On the solution of a model Boltzmann Equation via steepest descent in the 2–Wasserstein metric

E. A. Carlen* and W. Gangbo**

School of Mathematics
Georgia Institute of Technology
Atlanta, GA, 30332 U.S.A.

Abstract We study a model Boltzmann equation closely related to the BGK equation using a steepest descent method in the Wasserstein metric, and prove global existence of energy and momentum conserving solutions. We also show that the solutions converge to the manifold of local Maxwellians in the large time limit, and obtain other information on the behavior of the solutions. We show how the Wasserstein metric is natural for this problem because it is adapted to the study of both the free streaming and the "collisions".

Key words: Boltzmann equation, mass transfer, Wasserstein metric.

Contents

* carlen@math.gatech.edu. Work partially supported by U.S. N.S.F. grant DMS 92–07703
** gangbo@math.gatech.edu. Work partially supported by U.S. N.S.F. grant DMS-99-70520 and DMS-00-74037

Introduction

This paper concerns the solution of a spatially inhomogeneous model Boltzmann equation known as the kinetic Fokker–Planck equation, by means of "steepest descent" in the Wasserstein metric. This equation governs the evolution of a probability density f on the phase space, $T^d \times I\!R^d$, of an d dimensional torus. The case of greatest interest is $d = 3$, but in most parts of the analysis, the dimension is not particularly important. We will use x to denote position variables; i.e, points in T^d, and shall use v to denote momentum variables. (We consider only one species of particles, so we choose units in which the mass is unity, and will not distinguish between momentum and velocity.)

Given a probability density $f(x,v)$ on $T^d \times I\!R^d$, we define its spatial density $\rho(x)$ through

$$\rho(x) = \int_{I\!R^d} f(x,v)\mathrm{d}v \ , \tag{1.1}$$

and we define its conditional velocity distribution at x, $F(v;x)$, through

$$F(v;x) = \frac{f(x,v)}{\rho(x)} \tag{1.2}$$

for all x with $\rho(x) > 0$. The *bulk velocity* $u(x)$ is the mean of $F(\cdot;x)$:

$$u(x) = \int_{I\!R^d} vF(v;x)\mathrm{d}v \ , \tag{1.3}$$

and the *temperature* $\theta(x)$ is $1/d$ times the variance of $F(\cdot;x)$:

$$\theta(x) = \frac{1}{d}\int_{I\!R^d} |v-u(x)|^2 F(v;x)\mathrm{d}v \ . \tag{1.4}$$

The total energy $E(f)$ of the density is

$$E(f) = \frac{1}{2}\int_{T^d\times I\!R^d} |v|^2 f(x,v)\mathrm{d}v\mathrm{d}x \ = \frac{1}{2}\int_{T^d} \left[d\theta(x) + |u(x)|^2\right]\rho(x)\mathrm{d}v\mathrm{d}x \ , \tag{1.5}$$

and we shall be concerned exclusively with phase space densities whose energy is finite.

The following probability densities are central to kinetic theory:

Definition. *The Maxwellian density with bulk velocity u and temperature θ, $M_{u,\theta}$, is given by*

$$M_{u,\theta}(v) = (2\pi\theta)^{-d/2}\, e^{-|v-u|^2/2\theta} \ .$$

The Maxwellian with the same temperature $\theta(x)$ and bulk velocity $u(x)$ as $F(\cdot;x)$ is denoted $M_{F(\cdot;x)}$. Finally, the local Maxwellian corresponding to f is the density M_f on the phase space $T^d \times I\!R^d$ where

$$M_f(x,v) = \rho(x)M_{F(\cdot;x)}(v) \ . \tag{1.6}$$

These notational conventions will be used throughout the paper.

The equation studied here is

$$\frac{\partial}{\partial t} f(x,v,t) + \nabla_x\cdot(vf(x,v,t)) = \mathcal{L}_f f(x,v,t) \tag{1.7}$$

where the operator $\mathcal{L}$ is given by

$$\mathcal{L}_f\phi = \theta^p\nabla_v \cdot \left(M_f\nabla_v\left(\frac{\phi}{M_f}\right)\right) = \theta^p\nabla_v \cdot \left(\phi\nabla_v \ln\left(\frac{\phi}{M_f}\right)\right) \tag{1.8}$$

with p being a positive number. (The role of p will be discussed shortly). One easily sees that

$$\mathcal{L}_f\phi = \theta^p(x)\left(\Delta_v\phi + \nabla\cdot\left(\frac{v-u(x)}{\theta(x)}\phi\right)\right) \ . \tag{1.9}$$

The equation is, of course, non linear since the coefficients of $\mathcal{L}_f$ depend on f through certain of its local moments, namely the bulk velocity and the temperature.

The equation (1.7) is closely related to the Boltzmann equation

$$\frac{\partial}{\partial t}f(x,v,t) + \nabla_x\cdot(vf(x,v,t)) = \mathcal{Q}(f(x,v,t)) \ . \tag{1.10}$$

The right hand side $\mathcal{Q}(f)$ is the so–called collision kernel, and we will further discuss it below. In (1.7), $\mathcal{Q}(f)$ has been replaced by $\mathcal{L}_f f$. Among the most characteristic properties of any evolution equation are quantities that are monotone or conserved under the corresponding evolution. In replacing $\mathcal{Q}(f)$ by $\mathcal{L}_f f$, we have not changed the formal conservation and monotonicity properties of the Boltzmann equation. These are conservation of energy, momentum and mass, and increase of entropy:

Indeed, considering any sufficiently smooth solution f of (1.7). Then formally integrating by parts, one easily obtains

$$\frac{\mathrm{d}}{\mathrm{d}t}E(f) = 0 \tag{1.11}$$

so that the energy is conserved by the evolution. Similarly, if $U(f)$ denotes the total momentum; i.e.,

$$U(f) = \int_{T^d\times I\!R^d} vf(x,v)\mathrm{d}x\mathrm{d}v = \int_{T^d} u(x)\rho(x)\mathrm{d}x \tag{1.12}$$

one finds that this is conserved as well:

$$\frac{\mathrm{d}}{\mathrm{d}t}U(f) = 0 \ . \tag{1.13}$$

The conservation of mass, or in other words, $\int_{T^d\times I\!R^d} f\mathrm{d}x\mathrm{d}v$, is clear.

Finally, define the Boltzmann entropy $H(f)$ of f through

$$H(f) = -\int_{T^d\times I\!R^d} f(x,v)\ln f(x,v)\mathrm{d}x\mathrm{d}v \ . \tag{1.14}$$

This time, formal integration by parts leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}H(f) = \int_{T^d}\left[\int_{I\!R^d}\theta(x)\left|\nabla_v\ln F(\cdot;x) - \nabla_v\ln M_{F(\cdot;x)}\right|^2 F(v;x)\mathrm{d}v\right]\rho(x)\mathrm{d}x \ . \tag{1.15}$$

Thus, $H(f)$ is strictly increasing in unless $f = M_f$, and hence the analog of Boltzmann's H–theorem holds for this equation, together with conservation of energy and momentum.

These conservation laws, together with the H–theorem, are the features of the Boltzmann equation that are responsible, at least on a formal level, for its connection with the Euler Equations in a hydrodynamic scaling limit [11]. For this reason, it is important to maintain them in any model kinetic equation. Other such model kinetic equations, particularly the BGK equation, have been discussed extensively in the literature. While in the kinetic Fokker–Planck equation, the collision kernel $\mathcal{Q}(\mathrm{f})$ of (1.10) is replaced by $\mathcal{L}_f f$, in the BGK model it is replaced by a constant mutiple of $M_f - f$. See Cercignani's book [11] for further discussion of the BGK equation. It suffices to remark here that as far as existence, uniqueness and regularity, the present state of knowledge concerning the BGK equation is no better than for the Boltzmann equation itself. A theory leading to global existence of solutions of the Boltzmann equation has been developed by DiPerna and Lions [13], [14]. This is a very significant advance over what had been the state of the art; however, there is no uniqueness result, nor are the solutions shown to conserve energy.

To better understand the relationship between (1.7) and (1.10), we briefly recall a few facts about (1.10) and its origins. Consider the evolution of a gas consisting of a very large number of molecules moving in a large box with periodic boundaries, corresponding to T^d. The full microscopic state of the system is given by specifying all of the positions and velocities.

Such a specification is far too detailed for many purposes, and is at any rate computationally inaccessible, and so Boltzmann [4] sought only to describe something simpler: the single particle density, $f(x, v, t)$. Imagine all of the molecules to be individually labeled, and at time t, we randomly select the label of one of the molecules, which of course has a position x and a velocity v. Thus, in randomly selecting a label, one has randomly generated a point (x, v) of the phase space $T^d \times I\!R^d$, and $f(x, v, t)$ denotes the probability density of this random point on the phase space. One can then ask how this density will evolve in time.

There are two mechanisms at work in the evolution. First, there is *streaming*: In a short time step h a molecule at (x, v) moves to $(x + hv, v)$ provided there is no intervening collision. If there were no collisions, the equation would be

$$\frac{\partial}{\partial t} f(x, v, t) + v \cdot \nabla_x f(x, v, t) = 0 \tag{1.16}$$

and everything would be very simple and uninteresting. The second mechanism is provided by the collisions, which are assumed to be *local* and *binary*. That is, only molecules at the same location x can collide, and each colliding pair completes its collision before either of its members enter into a new one.

In the abscence of streaming, at each x, the local conditional velocity distribution $F(v; x, t)$ is updated by solving

$$\frac{\partial}{\partial t} F(v, t) = Q(F, F)(v, t) \tag{1.17}$$

independently at each x. The equation (1.17) is known as the spatially homogeneous Boltzmann equation, and there is a well developed theory of it [12].

While separately, each of the evolution mechanisms described by (1.16) and (1.17) is well understood, much less is known when they are both present, as in (1.10). Though considerable progress has been made by DiPerna and Lions, there is no uniqueness result for the solutions that these authors construct, and the sense in which they solve the equation does not permit one to conclude that they even conserve energy.

That said, we return to a brief description of the collision mechanism. The collision mechanism conserves the momentum and energy of a pair of colliding particles. For the case $d = 3$, given a pair of pre-collisional velocities v and w, the possible post–collisional velocities v^* and w^* are

$$v^* = \frac{v+w}{2} + \frac{|v-w|}{2}\sigma$$
$$w^* = \frac{v+w}{2} - \frac{|v-w|}{2}\sigma$$

where σ is a unit vector, i.e., $\sigma \in S^2$. If the two particles interact through a force which is inversely proportional to some power s of their separation, the rate of at which such such collisions occur at x is proportional to

$$F(\cdot;x)(w)F(v;x)b(|v-w|,\cos(\vartheta))$$

where $\cos(\vartheta) = \sigma \cdot (v-w)/|v-w|$. The function b has a complicated dependence on $\cos(\vartheta)$ for general s, but it is simply proportional to $|v-w|^{(s-5)/s}$. The case $s = 5$ is referred to as the case of Maxwellian molecules since Maxwell observed [24] that in this case many quantities are readily computed due to the lack of dependence on $|v-w|$. (He actually tried to argue that an inverse fifth power force law did in fact mediate molecular collisions, which is not unreasonable since Nature often seems to opt for simplicity, though in this case, further investigations have not borne Maxwell out.) Another case of special interest is the case of hard sphere collisions, which formally correspond to $s = \infty$. In this case, b is a constant multiple of $|v-w|\cos(\vartheta)$.

Since the underlying dynamics is time-reversible, the reverse collision occurs at an equal rate, and this leads to

$$\mathcal{Q}(F,F)(v) = \int_{I\!R^3}\int_{S^2} \left(F(v^*)F(w^*) - F(v)F(w)\right) b(|v-w|,\cos(\vartheta))\mathrm{d}\sigma\mathrm{d}w \ . \tag{1.18}$$

The term $\mathcal{Q}(f(x,v))$ in (1.10) is then given by

$$\mathcal{Q}(f(x,v)) = \rho^2(t,x)\mathcal{Q}(F(\cdot;x,t),F(\cdot;x,t))(v) \ .$$

Because b is proportional to $|v-w|^{(s-5)/s}$, $\mathcal{Q}$ satisfies a scaling law. That is, if

$$F_\lambda(v) = \lambda^d F(\lambda v) \ ,$$

then

$$\mathcal{Q}(F_\lambda,F_\lambda) = \lambda^{-(s-5)/s}\left[\mathcal{Q}(F,F)\right]_\lambda \ .$$

It is easily seen that

$$\mathcal{L}_{F_\lambda}F_\lambda(v) = \lambda^{2-2p}\left(\mathcal{L}_F F\right)_\lambda \ .$$

Thus for

$$p = 1 + \frac{s-5}{2s} \ , \tag{1.19}$$

$$\mathcal{L}_{F_\lambda}F_\lambda(v) = \lambda^{-(s-5)/s}\left(\mathcal{L}_F F\right)_\lambda \ .$$

Thus, by choosing the value of p in (1.19), we can match the scaling properties of the Boltzmann equation, as well as the conservation laws and the H–theorem. We note that hard sphere scaling corresponds to $p = 2$, and the case of Maxwellian molecules corresponds to $p = 1$. Interestingly enough, this case is especially nice for (1.7) as well as for (1.10).

The model equation (1.7) studied here is obtained by replacing the true Boltzmann collision mechanism for updating $F(\cdot;x)$ at each x by what amounts to a mechanism of steepest ascent in the entropy, under the constraint that the momentum and energy are conserved locally at each x, as required by the conservation laws. The replacement of the true collision mechanism by such a “constained gradient flow” for the entropy is natural if one seeks a simple replacement for the Boltzmann evolution that respects both the conservation laws and the H–theorem.

To explain our starting point, we recall a recent result [20] of Jordan, Kinderlehrer and Otto, who have shown how to regard the linear Fokker–Planck equation as gradient flow for the relative entropy functional.

Let $\mathcal{P}$ denote the set of probability densities on $I\!R^d$ with finite second moments; i.e., the set of all non–negative measurable functions F on $I\!R^d$ such that $\int_{I\!R^d} F(v)\mathrm{d}v = 1$ and $\int_{I\!R^d} |v|^2 F(v)\mathrm{d}v < \infty$. Equip $\mathcal{P}$ with the metric $W_2(F_0, F_1)$ where

$$W_2^2(F_0, F_1) = \inf_{\gamma\in\mathcal{C}(F_0,F_1)} \int_{I\!R^d\times I\!R^d} \frac{1}{2}|v-w|^2\gamma(\mathrm{d}v, \mathrm{d}w) \tag{1.20}$$

where $\mathcal{C}(F_0, F_1)$ consists of all couplings of F_0 and F_1; i.e., the set of all probability measures γ on $I\!R^d \times I\!R^d$ such that for all test functions η on $I\!R^d$

$$\int_{I\!R^d\times I\!R^d} \eta(v)\gamma(\mathrm{d}v, \mathrm{d}w) = \int_{I\!R^d} \eta(v)F_0(v)\mathrm{d}v$$

and

$$\int_{I\!R^d\times I\!R^d} \eta(w)\gamma(\mathrm{d}v, \mathrm{d}w) = \int_{I\!R^d} \eta(w)F_1(w)\mathrm{d}v \ .$$

The infimum in (1.20) is actually a minimum, and it is attained at a unique point γ_{F_0,F_1} in $\mathcal{C}(F_0, F_1)$. Recent results of Brenier, Caffarelli, Gangbo and McCann have shed considerable light on the nature of this minimizer, and we shall recall some of their results later. But for the purposes of the introduction, it suffices the metric, commonly called the Wasserstein metric, is defined.

