_p$, the argument added to $X+iY$ is $j\theta_p - (\alpha_p - j)\theta_p = (2j - \alpha_p)\theta_p$.

Since the choices of $k$, and of the different $j's$ are independent, the number of different solutions is $\quad 4 \prod_{p \equiv 1 \bmod 4} (\alpha_p + 1)$.

This description is constructive, and can easily be implemented as a computer program for tabulating the sets (21).

The key estimate remaining for the proof of Theorem 1 is the following estimate for averages of exponential sums:

**Proposition 3.** *If $4 \nmid k$ then $|S(m,k)| = 0$. If $4|k$ and $k \neq 0$, there exist $C$ and $b > 0$ such that*

$$\log\left(\frac{1}{X}\sum_{m \leq X} |S(m,k)|\right) \leq C - (1 - 2/\pi)\log\left(\frac{\log X}{(\log|k|)^2}\right)$$

*for $X$ sufficiently large and $\log|k| \leq b\sqrt{\log X}$.*

The proof is based on the observation that $|S(m,k)|/4$ is a *multiplicative* function, i.e. a function $f : \mathbb{N} \to \mathbb{C}$ such that $f(mn) = f(m)f(n)$ for all $m, n$ such that $(m,n) = 1$.

It turns out that the mean value of a multiplicative function, under fairly general circumstances, can be bounded in terms of an exponential of a sum over primes. The precise result that is proved and used in [12] is the following weak form of the *Halberstam-Richert inequality* (cf. [13]).

**Theorem 4.** *Let $f$ be a nonnegative multiplicative function such that*

$$\sum_{n\leq x} f(n) = O(x)\,, \tag{22}$$

*and $f(p^k) = O(k)$ for all primes $p$ and $k \geq 1$. Then there exists $C > 0$ such that*

$$\frac{1}{X}\sum_{m\leq X} f(m) \leq C\cdot \exp\left(\sum_{p\leq X}\frac{f(p)-1}{p}\right) + O(\frac{1}{\log X})$$

*for all sufficiently large $X$.*

One can check that $\frac{1}{4}|S(p,k)|$ is a multiplicative function that satisfies the conditions for Proposition 3, and so

$$\frac{1}{X}\sum_{m\leq X}|S(m,k)| \leq C\exp\left(\sum_{p\leq X}\frac{\frac{1}{4}|S(p,k)|-1}{p}\right) + O\left(\frac{1}{\log X}\right)\,.$$

It is also straightforward to check that

$$\frac{1}{4}|S(p,k)| = \begin{cases} 2|\cos(k\theta_p)| & \text{if } p\equiv 1 \mod 4,\\ 0 & \text{if } p\equiv 3 \mod 4\,,\end{cases}$$

where $\theta_p$ is the argument of the Gaussian prime $z$ such that $z\bar{z} = p$. Hence

$$\sum_{p\leq X}\frac{\frac{1}{4}|S(p,k)|-1}{p} = \sum_{\substack{p\leq X\\ p\equiv 1 \mod 4}}\frac{2|\cos(k\theta_p)|}{p} - \sum_{p\leq X}\frac{1}{p}\,.$$

This is the precise point where the angular distribution of Gaussian primes is important, and we rely on the following estimate, which is a corollary of a theorem by Kubilyus (see [18, 11])

**Theorem 5.** *If $k \in 4\mathbb{N}$ and $\log k \leq b\sqrt{\log x}$, then*

$$\sum_{\substack{p\leq x\\ p\equiv 1 \mod 4}}\frac{|\cos(k\theta_p)|}{p} \leq \frac{1}{\pi}\log\log x + (1-2/\pi)\log\log k + O(1).$$

Using this corollary, together with Merten's theorem see [14], Ch. 22.8,

$$\sum_{p\leq X}\frac{1}{p} = \log\log X + O(1)\,,$$

we find

$$\sum_{p\leq X}\frac{\frac{1}{4}|S(p,k)|-1}{p}\leq(2/\pi-1)\log\log x+2(1-2/\pi)\log\log k+O(1)\,.$$

## 4. Some numerical examples and remarks

From a numerical point of view, the discretisation discussed above would be far too costly: a discrete velocity model with $N$ velocities would at least correspond to a computational cost of $O(N)$ per time step, because one needs to compute a value for each velocity. When the collision term is computed by the sum (11), the cost is $O(N^2)$ times some logarithmic factor of $N$ (which comes from the summation over the points on the circles). And the calculation above showed that $N$ grows exponentially in terms of the accuracy, $N\sim\frac{1}{h}>>\exp(\varepsilon^{-c})$ for some positive constant $c$.