Next, let the entropy $H(F)$ be defined by

$$H(F) = -\int_{I\!R^d} F(v)\ln F(v)\mathrm{d}v \ . \tag{1.21}$$

This is well defined, with $-\infty$ as a possible value under the assumption that $\int_{I\!R^d} |v|^2F(v)\mathrm{d}v$ is finite. Given any Maxwellian density M on $I\!R^d$, the relative entropy of F with respect to M, $H(F|M)$, is defined by

$$H(F|M) = \int_{I\!R^d} \left(\frac{F(v)}{M(v)}\right) \ln\left(\frac{F(v)}{M(v)}\right) M(v)\mathrm{d}v \ . \tag{1.22}$$

Kinderlehrer Jordan and Otto introduced the following scheme for solving the linear Fokker–Planck equation: Fix an initial density F_0 with $\int_{I\!R^d} |v|^2F_0(v)\mathrm{d}v$ finite. Further fix a time step $h > 0$. Then inductively define F_k in terms of F_{k-1} by choosing F_k to minimize the functional

$$F \to \left[W_2^2(F_{k-1}, F) + hH(F|M)\right] \tag{1.23}$$

on $\mathcal{P}$. They show that there is a unique maximizer F_k, so that each F_k is well defined. Then they define a time dependent probability density $F^{(h)}(v, t)$ by putting

$$F^{(h)}(v, kh) = F_k$$

and linearly interpolating when t is not an integer multiple of h. Finally, they show that for each t

$$F(\cdot,t) = \lim_{h\to 0} F^{(h)}(\cdot,t)$$

exists weakly in L^1, and that the resulting time dependent probability density solves

$$\frac{\partial}{\partial t}F(v,t) = \nabla\cdot\left(F(v,t)\nabla\ln\left(\frac{F(v,t)}{M(v)}\right)\right) \tag{1.24}$$

with

$$\lim_{t\to 0} F(\cdot,t) = F_0\ .$$

The equation (1.24) has much in common with (1.17). Our object here is to combine this "collision mechanism" with streaming and to study the resulting model kinetic equation. Our approch is to implement a "splitting scheme" in which we alternately run the streaming and collisions in a succesion of time intervals of length $h > 0$. We shall obtain estimates on this process that are independent of h, and which lead to solutions of (1.7). A key feature in our approach is that the Wasserstein 2-metric is naturally suited to both the streaming and the substitute collision mechanism, as we shall explain further on.

As t tends to infinity on the other hand, $F(\cdot,t)$ tends to M, and the bulk velocity $u_{F(\cdot,t)}$ and the temperature $\theta_{F(\cdot,t)}$ of $F(\cdot,t)$ tend to the bulk velocity and temperature of M. Indeed, one easily sees that *if the initial data F_0 and M have the same bulk velocity and temperature*, then these quantities are conserved for solutions of (1.24). That is, the evolution given by

$$\frac{\partial}{\partial t}F(v,t) = \nabla\cdot\left(F(v,t)\nabla\ln\left(\frac{F(v,t)}{M_{F_0}(v)}\right)\right) \tag{1.25}$$

conserves energy and momentum, while, as easily seen, decreasing the entropy. At the discrete level, this evolution is obtained by choosing $M = M_{F_0}$ in (1.23).

This suggest the following scheme for the solution of (1.7): Fix an initial phase space density $f_0(x,v)$, and a time step $h > 0$. Inductively define $f_k(x,v)$ in terms of $f_{k-1}(x,v)$ through the following algorithm:

Definition. (The flow and descend algortithm) This algorithm consists of the following steps:

(1) First, run the streaming: Define $\tilde f_k(x,v)$ by

$$\tilde f_k(x,v) = f_{k-1}(x-hv,v)\ . \tag{1.26}$$

(2) Define $\rho_k(x) = \int_{I\!R^d} \tilde f(x,v)\mathrm{d}v$ and the precollision conditional velocity distribution $\tilde F_k(v;x)$ by

$$\tilde F_k(v;x) = \frac{\tilde f_k(x,v)}{\rho_k(x)}\ . \tag{1.27}$$

Then define $\tilde u_k(x)$ and $\tilde\theta_k(x)$ by

$$\tilde u_k(x) = \int_{I\!R^d} v\tilde F_k(v;x)\mathrm{d}v \qquad \text{and} \qquad \tilde\theta_k(x) = \frac{1}{d}\int_{I\!R^d} |v-\tilde u_k(x)|^2\tilde F_k(v;x)\mathrm{d}v\ . \tag{1.28}$$

(3) Now run the collisions though steepest descent of the relative entropy: For each x, let $F_k(v;x)$ minimize the functional

$$F \to \left[\frac{W_2^2(F, \tilde{F}_k(\cdot;x))}{\tilde{\theta}_k(x)} + hH(F|M_{\tilde{F}_k(\cdot;x)}) \right] \tag{1.29}$$

over $\mathcal{P}$. This functional, with the factor of $\tilde{\theta}$ in the denominator is scale invariant. As we shall see, it leads to (1.7) with $p = 1$ in (1.8). As explained around (1.19), this makes $\mathcal{L}_F F$ scale invariant, as is $\mathcal{Q}(F,F)$ in the case of Maxwellain molecules.

Notice that the "target Maxwellian" $M_{\tilde{F}_k}$ changes from step to step due to the effects of the streaming. This is what produces the non-linearity in our equation, and separates this evolution form the linear evolution studied in [20].

(4) Finally, one reconstructs $f_k(x,v)$ through

$$f_k(x,v) = \rho_k(x) F_k(v;x) \ , \tag{1.30}$$

which completes the passage from f_{k-1} to f_k. Then define a time dependent phase space probability density $f^h(x,v,t)$ through

$$f^h(x,v,kh) = f_k(x,v)$$

and by an interpolation when t is not an integer multiple of h.

We shall show that for each t, the weak L^1 limit

$$f(x,v,t) = \lim_{h \to 0} f^{(h)}(x,v,t)$$

exists and satisfies the evolution equation (1.7).

The difficulty in studying equations such as (1.7) arises largely from the fact that there is no direct mechanism producing regularity in x. The chief advantage of working in the Wasserstein metric here is that the distance between $f(x,v)$ and $f(x-hv,v)$ is easily controlled in terms of the energy without any recourse to estimates on spatial derivatives of $f(x,v)$, as we shall see.

On the other hand, several difficulties arise in connection with the discrete method. For example, the discrete Fokker–Planck evolution does not conserve the energy; at any finite h, there is energy dissipation. However, it is possible to bound the energy dissipation, and to show that the size of the effect vanished as h tends to zero.

Another natural approach to this evolution would be to replace the relative entropy $H(F|M_{\tilde{F}_k})$ by the entropy $H(F)$ itself. Then one would have to enforce energy conservation by an explicit constraint. That is, one would restrict the class of densities F to be considered in the variational problem in step 3 for the functional in (1.29) to the "submanifold" $\mathcal{S}_{(u,\theta)}$ of $\mathcal{P}$ consisting of densities. This constrained minimization problem is much more delicate due to the fact that $\mathcal{S}_{(u,\theta)}$ is not weakly closed in $\mathcal{P}$. Though we have proven existence and uniqueness of minimizers for this constrained problem in [10], there are open problems, discussed in in [10], about the nature of the minimizers that complicate the approach via constrained descent. Hence we have chosen the present route using the relative entropy and bounding energy dissipation in the present paper.

The paper is organized as follows. In section 2, we derive the Euler–Lagrange equation for $F_{k+1}(x;v)$, and establish the formal connection between the algorithm specified in (1.26) through (1.30). In sections 3 and 4, we establish, on the basis of the variational principle, several of the *a–priori* estimates needed to make the connection rigorous. In section 5, we prove crucial modulus of continuity in t estimates. In section 6, we establish a form of the velocity averaging lemma, and then finally in section 7 take the limit as h tends to zero and prove the main existence and regularity theorem. In section 8, we conclude by establishing that in the limit as t tends to infinity, the solutions of (1.7) tend to set $\mathcal{M}$ of local Maxwellian densities.

2 The Euler–Lagrange equation

The main purpose of this section is to derive the Euler Lagrange equation for the minimizer of the discrete–time problem, and to introduce the appropriate continuous time interpolation. The variational problem considered here is the following: For fixed $h > 0$, and a given density $\tilde{F}_k$, we seek to minimize the functional

$$\int_{TR^d} \rho_k(x) \left[\frac{W_2^2(\tilde{F}, F_k(\cdot;x))}{\tilde{\theta}_k(x)} + hH(F|M_{\tilde{F}_k(\cdot;x)}) \right] \mathrm{d}x \ . \tag{2.1}$$

Existence an uniqueness of the optimizers for this problem is proved in [20], where the Euler–Lagrange equation is derived as well. The differences here are that the "target Maxwellian" $M_{\tilde{F}_k}$ is fixed and independent of k in [20], and of course, the fact that there is no x dependence to deal with. At each fixed x and time step however, these differences are immaterial.

We begin by briefly sketching the derivation of the Euler–Lagrange equation in a notation suited to our application. We refer [20] for details. We will then use this to make a first formal connection between the discrete the scheme defined in (1.26) through (1.30) with the kinetic Fokker–Planck equation (1.7).

Lemma 2.1 (Explicite expression for optimal maps) *Let F_k be the unique minimizer to of the functional given in (2.1). Let $\tilde{\psi}_k$ be the convex function on $I\!R^d$ such that*

$$\nabla\tilde{\psi}_k \# F_k = \tilde{F}_k \ .$$

Then

$$\nabla\tilde{\psi}_k(v) = v + h\tilde{\theta}_k(x)\nabla_v \left(\ln \frac{F_k}{M_{\tilde{F}_k}} \right) \ . \tag{2.2}$$

Proof: Consider a smooth function $\xi : I\!R^d \to I\!R^d$ and define the flow $T_t(v) = v + t\xi(v)$. Use this to define the parameterized family of probability densities $G(t)$ by

$$G(t) = (T_t)\#F_k \ .$$

$$\frac{\mathrm{d}}{\mathrm{d}t} H(G(t)|M_{\tilde{F}_k})\bigg|_{t=0} = \int_{I\!R^d} \nabla_v \left(\ln \frac{F_k}{M_{\tilde{F}_k}} \right) \cdot \xi(v) F_k(v;x) \mathrm{d}v \ . \tag{2.3}$$

For a detailed justification of this computation, as well as what follow in the next paragraph, see [20] or, in similar notation, [10]. To compute the variation in the 2–Wasserstein distance, note that since $T_t\#F_k = G(t)$, $\nabla\tilde{\psi}_k \circ T_t^{-1}\#G(t) = \tilde{F}_k$. Thus

$$\begin{aligned} W_2^2(G(t),\tilde{F}_k) &\le \frac{1}{2}\int_{I\!R^d} |\nabla\tilde{\psi}_k \circ T_t^{-1}(v) - v|^2 G(t,v)\mathrm{d}v \\ &= \frac{1}{2}\int_{I\!R^d} |\nabla\tilde{\psi}_k - T_t(v)|^2 F_k(v;x)\mathrm{d}v \\ &\le W_2^2(F_k,\tilde{F}_k) - t\int_{I\!R^d} \left(\nabla\tilde{\psi}_k - v\right)\cdot\xi F_k(v;x)\mathrm{d}v \\ &+ \frac{t^2}{2}\int_{I\!R^d} |\xi|^2 F_k(v;x)\mathrm{d}v\ . \end{aligned}$$

From this, (2.3) and the fact that F_k is a minimizer, it follows that

$$\int_{I\!R^d} \left(\left(v - \nabla\tilde{\psi}_k(v)\right) + h\tilde{\theta}_k(x)\nabla_v\left(\ln\frac{F_k}{M_{F_k}}\right)\right)\cdot\xi(v)F_k(v;x)\mathrm{d}v \le 0\ . \tag{2.4}$$

Replacing ξ by $-\xi$, we obtain the desired equality. ■

We now proceed from the Euler–Lagrange equation to the weak form of the evolution equation. Let ϕ be any test function on $I\!R^d$. Then

$$\int_{I\!R^d} \phi(v)(F_k - \tilde{F}_k)\mathrm{d}v = \int_{I\!R^d} (\phi(v) - \phi(\nabla\tilde{\psi}_k(v))F_k\mathrm{d}v\ . \tag{2.5}$$

Now define

$$K_k[\phi](v) = \int_0^1\int_0^t D^2\phi(v + s(\nabla\tilde{\psi}_k(v) - v))\langle\nabla\tilde{\psi}_k(v) - v, \nabla\tilde{\psi}_k(v) - v\rangle\mathrm{d}s\mathrm{d}t\ , \tag{2.6}$$

so that we have

$$\phi(\nabla\tilde{\psi}_k(v)) = \phi(v) + \nabla\phi(v)\cdot\left(\nabla\tilde{\psi}_k(v) - v\right) + K_k[\phi](v)\ . \tag{2.7}$$

Combining (2.5) and (2.7), and then using (2.2), we obtain:

$$\begin{aligned} \int_{I\!R^d} \phi(v)(F_k - \tilde{F}_k)\mathrm{d}v &= \int_{I\!R^d} \nabla\phi(v)\cdot\left(\nabla\tilde{\psi}_k(v) - v\right)F_k(v;x)\mathrm{d}v \\ &\quad + \int_{I\!R^d} K_k[\phi](v)F_k(v;x)\mathrm{d}v \\ &= h\int_{I\!R^d} \nabla\phi(v)\cdot\left(\tilde{\theta}_k(x)\nabla\left(\ln\frac{F_k}{M_{\tilde{F}_k}}\right)\right)F_k(v;x)\mathrm{d}v \\ &\quad + \int_{I\!R^d} K_k[\phi](v)F_k(v;x)\mathrm{d}v\ . \end{aligned} \tag{2.8}$$

Let u_k and θ_k be defined by

$$u_k(x) = \int_{I\!R^d} vF_k(v;x)\mathrm{d}v \qquad \text{and} \qquad \theta_k(x) = \frac{1}{d}\int_{I\!R^d} |v - u_k(x)|^2 F_k(v;x)\mathrm{d}v\ , \tag{2.9}$$

which corresponds to the definition (1.28) of $\tilde{u}_k$ and $\tilde{\theta}_k$. We shall see in the next section, in (3.5) and (3.8), that for each x and k,

$$u_k(x) = \tilde{u}_k(x) \qquad \text{but} \qquad \theta_k(x) < \tilde{\theta}_k(x) \ .$$

Therefore

$$\begin{aligned}\int_{I\!R^d} \nabla\phi(v) \cdot \left(\tilde{\theta}_k(x)\nabla\left(\ln\frac{F_k}{M_{\tilde{F}_k}}\right)\right) F_k(v;x)\mathrm{d}v = \int_{I\!R^d} \nabla\phi(v) \cdot \left(\theta_k(x)\nabla\left(\ln\frac{F_k}{M_{F_k}}\right)\right) F_k(v;x)\mathrm{d}v \\ + (\tilde{\theta}_k(x) - \theta_k(x)) \int_{I\!R^d} \Delta_v \phi(v) F_k(v;x)\mathrm{d}v \ .\end{aligned}$$

Now let ϕ denote a test function on the whole phase space, $T^d \times I\!R^d$. Integrating both sides of (2.8) against $\rho_k(x)$, one obtains

$$\begin{aligned}\int_{T^d\times I\!R^d} (f_k - \tilde{f}_k)\phi \mathrm{d}v\mathrm{d}x =& h\int_{T^d\times I\!R^d} (\mathcal{L}_{f_k}\phi)\, f_k \mathrm{d}v\mathrm{d}x \\ &+ h\int_{T^d\times I\!R^d} (\tilde{\theta}_k(x) - \theta_k(x))\Delta_v\phi(v) f_k(v;x)\mathrm{d}x\mathrm{d}v \\ &+ \int_{T^d\times I\!R^d} K_k[\phi] f_k \mathrm{d}v\mathrm{d}x \ .\end{aligned} \tag{2.10}$$

Provided the terms in the last two lines of (2.10) are negligible, as we shall see, (2.10) bears a close resemblance to a weak form of the kinetic Fokker–Planck equation.