However, rather than estimating the computational cost in terms of the number of discretisation points used, it is more relevant to give the cost in terms of the desired accuracy, assuming that the discretization points are used in an optimal way. The discussion around (11) suggests that one can reduce the computational cost considerably without compromising the order of accuracy. The poor rate of convergence is due to the approximation of $G_v(w)$. Generalizing the formula (11) slightly, we can write

$$\int_{\mathbb{R}^2}G_v(w)\,dw\sim\frac{1}{\rho_h}\sum_{\zeta\in Z_h}G_v(h\zeta) \tag{23}$$

where $\rho_h$ is the local density of $Z_h$. For $Z_h=\{\zeta\in\mathbb{Z}^2\ \text{ s.t. }\ |h\zeta|\leq R\}$, one has $\rho_h=h^{-2}$.

An important reduction in computational cost could then presumably be obtained by replacing $Z_h$ by a much smaller carefully selected set in such a way that the integrals over the corresponding circles are well approximated.

In addition to the problem of keeping the overall accuracy, one would also need to address the question of spurious invariants, which we will do briefly before giving some numerical illustrations. Since our main concern in this work was to study how well the discretised collision operator agrees with the continuous one, we have not discussed the question of whether the models admits the correct number of conserved quantities, a rather delicate problem, which we will briefly discuss here.

By a collision invariant, we mean a function $\Psi(v)$ that satisfies

$$\begin{aligned}\forall(j,k,j',k')\text{ such that }\Gamma_{j,k}^{j',k'}&>0,\\ \Psi(v_j)+\Psi(v_k)&=\Psi(v_{j'})+\Psi(v_{k'})\end{aligned} \tag{24}$$

The only invariants should be the ones corresponding to the conservation of mass, momentum and energy, *i.e.*,

$$\Psi(v) = 1,\ \Psi(v) = b \cdot v\ \ (b \in \mathbb{R}^2),\ \text{and}\ \Psi(v) = |v|^2.$$

All other functions satisfying (24) are called spurious invariants.

That the present planar lattice model does not admit any spurious invariants, at least under some very modest requirements on the differential crossection follows from the fact that it can be constructed according to a general method for constructing "normal" models. The construction, which can be found in [4] is as follows:

Starting from a model which is known to possess the correct invariants, one adds one point in a suitable way. More precisely, suppose that a discrete velocity model consists of the velocities

$$\{v_1, ...., v_m\},$$

together with a set of $\Gamma_{j,k}^{j',k'}$. If a new velocity $v_{m+1}$ is added together with an augmented set of $\Gamma_{j,k}^{j',k'}$, such that for at least one choice of $j, k, j'$ (all different), $\Gamma_{j,k}^{j',m+1} > 0$, then

$$\{v_1, ...., v_m, v_{m+1}\},$$

is also an admissible model. In our situation, we see the model as a discretisation of the continuous Boltzmann equation. Under the mild assumption that the collision crossection is strictly positive, the above amounts to saying that the new velocity belongs to a circle which has at least three velocities from the original set of velocities.

With the model introduced in Section 1, only circles with centers at lattice points are considered, and then it is natural to use only lattice points $(x, y)$ such that $x + y$ is an even number. This corresponds to an integer lattice scaled by a factor $\sqrt{2}$ rotated by $\pi/2$.