To make the passage from discrete to continuous time, we require a properly chosen interpolation of our sequence of densities f_k to obtain the *discontinuous* time dependent density f^h where for t in $[t_k, t_{k+1})$,

$$f^h(x,v,t) = f_k(x - (t-t_k)v, v) \qquad \text{for} \quad t \in [t_k, t_{k+1}) \ , \tag{2.11}$$

and by convention, $t_k = kh$. Note that by this definition and (1.26),

$$\lim_{t\uparrow t_{k+1}} f^h(x,v,t) = \tilde{f}_{k+1}(x,v) \qquad \text{and} \qquad f^h(x,v,t_k) = f_k(x,v) \ .$$

Also, if ϕ is any test function on $T^d \times I\!R^d \times I\!R_+$, we have for t in $[t_k, t_{k+1})$,

$$\int_{T^d\times I\!R^d} f^h(x,v,t)\phi(x,v,t)\mathrm{d}x\mathrm{d}v = \int_{T^d\times I\!R^d} f_k(x,v)\phi(x + (t-t_k)v, v, t)\mathrm{d}x\mathrm{d}v \ .$$

It follows that

$$\begin{aligned}&\int_{t_k}^{t_{k+1}} \left[\int_{T^d\times I\!R^d} f^h \ \left(\frac{\partial\phi}{\partial t} + v\cdot\nabla_x\phi\right)\mathrm{d}x\mathrm{d}v\right]\mathrm{d}t = \\ &\int_{T^d\times I\!R^d} \left(\tilde{f}_{k+1}(x,v)\phi(x,v,t_{k+1}) - f_k(x,v)\phi(x,v,t_k)\right)\mathrm{d}x\mathrm{d}v \ .\end{aligned} \tag{2.12}$$

Now let $T = Nh$ for some positive integer N, and suppose that $\phi(\cdot,\cdot,t) = 0$ for $t = 0$ and $t = T$. Define

$$A(\phi) = \int_0^T \int_{T^d\times I\!R^d} f^h \ \left(\frac{\partial\phi}{\partial t} + v\cdot\nabla_v\phi\right)\mathrm{d}x\mathrm{d}v\mathrm{d}t \ .$$

Now sum (2.12) to obtain

$$\begin{aligned} A(\phi) &= \sum_{k=0}^{N-1} \int_{T^d\times \mathbb{R}^d} \Big(\tilde f_{k+1}(x,v)\phi(x,v,t_{k+1}) - f_k(x,v)\phi(x,v,t_k)\Big)\mathrm{d}x\mathrm{d}v \\ &= -\sum_{k=1}^{N-1} \int_{T^d\times \mathbb{R}^d} \Big(f_k(x,v) - \tilde f_k(x,v)\Big)\phi(x,v,t_k)\mathrm{d}x\mathrm{d}v\ . \end{aligned} \tag{2.13}$$

Using (2.10) in (2.13), one obtains the following:

Theorem 2.2: (Approximate transport equation: part I) *Let f_k be a sequence of probability densities given by the maximize and flow algorithm for a fixed time step $h > 0$. Let $f^h(x,v,t)$ be defined by (2.11). For any integer N, let $T = Nh$. Then for any test function ϕ on $T^d \times \mathbb{R}^d \times \mathbb{R}_+$ such that $\phi(\cdot,\cdot,t) = 0$ for $t = 0$ and $t = T$,*

$$\begin{aligned} \int_0^T \int_{T^d\times \mathbb{R}^d} f^h\ \left(\frac{\partial \phi}{\partial t} + v\cdot\nabla_v\phi\right)\mathrm{d}x\mathrm{d}v\mathrm{d}t = &-h\sum_{k=0}^{N-1}\int_{T^d\times \mathbb{R}^d} \left(\mathcal{L}_{f_k}\phi(x,\cdot,t_k)\right) f_k\mathrm{d}v\mathrm{d}x \\ &-h\sum_{k=0}^{N-1}\int_{T^d\times \mathbb{R}^d} (\tilde\theta_k(x) - \theta_k(x))\Delta_v\phi(x,\cdot,t_k) f_k\mathrm{d}x\mathrm{d}v \\ &-\sum_{k=0}^{N-1}\int_{T^d\times \mathbb{R}^d} K_k[\phi(x,\cdot,t_k)] f_k\mathrm{d}v\mathrm{d}x\ . \end{aligned} \tag{2.14}$$

where K_k is given in (2.6).

We shall have to show that the last two sums in (2.14) are negligible as h tends to zero. It is clear that for any $T > 0$, and any N with $Nh < T$,

$$\left|\sum_{k=0}^{N-1}\int_{T^d\times \mathbb{R}^d} (\tilde\theta_k(x) - \theta_k(x))\Delta_v\phi(x,\cdot,t_k) f_k\mathrm{d}x\mathrm{d}v\right| \le \|D_v^2\phi\|_{L^\infty(T^d\times\mathbb{R}^d\times\mathbb{R}_+)}\int_{T^d}\sum_{k=1}^{N-1}\left|\tilde\theta_k(x) - \theta_k(x)\right|\rho_k(x)\mathrm{d}x\ . \tag{2.15}$$

We shall show in Section 4 that the sum on the right vanishes as h tends to zero. In fact, as we shall see $\theta_k(x) - \theta_k(x) > 0$ for all x and k, so that the absolute value sign in the sum on the right is superfluous.

We close this section with a simple preparatory estimate on the summand in the final term in (2.14).

Lemma 2.3: *For each k,*

$$\int_{T^d\times \mathbb{R}^d} |K_k[\phi(x,\cdot,t_k)]|\, f_k\mathrm{d}v\mathrm{d}x \le \int_{T^d} \|D_v^2\phi(x,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^d\times\mathbb{R}_+)}\rho_k(x) W_2^2(F_k,\tilde F_k)\mathrm{d}x\ .$$

Proof: From (2.6), one has for each x

$$\int_{\mathbb{R}^d} |K_k[\phi(x,\cdot,t_k)]|\, F_k\mathrm{d}v \le \|D_v^2\phi(x,\cdot,\cdot)\|_{L^\infty(\mathbb{R}^d\times\mathbb{R}_+)} W_2^2(F_k,\tilde F_k).$$

We multiply both sides of the above inequality by ρ_k and integrate both sides of the subsequent inequality over T^d to conclude the proof of Lemma 2.3. ■

3 Displacement convexity and a-priori bounds

The main purpose of this section is to derive *a–priori* moment, L^p, and energy estimates for our discrete time evolution. All of these are deduced as consequences of McCann's displacement convexity. If Φ is a functional on the space $\mathcal{P}$ of probability densities G on $I\!R^d$ with finite second moments, then Φ is *displacement convex* in case for all t with $0 < t < 1$, and all G_0 and G_1 in $\mathcal{P}$,

$$\Phi(G_t) \leq t\Phi(G_0) + (1-t)\Phi(G_1) \tag{3.1}$$

where $t \mapsto G_t$ is the geodesic for the 2-Wasserstein metric joining G_0 and G_1. This is given by

$$G_t = \nabla\psi_t \# G_0$$

where ψ is the convex potential for the map $\nabla\psi$ that pushes G_0 forward onto G_1, and

$$\psi_t(v) = (1-t)\frac{|v|^2}{2} + t\psi(v) \ . \tag{3.2}$$

Then it follows that

$$\Phi(G_1) - \Phi(G_0) \geq \lim_{h\to 0^+} \frac{1}{h}\left(\Phi(G_h) - \Phi(G_0)\right) \ . \tag{3.3}$$

We will apply this for various choices of Φ with $G_0 = F_k$ and $G_1 = \tilde{F}_k$, which we can do since since (2.2) gives us the form of the potential ψ for which $\nabla\psi\# F_k = \tilde{F}_k$. Note that since $\tilde{F}_k$ precedes F_k in the evolution, (3.3) gives us a bound on the growth over one "collision" step of the quantity measured by Φ.

On the other hand, if we have an *a–priori* bound on how large the quantity measured by Φ can get, then (3.3) gives us a bound on the derivative on the right hand side. We will use (3.3) in both of these ways.

First, we consider moments. The functional

$$G \mapsto \int_{I\!R^d} W(v)G(v)\mathrm{d}v \tag{3.4}$$

is displacement convex whenever W is convex on $I\!R^d$, strictly so when W is strictly convex on $I\!R^d$. Therefore, taking $W(v) = |v|^{2m}$ for m an integer, we can get a bound on the growth of moments from (3.3). In this case, we could also proceed by direct computation, and this is convenient for small m. Indeed, we can compute the change in the first and second moments as follows:

For given $\tilde{F}_k$, let F_k be the optimizer in (2.1), and let $\tilde{\psi}_k$ be such that $\nabla\tilde{\psi}\# F_k = \tilde{F}_k$. First, by (2.2)

$$\begin{aligned} \tilde{u}_k = \int_{I\!R^d} v\tilde{F}_k\mathrm{d}v &= \int_{I\!R^d} \nabla\tilde{\psi}F_k\mathrm{d}v \\ &= \int_{I\!R^d} vF_k\mathrm{d}v + h\tilde{\theta}_k\int_{I\!R^d} \nabla(\ln F_k - \ln M_{\tilde{F}_k})F_k\mathrm{d}v \\ &= u_k + h(u_k - \tilde{u}_k) \ . \end{aligned} \tag{3.5}$$

Thus, $u_k = \tilde{u}_k$, and the passage from $\tilde{F}_k$ to F_k conserves momentum. Energy however, is dissipated, as the next calculation shows.

$$
\begin{aligned}
\int_{\mathbb{R}^d} |v|^2 \tilde{F}_k \mathrm{d}v &= \int_{\mathbb{R}^d} |\nabla\tilde{\psi}_k|^2 F_k \mathrm{d}v \\
&= \int_{\mathbb{R}^d} \left|v + h\tilde{\theta}_k \nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v \\
&= \int_{\mathbb{R}^d} |v|^2 F_k \mathrm{d}v + h2\tilde{\theta}_k \int_{\mathbb{R}^d} v\cdot\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right) F_k \mathrm{d}v \\
&\qquad + h^2\tilde{\theta}_k^2 \int_{\mathbb{R}^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v \\
&= \int_{\mathbb{R}^d} |v|^2 F_k \mathrm{d}v - h2d\left(\tilde{\theta}_k - \theta_k\right) + h^2\tilde{\theta}_k^2 \int_{\mathbb{R}^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v \ .
\end{aligned}
\tag{3.6}
$$

Therefore,

$$
\int_{\mathbb{R}^d} |v|^2 \tilde{F}_k(v;x)\mathrm{d}v = \int_{\mathbb{R}^d} |v|^2 F_k(v;x)\mathrm{d}v + \frac{h^2}{1+2h}\tilde{\theta}_k^2(x) \int_{\mathbb{R}^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v \ , \tag{3.7}
$$

and hence

$$
\theta_k(x) < \tilde{\theta}_k(x) \ . \tag{3.8}
$$

Integrating against $\rho_k(x)$, we see that the energy is monotonicly decreasing for the discrete scheme. Formally, the decrease is of order h^2, and one might therefore expect the energy to be conserved in the limit as h tends to zero. We shall establish this in the next sections, but first we need the *a–priori* inequalities that we now derive.

For higher moments, it is more convenient to use (3.3) . For any positive integer m, define

$$
I_{2m}(k) = \int_{T^d\times\mathbb{R}^d} |v - u_k(x)|^{2m} f_k(x,v)\mathrm{d}x\mathrm{d}v \ . \tag{3.9}
$$

By the conservation of momentum that we have just displayed, we may as well assume, making a common translation of both $\tilde{F}_k$ and F_k, that

$$
\tilde{u}_k = u_k = 0 \ . \tag{3.10}
$$

Then since $v \mapsto |v|^{2m}$ is a convex function of v, we are considering a functional of the form in (3.4), for which

$$
\int_{\mathbb{R}^d} |v|^{2m}\tilde{F}_k(v)\mathrm{d}v - \int_{\mathbb{R}^d} |v|^{2m} F_k(v)\mathrm{d}v \geq \lim_{h\to 0^+}\frac{1}{h}\int_{\mathbb{R}^d} |v|^{2m}\big(F_k^h(v) - F_k^0(v)\big)\mathrm{d}v \tag{3.11}
$$

where $F_k^t = \nabla\phi_t \# F_k$ and

$$
\phi_t = \frac{|v|^2}{2} + t\left(\tilde{\psi}_k - \frac{|v|^2}{2}\right) \tag{3.12}
$$

for $0 \leq t \leq 1$. From well–known formulas that can be found in [10] in a similar notation, it follows that any convex function W on $\mathbb{R}^d$,

$$
\lim_{h\to 0^+}\frac{1}{h}\int_{\mathbb{R}^d} W(v)\big(F_k^h(v) - F_k^0(v)\big)\mathrm{d}v = \int_{\mathbb{R}^d} \nabla W(v)\cdot(\nabla\tilde{\psi}_k(v) - v)F_k^0(v)\mathrm{d}v \ .
$$

Then using $W(v) = |v|^{2m}$ and the explicit form (2.2) for $\tilde{\psi}_k$, we have that the right hand side of (3.11) is equal to

$$2m \int_{I\!R^d} |v|^{2m-2} v \cdot \left(h\tilde{\theta}_k(x) \nabla_v \left(\ln \frac{F_k}{M_{\tilde{F}_k}} \right) \right) F_k(v) \mathrm{d}v \ ,$$

and hence one easily computes that

$$\begin{aligned} \int_{I\!R^d} |v|^{2m} \tilde{F}_k(v) \mathrm{d}v - \int_{I\!R^d} |v|^{2m} F_k(v) \mathrm{d}v \geq & h2m \int_{I\!R^d} |v|^{2m} F_k(v) \mathrm{d}v \\ & - h\tilde{\theta}_k(x) 2m(2m-2+d) \int_{I\!R^d} |v|^{2m-2} F_k(v) \mathrm{d}v \ . \end{aligned} \tag{3.13}$$

Observe that in particular, when $m = 1$, this reduces an inequality for $\theta_k(x)$ and $\tilde{\theta}_k(x)$, namely $\tilde{\theta}_k(x) \geq \theta_k(x)$ for all x and k, as deduced above.