Following the idea in [4] we construct a sequence of models $\{U_m\}$ inductively, and the model $U_{m+1}$ is constructed by adding all points in $\mathbb{Z}^2 \setminus U_m$, (or $Z_h \setminus U_m$) that belong to a circle which contains at least three different points from $U_m$. As the first generation in this construction we can choose an augmented Broadwell model consisting of the velocities $(\pm 1, 0)$, $(0, \pm 1)$, extended with the point $(0, 0)$, or, to satisfy the condition that the sums of the coordinates be even, $(\pm 1, \pm 1)$ and $(0, 0)$, or any suitably scaled and rotated version of this.

With this construction, one can see that in fact it is enough to use points on a very small number of circles. This would then be an example of how to reduce the computational cost, while keeping a physically correct model. As an example we consider a model allowing only circles with exactly 128 points. This model has

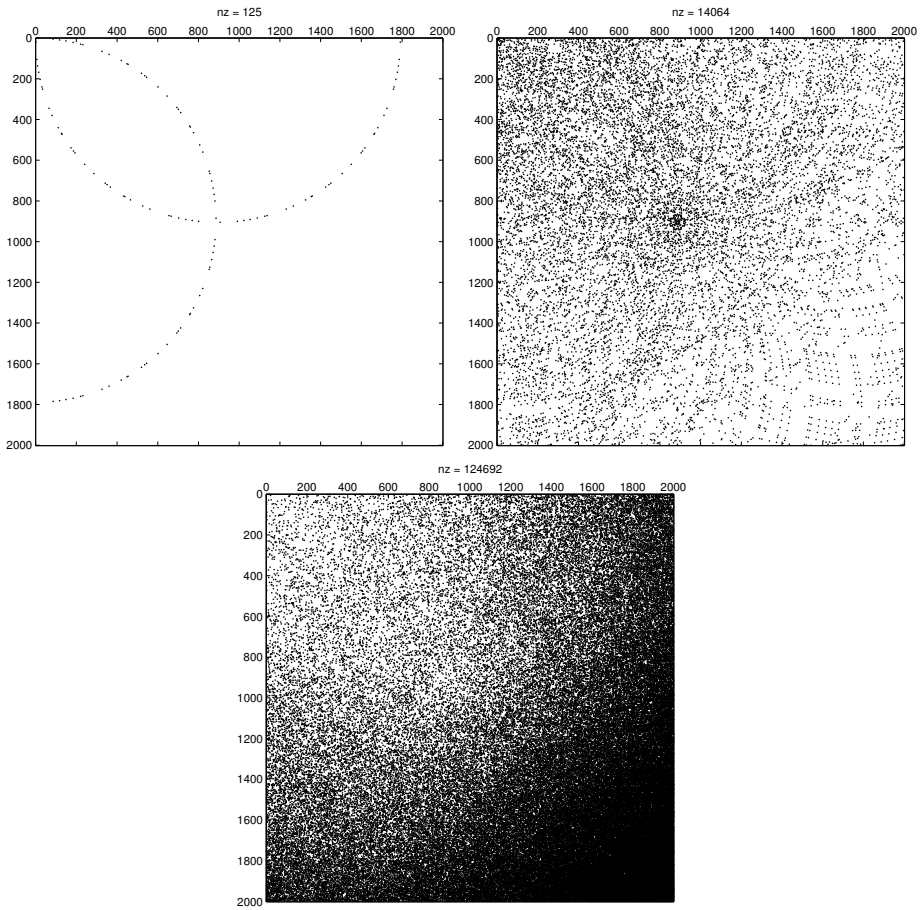


FIGURE 1. The points *added* in the second, third and fifth generation of the iteration from the implementation of the Bobylev-Cercignani method. In the fourth generation 1862118 of the in all 2 million points are added, and with the dot size used in the other plots, the square would be completely filled. The plots show one quadrant, with the origin in the upper left corner.

been chosen because within the chosen square (2000x2000 points), we don't find circles with a larger number of points. The first generation in this example is a model with the velocities $\pm(905, 885)$, and $\pm(-885, 905)$ and also $(0, 0)$. Then e.g. $(0, 0)$, $(905, 885)$, and $(885, -905)$ all are on the same circle with radius $\sqrt{801125}$, and hence one can add all the other $128 - 3$ on that same circle. In this way, because of the four fold symmetry of the problem, a first round of adding points to the Broadwell model gives $4 \times (128 - 3) = 500$ points new to the second generation. Figure 1 shows one quadrant of the second, third and fifth generation of this procedure; it is in this case enough with four iterations to obtain a model with all "even" (in the sense discussed above) points.

The conclusion of this is that allowing only circles with 128 points yields an admissible model, but it would correspond to a, from a physical point of view, very unrealistic differential crossection in the continuous case. The continuous model would have a very restricted differential crossection, and because of this the solutions could converge very slowly to equilibrium.

Finally, we give two examples of numerical calculations based on the discrete velocity model. The purpose of the simulation is to illustrate the equilibrium states. The exact time dependence of the solution is not important in this case, and a simple time stepping method has been choosen. With $f_m(\zeta) = f(h\zeta, m\Delta_t)$,

$$f_{m+1}(\zeta) = f_m(\zeta) + \Delta t \;\; Q(f_m, f_m)(\zeta),$$

and $\Delta t$ has been arbitrarily chosen to $0.1$. For reasons of computational cost, we restrict the calculation to integers $\zeta = (j, k)$ with $|j|, |k| < 100$. The iteration is computed with the formula

$$\begin{aligned} f_{m+1}(\zeta) =& f_m(\zeta) + \\ & \Delta t \sum_n \sum_{|\zeta_1|^2=n} \frac{1}{r_{\zeta+\zeta_1}(n)} \sum_{|\zeta_2|^2=n} \Big[ f_m(\zeta+\zeta_1+\zeta_2) f_m(\zeta+\zeta_1-\zeta_2) - \\ & \qquad f_m(\zeta+2\zeta_1) f_m(\zeta) \Big] \end{aligned} \tag{25}$$

This corresponds to carrying out the integration over $w$ in equation (6) with polar coordinates. Note that this summation counts all integers in the lattice exactly once, and that there is no need for a Jacobian as when changing to polar coordinates in a plane integral. The list of solutions to $|\zeta_1|^2 = n$ was tabulated in advance, using the techniques discussed in Section 3. The function $r_{\zeta+\zeta_1}(n)$ denotes the number of integer points on a circle with radius $\sqrt{n}$ as before, but counting only points inside the square domain for the simulation, and therefore it depends also on the center point $\zeta + \zeta_1$.

The graphs in Figure 2 show the result for a few of the iterates, for the case when the summation is carried out for circles with between 20 and 48 points (48 is the largest possible number of points on a circle in this case, and the restriction to circles with at least 20 was made to reduce computational cost). The corresponding values of $|\zeta_2|^2$ lie between 325 and 10000, i.e., to a differential crossection that is strictly zero in a ball $|w| < \sqrt{325}$. Because the density of circles with more than 20 points is larger in intervals of $n = |\zeta_1|^2$ with large $n$, this corresponds to hard potentials. However, the simulation is carried out without any particular differential crossection in mind.