Next, restore $\tilde{u}_k(x)$ and $u_k(x)$, and integrate both sides of (3.13) against $\rho_k(x)$ to obtain

$$\begin{aligned} & I_{2m}(k-1) - I_{2m}(k) \geq \\ & 2mh \left[I_{2m}(k) - (2m-2+d) \int_{T^d} \left(\tilde{\theta}_k(x) \int_{I\!R^d} |v - u_k(x)|^{2m-2} F_k(v) \mathrm{d}v \right) \rho_k(x) \mathrm{d}x \right] \ . \end{aligned}$$

By applying Hölder to remaining explicit integral, we obtain

$$\begin{aligned} \int_{T^d} \left(\tilde{\theta}_k(x) \int_{I\!R^d} |v - u_k(x)|^{2m-2} F_k(v) \mathrm{d}v \right) \rho_k(x) \mathrm{d}x \leq & \left(\int_{T^d} \tilde{\theta}_k^m(x) \rho_k(x) \mathrm{d}x \right)^{1/m} I_{2m}(k)^{(m-1)/m} \\ \leq & \frac{1}{d} I_{2m}(k-1)^{1/m} I_{2m}(k)^{(m-1)/m} \\ \leq & \frac{1}{dm} I_{2m}(k-1) + \frac{m-1}{dm} I_{2m}(k) \ , \end{aligned}$$

since, by Jensen's inequality,

$$\int_{I\!R^d} |v - \tilde{u}(x)|^{2m} \tilde{F}_k(v;x) \mathrm{d}v \geq \left(\int_{I\!R^d} |v - \tilde{u}(x)|^2 \tilde{F}_k(v;x) \mathrm{d}v \right)^m = d^m \tilde{\theta}_k(x)^m \ .$$

Therefore,

$$I_{2m}(k) \left[1 + \frac{2h}{d}(d - 2(m-1)^2) \right] \leq I_{2m}(k-1) \left[1 + \frac{2h}{d}(d + 2(m-1)) \right] \ .$$

We have now proved the following result:

Theorem 3.1: *For all k,*

$$\int_{T^d \times I\!R^d} v f_k(x,v) \mathrm{d}v \mathrm{d}x = \int_{T^d \times I\!R^d} v f_0(x,v) \mathrm{d}v \mathrm{d}x \tag{3.14}$$

and

$$\int_{T^d \times I\!R^d} |v|^2 f_k(x,v) \mathrm{d}v \mathrm{d}x \leq \int_{T^d \times I\!R^d} |v|^2 f_0(x,v) \mathrm{d}v \mathrm{d}x \ . \tag{3.15}$$

Moreover, for all k and all $m > 1$, there is a constant C depending only on m and d so that

$$I_{2m}(k) \leq (1 + Ch)^k I_{2m}(0) \ . \tag{3.16}$$

We shall also need the energy inequality that results from applying the displacement convexity argument to the entropy, which is displacement convex by McCann's criterion. The entropy function is not of the form (3.4), but rather of type

$$G \mapsto \int_{I\!R^d} g(G(v))\mathrm{d}v \ . \tag{3.17}$$

McCann has proved [26] a general criterion for the convexity of such functionals, and shown how to compute their derivatives along geodesics in the Wasserstein metric. The relevant formulas can be found in [10] in a notation similar to that used here.

Theorem 3.2: (Energy inequality) *For all k, we have that*

$$\begin{aligned} H(f_k) - H(f_{k-1}) &= H(f_k) - H(\tilde{f}_k) \\ &\geq h \int_{T^d \times I\!R^d} \tilde{\theta}_k(x) \left| \nabla_v \left(\ln \frac{f_k}{M_{\tilde{f}_k}} \right) \right|^2 f_k(x,v)\mathrm{d}x\mathrm{d}v \\ &\geq h \int_{T^d \times I\!R^d} \tilde{\theta}_k(x) \left| \nabla_v \left(\ln \frac{f_k}{M_{f_k}} \right) \right|^2 f_k(x,v)\mathrm{d}x\mathrm{d}v \ . \end{aligned} \tag{3.18}$$

Note. Notice the different subscripts on the two Maxwellians in (3.18). Both estimates will be used later, this is why they are all recorded here.

Proof: The first equality in (3.18) is easy to obtain. Next, because the entropy is a convex function of F, we have that

$$\int_{I\!R^d} \Big(\ln(\tilde{F}_k(v))\tilde{F}_k(v) - \ln\left(F_k(v)\right) F_k(v) \Big)\mathrm{d}v \geq \lim_{h\to 0} \frac{1}{h} \int_{I\!R^d} \ln\left(F_k^h(v)\right) F_k^h(v) - \ln\left(F_k^0(v)\right) F_k^0(v)\mathrm{d}v \tag{3.19}$$

where, as in the previous theorem, $F_k^t = \nabla\phi \# F_k$. Then, using McCann's differentiation formula, we obtain that the expression at the right handside of (3.19) is

$$\begin{aligned} \int_{I\!R^d} \nabla_v F_k(v;x) \cdot \left(h\tilde{\theta}_k(x) \nabla_v \left(\ln \frac{F_k}{M_{\tilde{F}_k}} \right) \right) \mathrm{d}v &= h\tilde{\theta}_k(x) \int_{I\!R^d} \left[\left|\nabla_v \ln F_k\right|^2 F_k + \nabla_v F_k \cdot \frac{v}{\tilde{\theta}_k(x)} \right] \mathrm{d}v \\ &= h\tilde{\theta}_k(x) \left[\int_{I\!R^d} \left|\nabla_v \ln F_k\right|^2 F_k \mathrm{d}v - \frac{d}{\tilde{\theta}_k(x)} \right] \\ &= h\tilde{\theta}_k(x) \int_{I\!R^d} \left|\nabla_v \ln(F_k/M_{\tilde{F}_k})\right|^2 F_k \mathrm{d}v \end{aligned} \tag{3.20}$$

We use (3.20) in (3.19), multiply both sides of the subsequent inequality by $\rho_k(x)$, and integrate over T^d to obtain the first estimate in (3.18).

Moreover, since $\theta_k(x) < \tilde{\theta}_k(x)$, we can continue the estimate for the penultimate terms as follows.

$$
\begin{aligned}
h\tilde{\theta}_k(x)\left[\int_{I\!R^d} |\nabla_v \ln F_k|^2 F_k \mathrm{d}v - \frac{d}{\tilde{\theta}_k(x)}\right] &\geq h\tilde{\theta}_k(x)\left[\int_{I\!R^d} |\nabla_v \ln F_k|^2 F_k \mathrm{d}v - \frac{d}{\theta_k(x)}\right] \\
&= h\tilde{\theta}_k(x)\int_{I\!R^d} |\nabla_v \ln(F_k/M_{F_k})|^2 F_k \mathrm{d}v
\end{aligned}
$$

which leads directly to the second estimate. ■

Finally, we need the analog of Theorem 3.1 for L^p norms. Note that $G \mapsto \|G\|_p^p$ is a functional of the type (3.17), and it is dispalcement convex for $p \geq 1$.

Theorem 3.3: (L^p bounds on f_k) *For all k, and all $1 \leq p \leq \infty$,*

$$
\int_{T^d \times I\!R^d} |f_k(x,v)|^p \mathrm{d}v\mathrm{d}x \leq (1 - d(p-1)h)^{-k} \int_{T^d \times I\!R^d} |f_0(x,v)|^p \mathrm{d}v\mathrm{d}x \tag{3.21}
$$

Note that this will give us an exponentially growing bound on the L^p norm of our solutions, uniform in h, provided the initial data is in L^p.

Proof: Again we exploit displacement convexity exactly as in the previous proof:

$$
\begin{aligned}
\int_{I\!R^d} |\tilde{F}_k|^p \mathrm{d}v - \int_{I\!R^d} |F_k|^p \mathrm{d}v &\geq h\tilde{\theta}_k(x)p(p-1)\int_{I\!R^d} (F_k)^{p-1}\nabla_v F_k \cdot \nabla_v \left(\ln \frac{F_k}{M_{F_k}}\right)\mathrm{d}v \\
&= h\tilde{\theta}_k(x)p(p-1)\int_{I\!R^d} (F_k)^{p-2}|\nabla_v F_k|^2 \mathrm{d}v + h(p-1)p\int_{I\!R^d} \nabla_v F_k^p \cdot v\mathrm{d}v \\
&\geq -hd(p-1)p\int_{I\!R^d} |F_k|^p \mathrm{d}v \ .
\end{aligned}
$$

Now multiplying both sides by $\rho_k(x)^p$ and integrating in x, we obtain that

$$
\int_{T^d \times I\!R^d} |\tilde{f}_k(x,v)|^p \mathrm{d}v\mathrm{d}x - \int_{T^d \times I\!R^d} |f_k(x,v)|^p \mathrm{d}v\mathrm{d}x \geq -hd(p-1)p\int_{T^d \times I\!R^d} |f_k(x,v)|^p \mathrm{d}v\mathrm{d}x \ .
$$

Finally, since streaming is measure preserving,

$$
\int_{T^d \times I\!R^d} |\tilde{f}_k(x,v)|^p \mathrm{d}v\mathrm{d}x = \int_{T^d \times I\!R^d} |f_{k-1}(x,v)|^p \mathrm{d}v\mathrm{d}x
$$

so that finally,

$$
\int_{T^d \times I\!R^d} |f_k(x,v)|^p \mathrm{d}v\mathrm{d}x \leq (1 - hd(p-1))^{-1} \int_{T^d \times I\!R^d} |f_{k-1}(x,v)|^p \mathrm{d}v\mathrm{d}x
$$

from which the result easily follows. ■

4 Conservation of energy in the small h limit

For given $\tilde{F}_k$, let F_k be the optimizer in (2.1), and let $\tilde{\psi}_k$ be a convex function such that $\nabla\tilde{\psi}\#F_k = \tilde{F}_k$. As in the previous section, we may assume without loss of generality that $\tilde{u}_k = u_k = 0$, making a common translation $\tilde{F}_k$ and F_k if necessary.

We know from (3.8) of the previous section that

$$\int_{I\!R^d} |v|^2 F_k \mathrm{d}v \leq \int_{I\!R^d} |v|^2 \tilde{F}_k \mathrm{d}v$$

and our goal in this section is to obtain the estimates needed to show that our evolution conserves energy in the limit as h tends to zero. Specifically, we prove:

Theorem 4.1: (Controlling energy dissipation) *Let f_0 be any initial density with*

$$\int_{T^d\times R^d} |v|^6 f_0(x,v)\mathrm{d}v\mathrm{d}x = A < \infty \ , \tag{4.1}$$

and let T be any positive number. Then there is finite constant C depending only on A and T so that

$$0 \leq \int_{T^d\times R^d} |v|^2 f_0(x,v)\mathrm{d}v\mathrm{d}x - \int_{T^d\times R^d} |v|^2 f_k(x,v)\mathrm{d}v\mathrm{d}x \leq C\left(H(f_k) - H(f_0)\right)^{3/5} h^{1/5}$$

for all k with $kh \leq T$.

Proof: The first inequality of the theorem follows from Theorem 3.1. When proving the second inequality, because $u_k = \tilde{u}_k$, by making a translation for each k, we may assume without loss of generality that $u_k = \tilde{u}_k = 0$. This will simplify a number of the expressions below.

We know from (3.7) that

$$\int_{I\!R^d} |v|^2 \tilde{F}_k(v;x)\mathrm{d}v \leq \int_{I\!R^d} |v|^2 F_k(v;x)\mathrm{d}v + h^2\tilde{\theta}_k^2(x)\int_{I\!R^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v \ . \tag{4.2}$$

Clearly, we shall need an upper bound on

$$\tilde{\theta}_k^2(x)\int_{I\!R^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v \ .$$

Henceforth in this section, C shall denote a constant depending only on T and A, but which changes from instance to instance.

By (4.1), Hölder's inequality and Theorem 3.1,

$$\int_{T^d}\left(\int_{I\!R^d} |v|^4 \tilde{F}_k(v;x)\mathrm{d}v\right)\rho_k(x)\mathrm{d}x \leq C \ . \tag{4.3}$$

for all k with $kh \leq T$.