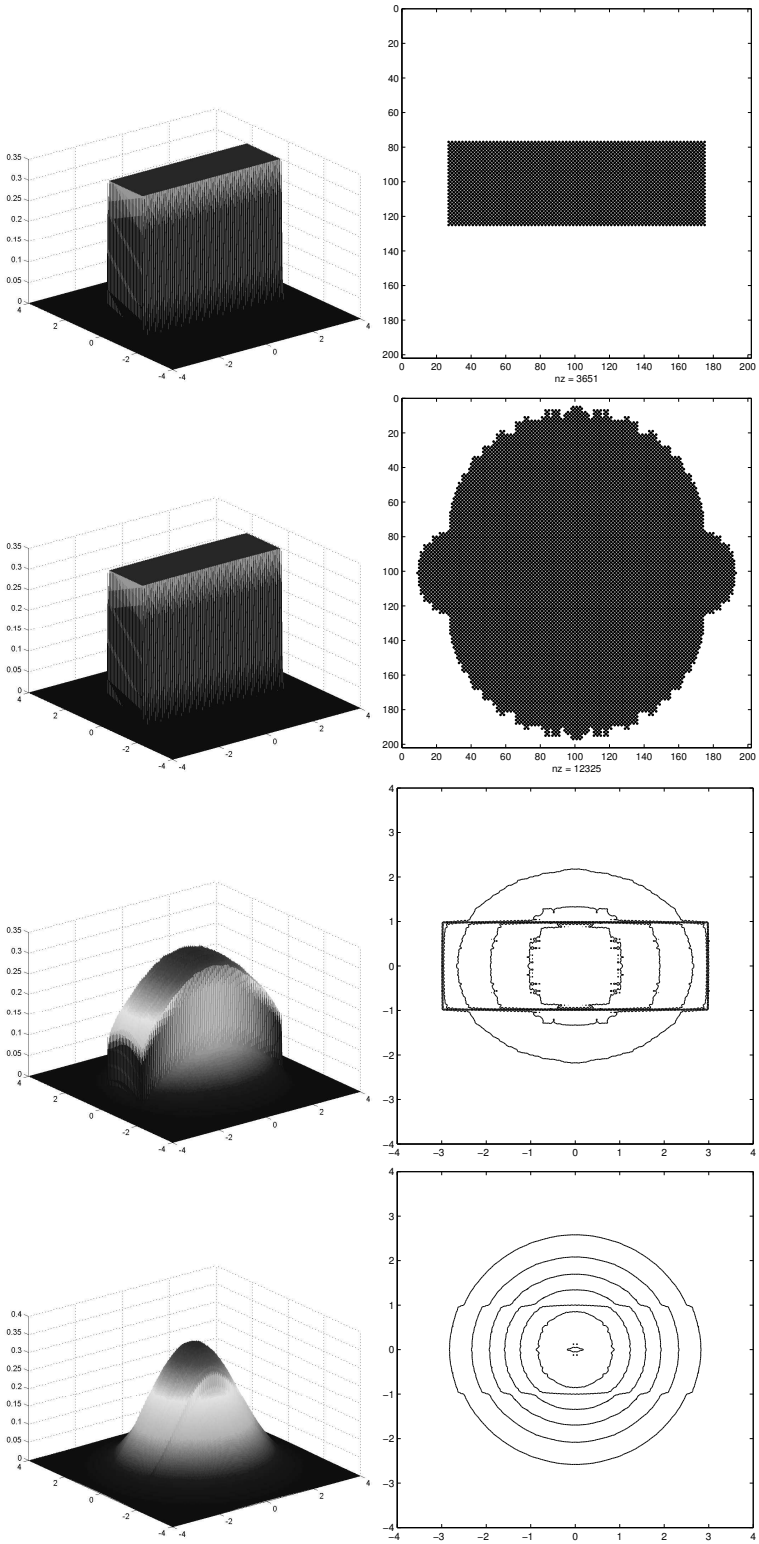


FIGURE 2. The initial data and iterations number 1, 200 and 800. The summation in equation (25) includes circles with at least 20 points. The graphs to the right show the support of the iterate or a contour plot.

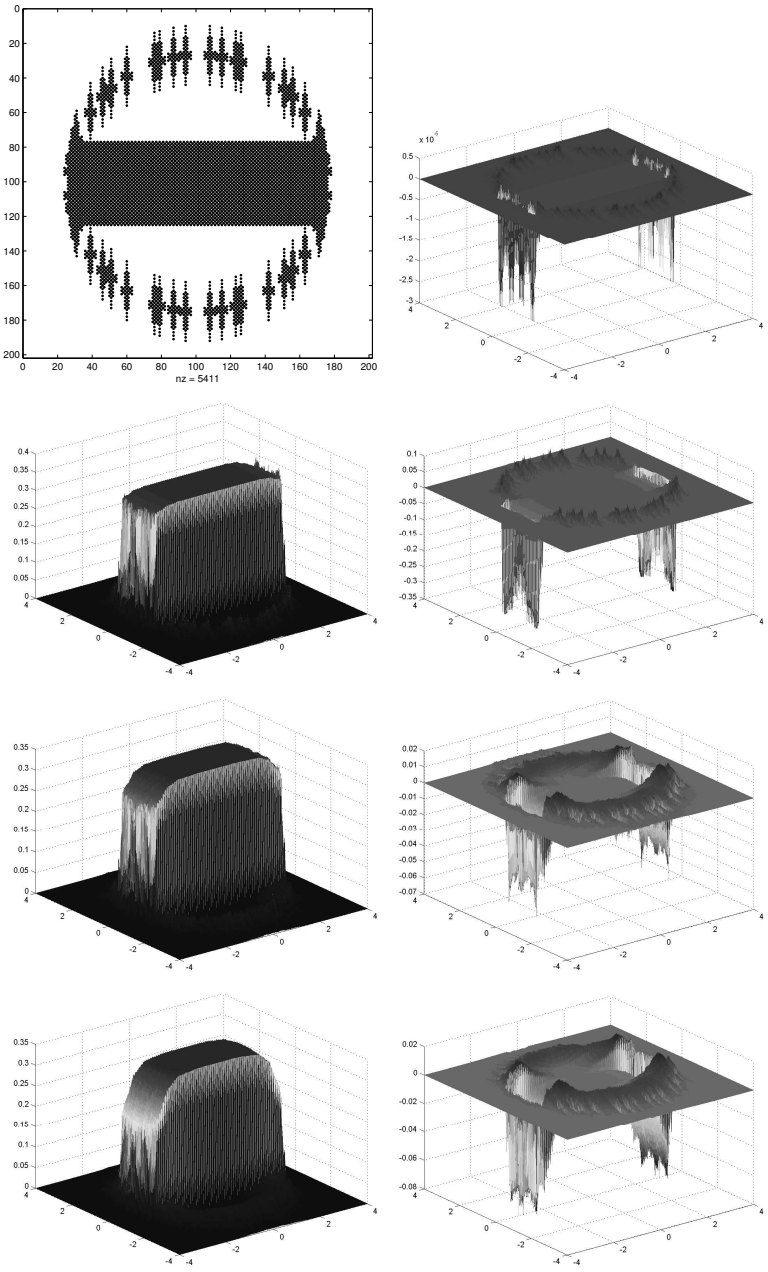


FIGURE 3. Initial data is as in Figure 1. Here only three different values of $n$ are included in eq. (25). The first two plots show the support of the first iterate and the difference $f_1 - f_0$. The next two plots show iterate number 300000 (i.e. $f_{300000}$) and the difference $f_{300000} - f_0$, and then iterate $f_{600000}$ and $f_{600000} - f_{300000}$. The last row shows $f_{900000}$ and $f_{900000} - f_{600000}$.

The plots in figure 3 illustrate that when taking only a small set of values for $n$ (in this case circles with either 40 or 48 points, in total three values for $n$), the rate of convergence to equilibrium is extremely slow. The model is physically not very realistic, as it corresponds to a differential crossection that is concentrated on only three values of $|v - v_*|$, and so it should be considered only as an illustration to the the discussion in the paper.

**Acknowledgment:** We would like to thank A. Bobylev, J. Brzezinski, A. Heintz, and Z. Rudnick for useful discussions.We acknowledge partial support from the National Science Foundation (DMS 0071503) (P.K.), the Swedish Research Council (P.K. & B.W.), the Royal Swedish Academy of Sciences (P.K.), and from the EC funded RTN network HYKE, Contract Number : HPRN-CT-2002-00282 (B.W.)

## REFERENCES

[1] A.V. Bobylev and N. Bernhoff, Discrete velocity models and dynamical systems, in em Lecture Notes on the Discretisation of the Boltzmann Equation, N.Bellomo and R.Gatignol, eds., 2003, pp.203-222.

[2] 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(5):639–644, 1995.

[3] A. Palczewski, J. Schneider, A. Bobylev: A consistency result for a discrete-velocity model of the Boltzmann equation. *SIAM J. Numer. Anal.* **34** (1997), no. 5, 1865–1883.