To apply (4.3), again do the computation that led to (4.2), except with fourth powers this time. In the interest of a readable notation, we write the map $\nabla\tilde{\psi}_k$ of (2.2) as $\nabla\tilde{\psi}_k = v + R$, where,

$$R = h\tilde{\theta}_k\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right) \ .$$

We use that $\nabla\tilde{\psi}_k \# F_k = \tilde{F}_k$ to conclude that

$$\begin{aligned}\int_{I\!R^d} |v|^4 \tilde{F}_k \mathrm{d}v &= \int_{I\!R^d} \left| v + h\tilde{\theta}_k \nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^4 F_k \mathrm{d}v \\ &\geq h^4\tilde{\theta}_k^4 \int_{I\!R^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^4 F_k\mathrm{d}v + 4\int_{I\!R^d}\left(|v|^2+|R|^2\right)(v\cdot R)F_k\mathrm{d}v \ ,\end{aligned}$$

where we have neglected all manifestly positive terms on the right. Then using the arithmetic–geometric mean inequality to estimate, for example,

$$\int_{I\!R^d} (v\cdot R)|R|^2 F_k \mathrm{d}v \leq \frac{1}{4}\int_{I\!R^d} |v|^4 F_k\mathrm{d}v + \frac{3}{4}\int_{I\!R^d} |R|^4 F_k \mathrm{d}v \ ,$$

we easily conclude the existence of a universal constant K so that

$$h^4\tilde{\theta}_k^4 \int_{I\!R^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^4 \leq K\left(\int_{I\!R^d}|v|^4\tilde{F}_k\mathrm{d}v + \int_{I\!R^d}|v|^4 F_k\mathrm{d}v\right) \ ,$$

and hence that

$$\tilde{\theta}_k^2 \int_{I\!R^d} \left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^4 \leq \frac{K}{h^4}\frac{1}{\tilde{\theta}_k^2(x)}\left(\int_{I\!R^d}|v|^4\tilde{F}_k\mathrm{d}v + \int_{I\!R^d}|v|^4\tilde{F}_k\mathrm{d}v\right) \ . \tag{4.4}$$

Now multiply both sides of (4.2) by $\rho_k(x)$, and integrate over T^d. We obtain

$$\begin{aligned}&\int_{T^d\times R^d} |v|^2\left(f_{k-1}(x,v) - f_k(x,v)\right)\mathrm{d}v\mathrm{d}x \\ &\qquad \leq h^2\int_{T^d}\left(\tilde{\theta}_k^2(x)\int_{I\!R^d}\left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k\mathrm{d}v\right)\rho_k(x)\mathrm{d}x \ .\end{aligned} \tag{4.5}$$

To estimate the integral on the right of (4.5) , we must use different arguments on the regions where the temperature $\theta_k(x)$ is large and where is small. Toward this end, define

$$A_\lambda = \{\ x\in T^d \mid \tilde{\theta}_k(x) > \lambda\ \} \ .$$

Then applying Hölder's inequality with dual indices p and p',

$$\begin{aligned}&\int_{A_\lambda}\left(\tilde{\theta}_k^2(x)\int_{I\!R^d}\left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k\mathrm{d}v\right)\rho_k(x)\mathrm{d}x \leq \\ &\left(\int_{T^d}\tilde{\theta}_k(x)^{p'}\rho_k(x)\mathrm{d}x\right)^{1/p'} \|G\|_{L^p(\rho_k)}\end{aligned} \tag{4.6}$$

where $\|G\|_{L^p(\rho_k)}$ is given by

$$\|G\|_{L^p(\rho_k)} = \left(\int_{A_\lambda}\left(\tilde{\theta}_k(x)\int_{I\!R^d}\left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k\mathrm{d}v\right)^p\rho_k(x)\mathrm{d}x\right)^{1/p} \ .$$

It follows from Theorem 3.2 that

$$\|G\|_{L^1(\rho_k)} \leq \frac{1}{h}\left(H(f_{k-1}) - H(f_k)\right) \ . \tag{4.7}$$

It follows from (4.4) that

$$\|G\|_{L^2(\rho_k)} \le \frac{C}{h^2\lambda} \ . \tag{4.8}$$

Then for $0 < s < 1$, let $p = 1+s$, so that by Hölder's inequlaity

$$\|G\|_{L^p(\rho_k)} \le \|G\|_{L^2(\rho_k)}^{2s/(1+s)} \|G\|_{L^1(\rho_k)}^{(1-s)/(1+s)} \ . \tag{4.9}$$

Combining (4.6), (4.7), (4.8) and (4.9),

$$\|G\|_{L^p(\rho_k)} \le C\left(\frac{1}{h^2\lambda}\right)^{2s/(1+s)} \left(\frac{H(f_{k-1}) - H(f_k)}{h}\right)^{(1-s)/(1+s)} \ .$$

Then since $p = 1+s$ so that $p' = (1+s)/s$, as long as $s \ge 1/2$,by Theorem 3.3 and (4.1)

$$\left(\int_{T^d} \tilde{\theta}_k(x)^{p'} \rho_k(x)\mathrm{d}x\right)^{1/p'} \le C$$

for all k with $kh \le T$. Altogether then, for $1/2 \le< s < 1$,

$$\begin{aligned}&\int_{A_\lambda}\left(\theta_k^2(x)\int_{I\!R^d}\left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v\right)\rho_k(x)\mathrm{d}x \le \\ &C\left(\frac{1}{h^2\lambda}\right)^{2s/(1+s)}\left(\frac{H(f_{k-1}) - H(f_k)}{h}\right)^{(1-s)/(1+s)}\end{aligned} \tag{4.10}$$

for all k with $kh \le T$. Here we have also used (4.6).

On the other hand, we have directly from Theorem 3.2 that

$$\int_{A_\lambda^c}\left(\theta_k^2(x)\int_{I\!R^d}\left|\nabla\left(\ln F_k - \ln M_{\tilde{F}_k}\right)\right|^2 F_k \mathrm{d}v\right)\rho_k(x)\mathrm{d}x \le \lambda\left(\frac{H(f_{k-1}) - H(f_k)}{h}\right) \ . \tag{4.11}$$

Now combining (4.5), (4.10) and (4.11), and optimizing over λ, one obtains

$$\begin{aligned}\int_{T^d\times R^d} |v|^2\left(f_{k-1}(x,v) - f_k(x,v)\right)\mathrm{d}v\mathrm{d}x &\le h^2 C h^{-4s/(1+3s)}\left(\frac{H(f_{k-1}) - H(f_k)}{h}\right)^{(1+s)/(1+3s)} \\ &\le C h^{(1+s)/(1+3s)}\left(H(f_{k-1}) - H(f_k)\right)^{(1+s)/(1+3s)} \ .\end{aligned}$$

With N denoting the first integer greater than T/h, one more applications of Hölder's inequality yields

$$\begin{aligned}\int_{T^d\times R^d} |v|^2\left(f_0(x,v) - f_k(x,v)\right)\mathrm{d}v\mathrm{d}x &\le C h^{(1+s)/(1+3s)}\sum_{j=1}^{N}\left(H(f_{k-1}) - H(f_k)\right)^{(1+s)/(1+3s)} \\ &\le C h^{(1+s)/(1+3s)}\left(H(f_{N-1}) - H(f_0)\right)^{(1+s)/(1+3s)} N^{2s/(1+3s)} \ .\end{aligned}$$

Choosing $s = 1/2$, we obtain the result. ■

5 Hölder continuity of approximate solutions

The main purpose of this section is to obtain bounds on

$$\sum_{k=0}^{N-1}\int_{T^d}\rho_k(x)W_2^2(\tilde{F}_k,F_k)\mathrm{d}x \quad \text{and} \quad \sum_{k=0}^{N-1}\int_{T^d}\rho_k(x)\frac{W_2^2(F_k,\tilde{F}_k)}{\theta_k(x)}\mathrm{d}x \tag{5.1}$$

that vanish as h, the time step, tends to 0, as well as several related sums.

This will be used for showing that we obtain a solution of (1.7) in the limit as h tends to 0, and for obtaining an *a–priori* estimate on the Hölder continuity of this solution. The first quantity on the right in (5.1) is estimated as in [20], with some complications due to fact we are working on the phase space and the energy is not conserved.

Lemma 5.1: For all N,

$$\sum_{k=0}^{N-1}\int_{T^d}\rho_k(x)\left[\frac{W_2^2(\tilde{F}_k,F_k)}{\theta_k(x)}\right]\mathrm{d}x \le h2\left(H(f_N)-H(f_0)\right) \ . \tag{5.2}$$

Proof: For each x, use $\tilde{F}_k(\cdot;x)$ as a trial function in our variational problem. We then clearly obtain

$$h\int_{T^d}\rho_k(x)H(\tilde{F}_k|M_{\tilde{F}_k}) \ge \int_{T^d}\rho_k(x)\left[\frac{W_2^2(\tilde{F}_k,F_k)}{\theta_k(x)}+hH(F_k|M_{\tilde{F}_k})\right]\mathrm{d}x \ . \tag{5.3}$$

This gives us

$$\begin{aligned}\int_{T^d}\rho_k(x)\left[\frac{W_2^2(\tilde{F}_k,F_k)}{\theta_k(x)}\right]\mathrm{d}x &\le h\left[\int_{T^d}\rho_k(x)\left(H(\tilde{F}_k|M_{\tilde{F}_k})-H(F_k|M_{\tilde{F}_k})\right)\right]\mathrm{d}x\\ &= h\int_{T^d}\rho_k(x)\left[\left(H(F_k)-H(\tilde{F}_k)\right)+\int_{I\!R^d}\frac{|v|^2}{2\tilde{\theta}_k}(\tilde{F}_k-F_k)\mathrm{d}v\right]\mathrm{d}x\end{aligned}$$

Note that by (3.7),

$$\int_{I\!R^d}\frac{|v|^2}{2\tilde{\theta}_k}(\tilde{F}_k-F_k)\mathrm{d}v \le h^2\tilde{\theta}_k(x)\int_{I\!R^d}\left|\nabla\left(\ln F_k-\ln M_{\tilde{F}_k}\right)\right|^2F_k\mathrm{d}v \ .$$

Conveniently, here the right hand side is linear in $\tilde{\theta}_k$, making the estimation much more direct than in section 4. By Theorem 3.2, this becomes

$$\int_{T^d}\rho_k(x)\left[\int_{I\!R^d}\frac{|v|^2}{2\tilde{\theta}_k}(\tilde{F}_k-F_k)\right]\mathrm{d}v \le h\left(H(f_k)-H(f_{k-1})\right).$$

Altogether, this gives us

$$\int_{T^d}\rho_k(x)\left[\frac{W_2^2(\tilde{F}_k,F_k)}{\theta_k(x)}\right]\mathrm{d}x \le h\left(H(f_k)-H(f_{k-1})\right)+h\int_{T^d}\rho_k(x)\left(H(F_k)-H(\tilde{F}_k)\right)\mathrm{d}x \ .$$

Also, since

$$H(f_k) = -\int_{T^d}\rho_k\ln\rho_k\mathrm{d}x+\int_{T^d}\rho_k(x)H(F_k)\mathrm{d}x$$

and

$$H(\tilde{f}_k) = H(f_{k-1}) \ , \tag{5.4}$$

one obtains

$$\int_{T^d} \rho_k(x) \left[\frac{W_2^2(\tilde{F}_k, F_k)}{\theta_k(x)} \right] \mathrm{d}x \le 2h \left(H(f_k) - H(f_{k-1}) \right) \ ,$$

and Lemma 5.1 follows upon telescoping the sum on the right. ■

We now turn to the left member in (5.1):

$$\sum_{k=0}^{N-1} \int_{T^d} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \ . \tag{5.5}$$

Clearly, large values of $\theta_k(x)$ pose a problem if we try to estimate (5.5) in terms of the left had side of (5.2). To deal with this problem, we require moment estimates on $\theta_k(x)$.

Lemma 5.2: *Suppose that for some $r > 1$ and all $k \le N$,*

$$\int_{T^d} \theta_k^r(x) \rho_k(x) \mathrm{d}x \le C \ . \tag{5.6}$$

Then there is a constant K depending only on C so that

$$\int_{T^d} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \le K \left(\int_{T^d} \rho_k(x) \left[\frac{W_2^2(\tilde{F}_k, F_k)}{\theta_k(x)} \right] \mathrm{d}x \right)^{(r-1)/r} \tag{5.7}$$

for all $k \le N$. Moreover, with the same constant K,

$$\sum_{k=0}^{N-1} \int_{T^d} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \le K N^{1/r} \left(2h[H(f_{N-1}) - H(f_0)] \right)^{(r-1)/r} \ . \tag{5.8}$$

Proof: For any $a > 0$,

$$\begin{aligned}
\int_{T^d} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \le & \int_{\theta_k < a} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \\
& + \int_{\theta_k \ge a} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \\
\le & \int_{\theta_k < a} \frac{a}{\theta_k(x)} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x + \int_{\theta \ge a} \left(\frac{\theta_k(x)}{a} \right)^{r-1} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x \\
\le & \int_{T^d} \frac{a}{\theta_k(x)} \rho_k(x) W_2^2(\tilde{F}_k, F_k) \mathrm{d}x + \int_{T^d} \frac{2d\theta_k^r(x)}{a^{r-1}} \rho_k(x) \mathrm{d}x \ ,
\end{aligned}$$

using in the last line the fact that $W_2^2(\tilde{F}_k, F_k) \le 2d\theta_k(x)$. One then optimizes this by chosing

$$a^r = 2dC(r-1) \left[\int_{T^d} \rho_k(x) \left[\frac{W_2^2(\tilde{F}_k, F_k)}{\theta_k(x)} \right] \mathrm{d}x \right]^{-1} \ ,$$

and obtains (5.7).

Next, by this and Hölder's inequality,

$$\begin{aligned}
\sum_{k=0}^{N-1} \int_{T^d} \rho_k(x) W_2^2(F_k, \tilde{F}_k) \mathrm{d}x &\le \sum_{k=0}^{N-1} K \left(\int_{T^d} \rho_k(x) \left[\frac{W_2^2(F_k, \tilde{F}_k)}{\theta_k(x)} \right] \mathrm{d}x \right)^{r-1/r} \\
&\le K N^{1/r} \left(\sum_{k=0}^{N-1} \int_{T^d} \rho_k(x) \left[\frac{W_2^2(F_k, \tilde{F}_k)}{\theta_k(x)} \right] \mathrm{d}x \right)^{(r-1)/r} \\
&\le K N^{1/r} \left(h2[H(f_{N-1}) - H(f_0)] \right)^{(r-1)/r} \ ,
\end{aligned} \tag{5.9}$$

where in the last line we have used (5.2). ■

These lemmas may be applied to prove an estimate on the convergence of our approximate solutions:

Theorem 5.3: (Approximate transport equation: part II) *Let f_k be a sequence of probability densities given by the maximize and flow algorithm, with $p = 1$, for a given time step $h > 0$. Let $f^h(t, x, v)$ be defined by (2.11). For any integer N, let $T = Nh$. Suppose moreover that (5.6) holds for $r > 2$. Then for any test function ϕ on $I\!R_+ \times T^d \times I\!R^d$ such that $\phi(t, \cdot, \cdot) = 0$ for $t = 0$ and $t = T$, there is a constant C_T depending on T, ϕ and r, but not on h, so that*

$$\begin{aligned}
&\left| \int_{T^d \times I\!R^d} \left[\int_0^T f^h \left(\frac{\partial \phi}{\partial t} + v \cdot \nabla_v \phi \right) \mathrm{d}t + h \sum_{k=0}^{N-1} (\mathcal{L}_{f_k} \phi)) f_k \right] \mathrm{d}v \mathrm{d}x \right| \le \\
&C_T \left(h^{1-\frac{1}{r}} + h^{\frac{1}{5}} \right) .
\end{aligned} \tag{5.10}$$

Proof: The starting point is Theorem 2.2. In light of (3.8), the middle term on the right side of (2.14) is bounded by

$$\|D_v^2 \phi\|_\infty \left(H(f_{N-1}) - H(f_0) \right) .$$

We use a classical inequality that compare the entropy and the kinetic energy of any density function f:

$$H(f) \le 2\pi E(f). \tag{5.11}$$

We apply (5.11) to f_{N-1} and use Theorem 4.1 to prove that the middle term on the right side of (2.14) is bounded by $C_T h^{\frac{1}{5}}$. By Lemma 2.3, the final term on the right side of (2.14) is bounded by

$$\|D_v^2 \phi\|_\infty \sum_{k=0}^{N-1} \int_{T^d} \rho_k(x) W_2^2(F_k, \tilde{F}_k) \mathrm{d}x \ . \tag{5.12}$$

We use (5.11) in the last inequality of Lemma 5.2, use that $T = Nh$, to control the final term on the right side of (2.14), by $C_T h^{1-\frac{1}{r}}$. ■

The same sort of reasoning leads to an *a–priori* Hölder continuity result for f^h in terms of the Wasserstein 2–metric, W_2.

Theorem 5.4 (Hölder Continuity in time): *For all s and t with $0 < s < t < T = Nh$,*

$$W_2(f^h(t, \cdot, \cdot), f^h(s, \cdot, \cdot)) \le \sqrt{t - s + h} \sqrt{E(f_0)(H(M_\infty) - H(f_0))} + (t - s + 3h) \sqrt{E(f_0)} \ , \tag{5.13}$$

Where M_∞ is the Maxwellian density with spatially homogenous ρ, u and θ, and with the same total energy and momentum as f_0.

Proof: For any times $t > s > 0$. Let $M = [s]$, the integer part of s, and let $N = [t]$, so that

$$\begin{aligned} &W_2(f^h(t,\cdot,\cdot), f^h(s,\cdot,\cdot)) \le \\ &W_2(f^h(t,\cdot,\cdot), f_N(\cdot,\cdot)) + W_2(f_N(\cdot,\cdot), f_M(\cdot,\cdot)) + W_2(f_M(\cdot,\cdot), f^h(s,\cdot,\cdot)) \ . \end{aligned} \tag{5.14}$$

Let $L : T^d \times I\!R^d \to T^d \times I\!R^d$ be given by $L(x,v) = (x + (t - t_N)v, v)$. Then $L\#f_N = f^h(t,\cdot,\cdot)$, and hence

$$\begin{aligned} &W_2^2(f^h(t,\cdot,\cdot), f_N(\cdot,\cdot)) \le \frac{1}{2}\int_{T^d\times I\!R^d} |L(x,v) - (x,v)|^2 f_N(x,v)\mathrm{d}x\mathrm{d}v \le \\ &h^2\frac{1}{2}\int_{T^d\times I\!R^d} |v|^2 f_N(x,v)\mathrm{d}x\mathrm{d}v = h^2 E(f_0) \ , \end{aligned} \tag{5.15}$$

and hence

$$W_2(f^h(t,\cdot,\cdot), f_N(\cdot,\cdot)) \le W_2(f^h(s,\cdot,\cdot), f_M(\cdot,\cdot)) \le h\sqrt{E(f_0)} \ . \tag{5.16}$$

By the same argument,

$$W_2(f^h(s,\cdot,\cdot), f_M(\cdot,\cdot)) \le h\sqrt{E(f_0)} \ . \tag{5.17}$$

Next, for each $k = M+1, \ldots, N$,

$$W_2(f_{k-1}, f_k) \le W_2(f_{k-1}, \tilde f_k) + W_2(\tilde f_k, f_k)$$

and by the same reasoning as above once again,

$$W_2(f_{k-1}, \tilde f_k) \le h\sqrt{E(f_0)} \ . \tag{5.18}$$

Then since f_k and $\tilde f_k$ have the same spatial density ρ_k,

$$\begin{aligned} W_2(f_k, \tilde f_k) &\le \int_{T^d} W_2(F_k, \tilde F_k)\rho_k(x)\mathrm{d}x \\ &\le \int_{T^d} \theta_k^{1/2}(x)\left[\frac{W_2^2(f_k, \tilde f_k)}{\theta_k(x)}\right]^{1/2} \rho_k(x)\mathrm{d}x \\ &\le \left(\int_{T^d} \theta_k(x)\rho_k(x)\mathrm{d}x\right)^{1/2}\left(\int_{T^d}\left[\frac{W_2^2(f_k, \tilde f_k)}{\theta_k(x)}\right]\rho_k(x)\mathrm{d}x\right)^{1/2} \ . \end{aligned} \tag{5.19}$$

Hence, combining (5.18) and (5.19), summing, using Lemma 5.1, and the fact that $\frac{d}{2}\int_{T^d}\rho_k\theta_k dx$ is bounded by the initial energy $E(f_0)$ we conclude that

$$\begin{aligned} \sum_{k=M}^{N-1} W_2(f_k, \tilde f_k) &\le \sqrt{(N-M)E(f_0)}\sqrt{h(H(f_N) - H(f_M))} + (N-M)h\sqrt{E(f_0)} \\ &\le \sqrt{t - s + h}\sqrt{E(f_0)(H(f_N) - H(f_M))} + (t - s + h)\sqrt{E(f_0)} \ . \end{aligned} \tag{5.20}$$

Since for any $M < N$, $H(f_N) - H(f_M) \le H(M_\infty) - H(f_0)$, the result follows readily from (5.16), (5.17) and (5.20). ■

6 Velocity averaging lemma

In this section we prove a discrete time version of the velocity averaging lemma. Our starting point is (2.13), which, leaving out the intermediate steps is

$$
\begin{aligned}
&\int_0^T \int_{T^d\times I\!R^d} f^h \left(\frac{\partial \phi}{\partial t} + v\cdot\nabla_v \phi\right) \mathrm{d}x\mathrm{d}v\mathrm{d}t \\
&= -\sum_{k=1}^{N-2} \int_{T^d\times I\!R^d} \Big(f_k(x,v) - \tilde{f}_k(x,v)\Big)\, \phi(t_k,x,v)\mathrm{d}x\mathrm{d}v \\
&\quad + \int_{T^d\times I\!R^d} \tilde{f}_N(x,v)\phi(T,x,v)\mathrm{d}x\mathrm{d}v - \int_{T^d\times I\!R^d} \tilde{f}_0(x,v)\phi(0,x,v)\mathrm{d}x\mathrm{d}v\ .
\end{aligned}
\tag{6.1}
$$

The regularity properties that we investigate here are local, and it suffices to consider test functions that vanish for t outside $(0,T)$, so that we may neglect the boundary terms.

To begin, define the distribution κ on $T^d \times I\!R^d \times (0,T)$ by

$$
\kappa(x,v,t) = \sum_{k=1}^{N-2} \Big(f_k(x,v) - \tilde{f}_k(x,v)\Big)\, \delta(t-t_k)\ .
$$

Consider (x,t) as a variable in R^{d+1}, and introduce the $d{+}1$ dimensional vector $a(v) = (v_1, v_2, \ldots, v_d, 1)$ in terms of which we may rewrite (6.1) as

$$
\mathrm{div}_{(x,t)}\left(a(v)f^h(x,v,t)\right) = \kappa(x,v,t)\ . \tag{6.2}
$$

Next fix any numbers β and m with $0<\beta<1$, and m an even positive integer. Define the function g by

$$
g(x,v,t) = (1-\Delta_{(x,t)})^{-\beta/2}(1-\Delta_v)^{-m/2}\kappa(x,v,t)\ . \tag{6.3}
$$

Then we can rewrite (6.2) as

$$
\mathrm{div}_{(x,t)}\left(a(v)f^h(x,v,t)\right) = (1-\Delta_z)^{\beta/2}(1-\Delta_v)^{m/2}g(x,v,t)\ . \tag{6.4}
$$

Now let ψ be a smooth and compactly supported function on R^d, and define

$$
\rho^h_\psi(x,t) = \int_{I\!R^d} f^h(x,v,t)\psi(v)\mathrm{d}v\ . \tag{6.5}
$$

A theorem due to DiPerna, Lions, Meyer [15] and Bézard [2], presented in this form in [3], Theorem 1.6, asserts that for f and g satisfying (6.4), and both belonging to L^p, some $1<p\le 2$, there is a constant C so that

$$
\|(1-\Delta_{(x,t)})^{s/2}\rho^h_\psi\|_{L^p(T^d\times[0,T])} \le C\left(\|f^h\|_{L^p(T^d\times I\!R^d\times[0,T])} + \|g\|_{L^p(T^d\times I\!R^d\times[0,T])}\right)\ , \tag{6.6}
$$

where

$$
s = \frac{1-\beta}{(m+1)p'}
$$

and C is independent of f and g. (As stated in [3], the theorem requires that $\partial^\alpha a$ is bounded for all $|\alpha|\le m$. However, inspection of the proof reveals that a uniform bound on $\partial^\alpha a$ is required only on the support on ψ, and this holds in our case.)

To apply this in the present setting, we first prove the following lemma:

Lemma 6.1: *For any p with $1 < p < \infty$, let β be a number with $(p-1)/p < \beta < 1$, and let $h(x)$ belong to $L^p(\mathbb{R}^d)$. Define $g(x,t)$ by*

$$g(x,t) = (1-\Delta_{x,t})^{-\beta/2}\,(h(x)\delta(t))$$

where $\delta(t)$ denotes the Dirac mass at $t = 0$. Then there is a constant C depending only on d, p' and β so that $\|g\|_p \le C\|h\|_p$.

Proof: Whether or not $\left\|(1-\Delta_{x,t})^{-\beta/2}\,(h(x)\delta(t))\right\|_{L^p(\mathbb{R}^{d+1})}$ is finite, there is a function ϕ on $\mathbb{R}^{d+1}$ with $\|\phi\|_{L^{p'}(\mathbb{R}^{d+1})} = 1$ so that

$$\begin{aligned}\left\|(1-\Delta_{x,t})^{-\beta/2}\,(h(x)\delta(t))\right\|_{L^p(\mathbb{R}^{d+1})} &= \int_{\mathbb{R}^{d+1}} \phi(x,t)(1-\Delta_{x,t})^{-\beta/2}\,(h(x)\delta(t))\,\mathrm{d}x\mathrm{d}t\\ &= \int_{\mathbb{R}^d} \left(\left((1-\Delta_{x,t})^{-\beta/2}\phi\right)(x,0)\right) h(x)\mathrm{d}x \\ &\le \left\|\tau\left((1-\Delta_{x,t})^{-\beta/2}\phi\right)\right\|_{L^{p'}(\mathbb{R}^{d+1})} \|h\|_{L^p(\mathbb{R}^{d+1})}\ ,\end{aligned} \tag{6.7}$$

where τ denotes the trace map on the hyperplane $t = 0$. A standard trace theroem for Bessel potentials (see [31], section VI.4.2, or [32]), there is a constant C depending only on d, p' and β so that for any $\beta > 1/p'$,

$$\left\|\tau\left((1-\Delta_{x,t})^{-\beta/2}\phi\right)\right\|_{p'} \le C\|\phi\|_{p'}\ .$$

This yields the result. ∎

Lemma 6.2: (Controlling moments of f^p) *For any probability density f on $T^d \times \mathbb{R}^d$, let $\rho(x) = \int_{\mathbb{R}^d} f(x,v)\mathrm{d}v$. Suppose that for some $1 < q < \infty$,*

$$\int_{T^d\times\mathbb{R}^d} (1+|v|^2)f(x,v)\mathrm{d}x\mathrm{d}v = A < \infty \qquad \text{and} \qquad \int_{T^d\times\mathbb{R}^d} f^q(x,v)\mathrm{d}x\mathrm{d}v = B < \infty\ .$$

Then for any $0 < \alpha < 1$, and $p = \frac{q}{1-\alpha+\alpha q}$, there is a constant C depending only on α, q, A and B so that

$$\int_{T^d\times\mathbb{R}^d} |(1+|v|^2)^\alpha f(x,v)|^p\mathrm{d}x\mathrm{d}v \le C\ .$$

Under the additional condition that $2\alpha p' > d$, there is a constant C of the same type so that

$$\|\rho\|_{L^p(T^d)} \le C\ . \tag{6.8}$$

Furthermore, for any non-negative function y on T^d, and any r with $1 < r < \infty$, and any s with $0 < s < 1/r$

$$\|y\rho\|_r \le \left(\int_{T^d} y^{1/s}\rho\mathrm{d}x\right)^s \left(\|\rho\|_{(r-sr)/(1-sr)}\right)^{1-s}\ .$$

Proof: Using Hölder's inequality with dual indices $r = 1/\alpha q$ and $r' = 1/(1-\alpha q)$,

$$\int_{T^d\times I\!R^d} |(1+|v|^2)^\alpha f(x,v)|^q \mathrm{d}x\mathrm{d}v = \int_{T^d\times I\!R^d} |(1+|v|^2) f(x,v)|^{\alpha q} f(x,v)^{q-\alpha q}\mathrm{d}x\mathrm{d}v$$
$$\leq A^{\alpha q} B^{1-\alpha q} \ .$$

Now for any function g on T^d,

$$\begin{aligned}\int_{T^d} g(x)\rho(x)\mathrm{d}x &= \int_{T^d\times I\!R^d} g(x) f(x,v)\mathrm{d}x\mathrm{d}v \\ &= \int_{T^d\times I\!R^d} g(x)(1+|v|^2)^{-\alpha}(1+|v|^2)^\alpha f(x,v)\mathrm{d}x\mathrm{d}v \\ &\leq \left(\int_{T^d\times I\!R^d} |g(x)(1+|v|^2)^{-\alpha}|^{p'}\mathrm{d}x\mathrm{d}v\right)^{1/p'} \left(\int_{T^d\times I\!R^d} |(1+|v|^2)^\alpha f(x,v)|^p \mathrm{d}x\mathrm{d}v\right)^{1/p}\end{aligned}$$

As long as $2\alpha p' > d$, the first integral on the right is bounded by a constant multiple of $\|g\|_{L^{p'}(T^d)}$, and this proves (6.8)

Finally, for any non-negative function y on T^d,

$$\int_{T^d} y^r\rho^r\mathrm{d}x = \int_{T^d} y^r \rho^{sr}\rho^{r-sr}\mathrm{d}x \leq \left(\int_{T^d} y^{1/s}\rho\mathrm{d}x\right)^{sr}\left(\int_{T^d}\rho^{(r-sr)/(1-sr)}\mathrm{d}x\right)^{1-sr}$$

where we have used Hölder's inequality with exponents $1/sr$ and $1/(1-sr)$. This yields the result. ■

Lemma 6.3: (Controlling the transport term $\mathrm{div}_{(x,t)}[af^h]$) *For any p with $1 < p < \frac{4+d}{2+d}$, any $T > 0$, and any $m > 2 + 2d/p'$ even integer, there exists a constant C depending only on d, T, $\int_{T^d\times I\!R^d} |v|^6 f_0\mathrm{d}x\mathrm{d}v$, and $\|f_0\|_2$ so that uniformly in N and h with $Nh < T$,*

$$\sum_{k=1}^{N-1} \|(1-\Delta_v)^{-m/2}(\tilde{f}_k - f_k)\|_{L^p(T^d\times I\!R^d)} \leq C \ .$$

Proof: As in the proof of Lemma 6.1,there is a function ϕ on $T^d\times I\!R^d$ with $\|\phi\|_{L^{p'}(T^d\times I\!R^d)} = 1$ so that

$$\|(1-\Delta_v)^{-m/2}(\tilde{f}_k - f_k)\|_p = \int_{I\!R^{2d}} (1-\Delta_v)^{-m/2}\phi(x,v)(\tilde{f}_k(x,v) - f_k(x,v))\mathrm{d}x\mathrm{d}v \ . \tag{6.9}$$

Let $\psi(x,v)$ be given by

$$\psi(x,v) = (1-\Delta_v)^{-m/2}\phi(x,v) = (1-\Delta_v)^{-1}(1-\Delta_v)^{-(m-2)/2}\phi(x,v) \ .$$

Choose m large enough that $m-2 > d/p'$. Then for each x, $(1-\Delta_v)^{-(m-2)/2}\phi(x,\cdot)$ is uniformly bounded and Hölder continuous with the bound depending only on $\|\phi(x,\cdot)\|_{L^{p'}(I\!R^d)}$. This follows from properties of Bessel potentials; see the appendix for details. This in turn implies [23 Theorem 10.3] that the second derivative of $(1-\Delta_v)^{-1}(1-\Delta_v)^{-(m-2)/2}\phi$ are uniformly bounded and Hölder continuous. Thus,

$$\|\psi(x,\cdot)\|_\infty + \|\nabla_v\psi(x,\cdot)\|_\infty + \|D_v^2\psi(x,\cdot)\|_\infty < C\|\phi(x,\cdot)\|_{p'} \tag{6.10}$$

with C finite and depending only on the dimension, p and the choice of m.