[4] A. Bobylev, C. Cercignani, Discrete velocity models without nonphysical invariants. *J. Statist. Phys.* **97** (1999),

[5] T. Carleman, *Problèmes mathématiques dans la théorie cinétique des gaz*, Almqvist & Wiksell, Uppsala (1957).

[6] J. Cilleruelo, The distribution of lattice points on circles, *J. Number Theory* **43** (1993), 198-202.

[7] C. Cercignani, R. Illner, M. Pulvirenti: *The mathematical theory of dilute gases*, Springer Verlag (1994).

[8] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. Revised and with a preface by Hugh L. Montgomery.

[9] L. Desvillettes, S. Mischler: About the splitting algorithm for Boltzmann and B.G.K. equations. *Math. Models Methods Appl. Sci.* **6** (1996), no. 8, 1079–1101.

[10] L. Desvillettes, B. Wennberg: Regularity of solutions to the spatially homogeneous Boltzmann equation without cutoff, *Comm. PDE* **29** no 1&2 (2003), 133 –156.

[11] P. Erdős and R. R. Hall. On the angular distribution of Gaussian integers with fixed norm. *Discrete Math.*, 200(1-3):87–94, 1999. Paul Erdős memorial collection.

[12] L. Fainsilber, P. Kurlberg, and B. Wennberg. Lattice points in circles and Discrete velocity models for the Boltzmann equation preprint `http://arxiv.org/pdf/math.AP/0405171`, submitted 2004.

[13] H. Halberstam and H.-E. Richert. On a result of R. R. Hall. *J. Number Theory.*, 11(1):76–89, 1979.

[14] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers*. Oxford, at the Clarendon Press, 1954. 3rd ed.

[15] V. Panferov, A. Heintz, A new consistent discrete-velocity model for the Boltzmann equation, *Math. Methods Appl. Sci.* **25** (2002), no. 7, 571–593.
[16] H. Iwaniec, Fourier coefficients of modular forms of half-integral weight, *Invent. Math.*, **87** (1987), no. 2, 385–401.
[17] I. Kátai and I. Környei. On the distribution of lattice points on circles. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, 19:87–91 (1977), 1976.
[18] I. Kubilyus. The distribution of Gaussian primes in sectors and contours. *Leningrad. Gos. Univ. Uč. Zap. Ser. Mat. Nauk*, 137(19):40–52, 1950.
[19] A. Palczewski, J. Schneider, Existence, stability, and convergence of solutions of discrete velocity models to the Boltzmann equation. *J. Statist. Phys.* **91** (1998), no. 1-2, 307–326.
[20] C. Pommerenke, Über die Gleichverteilung von Gitterpunkten auf $m$-dimensionalen Ellipsoiden, *Acta Arith.* **5** (1959), 227–257.
[21] F. Rogier, J. Schneider: A direct method for solving the Boltzmann equation, *Transport Theory Statist. Phys.* **23** (1994) 313–338.
[22] V.V. Vedenyapin, Velocity inductive construction for mixtures, Transport Theory, Stat. Phys. **28** no 7, (1999), 727–742.
[23] V.V. Vedenyapin, S. A. Amosov, and L. Toscano, Invariants for Hamiltonians and knetic equations, *Russ. Math. Surv.* **54** (1999), 1056–1057.

DEPARTMENT OF MATHEMATICS, CHALMERS UNIVERSITY OF TECHNOLOGY, SE-412 96 GOTHENBURG, SWEDEN
*E-mail address*: laura@math.chalmers.se

DEPARTMENT OF MATHEMATICS, ROYAL INSTITUTE OF TECHNOLOGY, SE-10044 STOCKHOLM, SWEDEN
*E-mail address*: kurlberg@math.kth.se

DEPARTMENT OF MATHEMATICS, CHALMERS UNIVERSITY OF TECHNOLOGY, SE-412 96 GOTHENBURG, SWEDEN
*E-mail address*: wennberg@math.chalmers.se