Then, by (2.10), and Lemma 2.3,

$$\begin{aligned}
\left|\int_{\mathbb{R}^{2d}}\left((1-\Delta_v)^{-m/2}\phi\right)(\tilde{f}_k-f_k)\mathrm{d}x\mathrm{d}v\right| &= \left|\int_{T^d\times\mathbb{R}^d}(f_k-\tilde{f}_k)\psi\mathrm{d}v\mathrm{d}x\right| \\
&\le h\left|\int_{T^d\times\mathbb{R}^d}\left(\mathcal{L}_{\tilde{f}_k}\psi\right)f_k\mathrm{d}v\mathrm{d}x\right| \\
&+ \int_{T^d}\|D_v^2\psi(x,\cdot)\|_\infty\rho_k(x)W_2^2(F_k,\tilde{F}_k)\mathrm{d}x \\
&\le h\left|\int_{T^d\times\mathbb{R}^d}\left(\mathcal{L}_{\tilde{f}_k}\psi\right)f_k\mathrm{d}v\mathrm{d}x\right| \\
&+ C\int_{T^d}\|\phi(x,\cdot)\|_{p'}\rho_k(x)W_2^2(F_k,\tilde{F}_k)\mathrm{d}x \\
&\le h\left|\int_{T^d\times\mathbb{R}^d}\left(\mathcal{L}_{\tilde{f}_k}\psi\right)f_k\mathrm{d}v\mathrm{d}x\right| \\
&+ C\|\phi\|_{L^{p'}(T^d\times\mathbb{R}^d)}\left(\int_{T^d}\left(\rho_k(x)W_2^2(F_k,\tilde{F}_k)\right)^p\mathrm{d}x\right)^{1/p} .
\end{aligned} \tag{6.11}$$

Here, because $u_k=\tilde{u}_k$, $\mathcal{L}_{\tilde{f}_k}$ is $\mathcal{L}_{f_k}$ with $\theta_{f_k}\Delta_v$ replaced by $\tilde{\theta}_{f_k}\Delta_v$.

Next we estimate the last term in (6.11). Fix k, and let $e(x)$ denote $\int_{\mathbb{R}^d}|v|^2\tilde{F}_k(v;x)\mathrm{d}v$, which is essentially the local energy density at the kth time step. Then since $\int_{\mathbb{R}^d}|v|^2F_k(v;x)\mathrm{d}v\le\int_{\mathbb{R}^d}|v|^2\tilde{F}_k(v;x)\mathrm{d}v$,

$$W_2^2(F_k,\tilde{F}_k)\le e(x)$$

for each x. Now choose any $\bar{s}$ with $0<\bar{s}<1/p$. Then

$$\begin{aligned}
\int_{T^d}\left(\rho_k(x)W_2^2(F_k,\tilde{F}_k)\right)^p\mathrm{d}x &\le \int_{T^d}\left(\rho_k(x)W_2^2(F_k,\tilde{F}_k)\right)^{\bar{s}p}(\rho_k(x)e(x))^{(1-\bar{s})p}\mathrm{d}x \\
&\le \left(\int_{T^d}\rho_k(x)W_2^2(F_k,\tilde{F}_k)\mathrm{d}x\right)^{\bar{s}p}\left(\int_{T^d}(\rho_k(x)e(x))^r\mathrm{d}x\right)^{1-\bar{s}p}
\end{aligned} \tag{6.12}$$

where $r=p(1-\bar{s})/(1-\bar{s}p)$. Recall that by Theorem 3.1, the sixth moment at time $t<T$ is controlled by the initial sixth moment, and so, $\int_{T^d}e(x)^3\rho_k(x)\mathrm{d}x$ is controlled by the initial sixth moment of the velocity uniformly in k for $hk<T$. Therefore, for such values of k, there is a constant C depending only on T and the initial sixth moment so that if we apply Lemma 6.2 with $s=1/3$ and $r\in(0,3)$, we obtain

$$\left(\int_{T^d}(\rho_k(x)e(x))^r\mathrm{d}x\right)^{1/r}\le C\left(\|\rho_k\|_{\frac{2r}{3-r}}\right)^{2/3} . \tag{6.13}$$

Now take $s=1/2$ in (6.12) so that $r=p/(2-p)$. Then $2r/(3-r)=p/(3-2p)$. By Theorem 3.3 and Lemma 6.2, $\|\rho_k\|_{\frac{2r}{3-r}}$ is bounded uniformly in k for all $kh<T$ by a constant C depending only on T and $\|f_0\|_2$. Combining (6.12) and (6.13) we obtain

$$\left(\int_{T^d}\left(\rho_k(x)W_2^2(F_k,\tilde{F}_k)\right)^p\mathrm{d}x\right)^{1/p}\le C\int_{T^d}\left(\rho_k(x)W_2^2(F_k,\tilde{F}_k)\right)^{1/2} \tag{6.14}$$

uniformly in k with $kh < T$ for a constant C depending only on T, the initial sixth moment and $\|f_0\|_2$, which completes our estimation of the last term in (6.11).

We now estimate the penultimate term. Note that

$$\left|\int_{I\!R^d} \mathcal{L}_{\tilde{f}_k}(\psi(x,\cdot))dv\right| \leq \tilde{\theta}_k(x)\|D_v^2\psi(x,\cdot)\|_\infty + \|\psi(x,\cdot)\|_\infty + (1+d\theta_k)\|\nabla_v\psi(x,\cdot)\|_\infty .$$

This, combined with (6.10) yields that there exists a constant C depending only on d and p such that,

$$\begin{aligned}\left|\int_{T^d\times I\!R^d}\left(\mathcal{L}_{\tilde{f}_k}\psi\right) f_k \mathrm{d}v\mathrm{d}x\right| &\leq C\int_{T^d}\|\phi(x,\cdot)\|_{p'}(1+e(x))\rho_k(x)\mathrm{d}x\\ &\leq \|\phi\|_{p'}\|(1+e(\cdot))\rho_k(\cdot)\|_p \ .\end{aligned}$$

Applying Lemma 6.2 with $s = 1/3$, we obtain that

$$\left|\int_{T^d\times I\!R^d}\left(\mathcal{L}_{\tilde{f}_k}\psi\right) f_k \mathrm{d}v\mathrm{d}x\right| \leq C \tag{6.15}$$

uniformly in k with $kh < T$ for a constant C depending only on T, the initial sixth moment and $\|f_0\|_2$, which completes our estimation of the penultimate term in (6.11). Combining (6.9), (6.11), (6.14) and (6.15), we have

$$\|(-\Delta_v)^{-m/2}(\tilde{f}_k - f_k)\|_p \leq C\left[h + \int_{T^d}\left(\rho_k(x)W_2^2(F_k,\tilde{F}_k)\right)^{1/2}\right] \ . \tag{6.16}$$

By Lemma 5.2 with $r = 2$, there is a constant K so that

$$\sum_{k=1}^{N-2}\int_{T^d}\rho_k(x)W_2^2(F_k,\tilde{F}_k)\mathrm{d}x \leq K\left(Nh2(H(f_N)-H(f_0))\right)^{1/2} \ .$$

Because of the standard inequality (5.11), this quantity is plainly controlled by the energy of f_0. Hence we may sum the right hand side of (6.16) over k with $kh < T$, and we obtain a bound that is independent of h. ■

Theorem 6.4: (Strong compactness of averaging functions) *Let ψ be any compactly supported function on $I\!R^d$, and define*

$$\rho_\psi^h(x,t) = \int_{I\!R^d} f^h(x,v,t)\psi(v)\mathrm{d}v \ .$$

Then for any p with $1 < p < \frac{3+d/2}{2+d/2}$, and any $T > 0$, there is a constant C depending only on d, T, $\int_{T^d\times I\!R^d}|v|^6 f_0\mathrm{d}x\mathrm{d}v$, $\|f_0\|_2$ and ψ so that uniformly in N and h with $Nh < T$, then there is an $s > 0$ with

$$\|(1-\Delta_{x,t})^{s/2}\rho_\psi^h\|_p \leq C \ .$$

(In fact, $s \in (0,1)$ is a constant depending only on d, T, $\int_{T^d\times I\!R^d}|v|^6 f_0\mathrm{d}x\mathrm{d}v$, $\|f_0\|_2$, p, and ψ.)

Proof: We simply need to apply the lemmas to verify (6.6) in the cited theorem of DiPerna-Lions-Meyer-Bézard. First, we know that $\|f(\cdot,\cdot,t)\|_p$ is bounded uniformly on $[0,T]$ in terms of $\|f_0\|_p$, and hence we have a bound on $\|f\|_{L^p(T^d\times I\!R^d\times[0,T])}$. As for g, Lemma 6.1 says that there is a constant C so that

$$\|f\|_{L^p(T^d\times I\!R^d\times[0,T])} \le \sum_{k=1}^{N-1} \|(1-\Delta_v)^{-m/2}(\tilde{f}_k - f_k)\|_p \ .$$

Note that the dependence on v is suppressed in Lemma 6.1, but

$$\|g\|_{L^p(T^d\times I\!R^d\times[0,T])} = \left(\int_{I\!R^d} \left(\|g(\cdot,v,t)\|_{L^p(T^d\times[0,T])}\right)^p\right)^{1/p} \ .$$

Now Lemma 6.3 give us the desired bound on $\|g\|_{L^p(T^d\times I\!R^d\times[0,T])}$. ■

As we shall show in the next section, this gives us the compactness we need to show that the when we let h tend to zero, the limit of the temperatures is the temperature of the limit. With that in hand, the rest follows directly.

7 Existence of solutions

In this section we prove that a sequence of our discrete time approximations f^h converges to a solution f of the kinetic Fokker–Planck equation.

Theorem 7.1(Main result: existence of a solution) *Let f_0 be a probability density on $T^d\times I\!R^d$ such that*

$$\int_{T^d\times I\!R^d} |v|^6 f_0(x,v)\mathrm{d}x\mathrm{d}v < \infty \ ,$$

$$\int_{T^d\times I\!R^d} |f_0|^2(x,v)\mathrm{d}x\mathrm{d}v < \infty \ .$$

Then for any $T>0$. Then there is a sequence of time steps h_k with $\lim_{k\to\infty} h_k = 0$ so that $\lim_{k\to\infty} f^{h_k} = f$ weakly in $L^1(T^d\times I\!R^d\times[0,T])$ and where f is a weak solution of the kinetic Fokker–Planck equation (1.7), and

$$\lim_{t\to 0} f(\cdot,\cdot,t) = f_0(\cdot,\cdot) \ ,$$

weakly in $L^1(T^d\times I\!R^d)$. Moreover,

$$\int_{T^r\times I\!R^d} |v|^2 f(x,v,t)\mathrm{d}x\mathrm{d}v = \int_{T^r\times I\!R^d} |v|^2 f_0(x,v)\mathrm{d}x\mathrm{d}v$$

for all t and there is a constant C depending only on the energy and entropy of the initial data f_0 so that

$$W_2(f(\cdot,\cdot,t), f(\cdot,\cdot,s)) \le C(\sqrt{|t-s|} + |t-s|) \ .$$

Finally, for all $t>0$,

$$H(f(\cdot,\cdot,t) - H(f(\cdot,\cdot,0) \ge \int_0^t \int_{T^d} \theta(x,t)\left[\int_{I\!R^d} \left|\nabla_v \ln f(\cdot,\cdot,t) - \nabla_v \ln M_{f(\cdot,\cdot,t)}\right|^2 f(x,v,t)\mathrm{d}v\right]\mathrm{d}x\mathrm{d}t \ .$$

Proof: Most of the work required to prove this result has been done in the previous sections of this manuscript. In fact, because the velocity averaging lemma provides the strong compactness needed to deal with the nonlinear term $\mathcal{L}_{f_k}\phi$, Theorem 5.3, almost gives us the convergence result.

The starting point is the fact that the family of functions $\{f^h \mid h > 0\}$ is weakly compact in $L^p(T^d \times I\!R^d \times [0,T])$ for any $1 < q < 2$. This is a direct consequence of our uniform L^2 bound on the f^h that is provided by Theorem 3.3, and the uniform bound on the set of entropies $\{H(f^h)\}_{h>0}$. We may therefore select a sequence of times steps h_k tending to zero so that

$$\lim_{k\to\infty} f^{h_k} = f$$

weakly in $L^q(T^d \times I\!R^d \times [0,T])$, and hence locally in $L^1(T^d \times I\!R^d \times [0,T])$.

Now let $\rho_k(x,t) = \int_{I\!R^d} f^{h_k}(x,v,t)\mathrm{d}v$ and $\rho_k(x,t) = \int_{I\!R^d} f(x,v,t)\mathrm{d}v$. Also let

$$\rho_k(x,t)u_k(x,t) = \int_{I\!R^d} v f^{h_k}(x,v,t)\mathrm{d}v \qquad \text{and} \qquad \rho_k(x,t)\theta_k(x,t) = \frac{1}{d}\int_{I\!R^d} |v - u_k(x,t)|^2 f^{h_k}(x,v,y)\mathrm{d}v\ ,$$

and let

$$\rho(x,t)u(x,t) = \int_{I\!R^d} v f(x,v,y)\mathrm{d}v \qquad \text{and} \qquad \rho(x,t)\theta(x,t) = \frac{1}{d}\int_{I\!R^d} |v - u(x,t)|^2 f(x,v,t)\mathrm{d}v\ .$$

Note the change in notation: Here k indexes an approximate solution f^{n_k}, and not a particular time step.

We cannot directly apply Theorem 6.4 to, say, $\int_{I\!R^d} v f^{h_k}(x,v,t)\mathrm{d}v$, since this would correspond to choosing $\psi(v) = v$ in that Theorem, and ψ is required to have compact support. Consider instead $\psi(v) = \chi(v)v$ where χ is a smooth radial decreasing function with values in $[0,1]$ that equals 1 identically on the ball of radius R, and vanishes outside the ball of radius $2R$. For any $\epsilon > 0$, using the energy bound we can choose a value of R so that $\int(1-\chi(v))|v|f^h(x,v,t) < \epsilon$ uniformly in h and t. This gives us strong compactness of the $\rho_k u_k$. In a like manner, we obtain strong compactness of the ρ_k. Passing to subsequences, we may arrange that both sequences $\{\rho_k\}$ and $\{\rho_k u_k\}$ converge almost everywhere. Hence we have that $\{u_k\}$ converges almost everywhere on $\{\rho > 0\}$ to u. Hence, we may select a further subsequence that converges strongly in $L^q(T^d \times [0,T])$.

We now claim that for any test function ϕ on $T^d \times I\!R^d \times [0,T]$

$$\int_{T^d\times I\!R^d\times[0,T]} \nabla_x\phi(x,v,t)u(x,t)f(x,v,t)\mathrm{d}v\mathrm{d}x\mathrm{d}t = \lim_{k\to\infty}\int_{T^d\times I\!R^d\times[0,T]} \nabla_x\phi(x,v,t)u_k(x,t)f^{h_k}(x,v,y)\mathrm{d}x\mathrm{d}v\mathrm{d} \tag{7.1}$$

where here and in what follows, k progesses along our most recently selected subsequence.

By Egoroff's theorem, for any $\epsilon > 0$, there is a further subsequence so that u_k converges to u uniformly on a substet G_ϵ of $T^d \times [0,T]$ whose complement B_ϵ has a ρ–measure of no more than ϵ. We have that

$$\begin{aligned}
&\int_{T^d\times I\!R^d\times[0,T]} \nabla_x\phi(x,v,t)u_k(x,t)f^{h_k}(x,v,t)\mathrm{d}x\mathrm{d}v\mathrm{d}t \\
&\qquad = \int_{B_\epsilon\times I\!R^d} \nabla_x\phi(x,v,t)u_k(x,t)f^{h_k}(x,v,t)\mathrm{d}x\mathrm{d}v\mathrm{d}t \\
&\qquad + \int_{G_\epsilon\times I\!R^d} \nabla_x\phi(x,v,t)(u_k(x,t)-u(x,t))f^{h_k}(x,v,t)\mathrm{d}x\mathrm{d}v\mathrm{d}t \\
&\qquad + \int_{G_\epsilon\times I\!R^d} \nabla_x\phi(x,v,t)(u(x,t))f^{h_k}(x,v,t)\mathrm{d}x\mathrm{d}v\mathrm{d}t \\
&\qquad = I_1 + I_2 + I_3\ .
\end{aligned}$$

Since we have a uniform L^r bound on $u_k(x,t)\int_{\mathbb{R}^d} f^{h_k}(x,v,t)\mathrm{d}v$ for some $r>1$, by Theorem 3.1 and Lemma 6.2, it follows from Hölder's inequality that for some constant C depending on ϕ,

$$|I_1| \le C\epsilon^{(r-1)/r} \ .$$

By the uniform convergence on G_ϵ of $\{u_k\}_{k\ge 1}$, and the weak compactness of $\{f^{h_k}\}_{k\ge 1}$,

$$\lim_{k\to\infty}\int_{G_\epsilon\times\mathbb{R}^d} \nabla_x\phi(x,v)(u_k(x,t)-u(x,t))f^{h_k}(x,v,y)\mathrm{d}x\mathrm{d}v\mathrm{d}t = 0 \ ,$$

and so I_2 tends to zero as k increases. Finally, by the weak convergence

$$\lim_{k\to\infty}\int_{G_\epsilon\times\mathbb{R}^d} \nabla_x\phi(x,v)(u(x,t))f^{h_k}(x,v,y)\mathrm{d}x\mathrm{d}v = \int_{G_\epsilon\times\mathbb{R}^d} \nabla_x\phi(x,v)(u(x,t))f(x,v,y)\mathrm{d}x\mathrm{d}v \ .$$

But in the same way we bounded I_1, we conclude that

$$\left|\int_{B_\epsilon\times\mathbb{R}^d} \nabla_x\phi(x,v)(u(x,t))f(x,v,y)\mathrm{d}x\mathrm{d}v\right| \le C\epsilon^{(r-1)/r} \ .$$

Since ϵ is arbitrary, (7.1) is established.

The same sort of argument now shows that

$$\int_{T^d\times\mathbb{R}^d} \Delta_x\phi(x,v)\theta(x,t)f(x,v,t)\mathrm{d}v\mathrm{d}x = \lim_{k\to\infty}\int_{T^d\times\mathbb{R}^d} \Delta_x\phi(x,v)\theta_k(x,t)f^{h_k}(x,v,y)\mathrm{d}x\mathrm{d}v \tag{7.2}$$

The only difference is that we must use the propagation of higher moments proved in Theroem 3.1 instead of just the energy bound. The structure of the argument is the same, though.

Next, it follows immediately from Theorem 5.3 that f satisfies

$$\lim_{k\to\infty}\left|\int_{T^d\times\mathbb{R}^d}\left[\int_0^T f^{h_k}\left(\frac{\partial\phi}{\partial t}+v\cdot\nabla_v\phi\right)\mathrm{d}t + \int_0^T f^{h_k}\left(\theta_k\Delta_v\phi+(v-u_k)\cdot\nabla_v\phi\right)\mathrm{d}t\right]\mathrm{d}v\mathrm{d}x\right| = 0$$

and

$$\begin{aligned}\lim_{k\to\infty}&\int_{T^d\times\mathbb{R}^d}\left[\int_0^T f^{h_k}\left(\frac{\partial\phi}{\partial t}+v\cdot\nabla_v\phi\right)\mathrm{d}t + \int_0^T f^{h_k}\left(\theta_k\Delta_v\phi+(v-u_k)\cdot\nabla_v\phi\right)\mathrm{d}t\right]\mathrm{d}v\mathrm{d}x\\ &=\int_{T^d\times\mathbb{R}^d}\left[\int_0^T f\left(\frac{\partial\phi}{\partial t}+v\cdot\nabla_v\phi\right)\mathrm{d}t + \int_0^T f\left(\theta\Delta_v\phi+(v-u)\cdot\nabla_v\phi\right)\mathrm{d}t\right]\mathrm{d}v\mathrm{d}x \ .\end{aligned}$$

The conclusion is that

$$\int_{T^d\times\mathbb{R}^d}\left[\int_0^T f\left(\frac{\partial\phi}{\partial t}+v\cdot\nabla_v\phi\right)\mathrm{d}t + \int_0^T f\left(\theta\Delta_v\phi+(v-u)\cdot\nabla_v\phi\right)\mathrm{d}t\right]\mathrm{d}v\mathrm{d}x = 0 \ ,$$

and hence f is a weak solution of (1.7).

The conservation of the energy follows from Theorem 4.1 and (5.11) which ensure that the discrete solutions do a better and better job of conserving energy as h tends to zero. We have also used the strong compactness of $\{\rho_k|u_k|^2\}_{k\ge 0}$ and $\{\rho_k\theta_k\}_{k\ge 0}$.

We now deal with the entropy inequality. Since the entropy is a concave function, and hence weakly upper semicontinuous, for $t > 0$,

$$\liminf_{h\to 0} H(f^h(\cdot,\cdot,t) \le H(f(\cdot,\cdot,t) \ .$$

The entropy production for f^{h_k} at time t is

$$\int_{T^d} \theta_k(x,t) \left[\int_{I\!R^d} \left|\nabla_v \ln f^{h_k}(\cdot,\cdot,t) - \nabla_v \ln M_{f^{h_k}(\cdot,\cdot,t)}\right|^2 f^{h_k}(x,v,t)\mathrm{d}v\right] \mathrm{d}x$$

which can be written

$$\int_{T^d} \theta_k(x,t)\rho_k(x,t) \left(I(F^k(\cdot,x,t)) - I(M_{F^k}(\cdot,x,t))\right) \mathrm{d}x$$

where $I(F)$ is the *Fisher information*

$$I(F) = \int_{T^d} |\nabla \ln F|^2 F \mathrm{d}v \ .$$

It is well known that

$$I(F) = sup_{V\in \mathcal{C}(I\!R^d)} \left\{ \int_{I\!R^d} VF\mathrm{d}v - \lambda(V) \right\}$$

where $\lambda(V)$ is the fundamental eigenvalue of the operator $-\Delta + V$. It follows that for any continuous and compactly supported function V on $T^d \times I\!R^d$ that if we let $\lambda(V,x)$ denote the fundamental eigenvalue of $-\Delta_v + V(x,\cdot)$, then

$$\int_{T^d} \theta_k(x,t)\rho_k(x,t)I(F^k(\cdot,x,t))\mathrm{d}x \ge \int_{T^d}\int_{I\!R^d} \left(\int_{I\!R^d} f^{h_k} V\mathrm{d}v - \lambda(V,x)\rho_k\right)\mathrm{d}x \ .$$

Using the strong convergence of $\{\theta_k\}$ and $\{\rho_k\}$, and weak convergence of $\{f^{h_k}\}$ in appropriate L^p spaces, we have

$$\begin{aligned}\liminf_{k\to\infty}\int_{T^d} \theta_k(x,t)\rho_k(x,t)I(F^k(\cdot,x,t))\mathrm{d}x &\ge \liminf_{k\to\infty}\int_{T^d\times I\!R^d} \theta_k(\int_{I\!R^d} fV\mathrm{d}v - \lambda(V,x)\rho_k)\mathrm{d}x \\ &= \int_{T^d\times I\!R^d} \theta(fV - \lambda(V,x))\mathrm{d}x\mathrm{d}v\end{aligned}$$

Now taking the sup over all such V, we get

$$\liminf_{k\to\infty}\int_{T^d} \theta_k(x,t)\rho_k(x,t)I(F^k(\cdot,x,t))\mathrm{d}x \ge \int_{T^d} \theta(x,t)\rho(x,t)I(F(\cdot,x,t))\mathrm{d}x \ .$$

Finally, since $\theta_k I(M_{F^k}) = d$, identically in k, there is nothing to estimate there. Hence,

$$\liminf_{k\to\infty}\int_{T^d} \theta_k(x,t)\rho_k(x,t)(I(F^k(\cdot,x,t)) - I(M_{F^k}))\mathrm{d}x \ge \int_{T^d} \theta(x,t)\rho(x,t)(I(F(\cdot,x,t)) - I(M_F))\mathrm{d}x \ .$$

This shows that while the entropy itself can only increase “jump upwards” in the limit, the entropy production can only “jump downwards”, and so the entropy production inequality holds. (Note that $H(f^{h_k}(\cdot,\cdot,0)) = H(f_0(\cdot,\cdot))$ independent of k).

It remains observe that $\lim_{t\to 0} f(\cdot,\cdot,t) = f_0(\cdot,\cdot)$ weakly; this follows directly from the uniform modulus of continuity in t of the approximate solutions f^h. ■

8 Convergence to equilibrium

In this section, we study the sense in which the solutions that we have constructed, tend toward the set $\mathcal{M}$ of local Maxwellian densities as t tends to infinity. Our main tools in this effort will be the Hölder continuity of the solutions in the Wasserstein metric, and Talagrand's inequality, which tells us that [34], [36]

$$W_2^2(F, M_F) \le \theta_F H(F, M_F)$$

for any probability density F on $I\!R^d$. Here, θ is $1/d$ times the variance of F, as defined in (1.4).

We have from Theorem 7.1, and the logarithmic Sobolev inequality, and then Talgrand's inequality, for any $T > 0$,

$$\begin{aligned} H(f(\cdot,\cdot,T) - H(f(\cdot,\cdot,0) &\ge \int_0^T \int_{T^d} \theta(x,t) \left[\int_{I\!R^d} \left|\nabla_v \ln f(\cdot,\cdot,t) - \nabla_v \ln M_{f(\cdot,\cdot,t)}\right|^2 f(x,v,t)\mathrm{d}v \right] \mathrm{d}x\mathrm{d}t \\ &\ge \int_0^T \int_{T^d} H\left(F(\cdot;x,t)|M_{F(\cdot;x,t)}\right) \rho(x,t)\mathrm{d}x\mathrm{d}t \\ &\ge \int_0^T \int_{T^d} \frac{W_2^2(F(\cdot;x,t), M_{F(\cdot;x,t)})}{\theta(x,t)} \rho(x,t)\mathrm{d}x\mathrm{d}t\ . \end{aligned} \tag{8.1}$$

We need to control the effect of large values of $\theta(x,t)$, and since we need to do this uniformly in time, we can only use the energy bound, and not our bounds on higher velocity moments since they are not uniform in time.

For any $a > 0$,

$$\begin{aligned} \int_{T^d} \frac{W_2^2(F(\cdot;x,t), M_{F(\cdot;x,t)})}{\theta(x,t)} \rho(x,t)\mathrm{d}x &\ge \int_{\theta\le a^2} \frac{W_2^2(F(\cdot;x,t), M_{F(\cdot;x,t)})}{\theta(x,t)} \rho(x,t)\mathrm{d}x \\ &\ge \frac{1}{a^2}\left(\int_{\theta\le a^2} W_2(F(\cdot;x,t), M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x \right)^2 \end{aligned} \tag{8.2}$$

where we have used Jensen's inequality in the last line. To proceed, we shall make use of the inequality

$$W_2^2(F, M_F) \le 2d\theta_F$$

which follows directly from the definition of the left hand side. Hence,

$$\begin{aligned} &\int_{\theta\le a^2} W_2(F(\cdot;x,t), M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x \\ &\ge \int_{T^d} W_2(F(\cdot;x,t), M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x - \int_{\theta > a^2} W_2(F(\cdot;x,t), M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x \\ &\ge \int_{T^d} W_2(F(\cdot;x,t), M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x - \int_{\theta > a^2} \sqrt{2d\theta(x,t)}\rho(x,t)\mathrm{d}x \\ &\ge \int_{T^d} W_2(F(\cdot;x,t), M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x - \frac{\sqrt{2d}}{a}\left(\int_{T^d} \theta(x,t)\rho(x,t)\mathrm{d}x \right) \end{aligned} \quad . \tag{8.3}$$

Since the last integral in (8.3) is bounded uniformly in time by the energy, we have that for some constant C depending only on the initial energy,

$$\begin{aligned}&\int_{\theta\le a^2} W_2(F(\cdot;x,t),M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x\\ &\ge \int_{T^d} W_2(F(\cdot;x,t),M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x - C/a.\end{aligned}\tag{8.4}$$

Now choose

$$a = \frac{2C}{\int_{T^d} W_2(F(\cdot;x,t),M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x}\ .\tag{8.5}$$

Combining (8.2) and (8.4), and using this value of a, we get

$$\begin{aligned}\int_{T^d} \frac{W_2^2(F(\cdot;x,t),M_{F(\cdot;x,t)})}{\theta(x,t)}\rho(x,t)\mathrm{d}x &\ge C\left(\int_{T^d} W_2(F(\cdot;x,t),M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x\right)^4\\ &\ge C\left(\int_{T^d} W_2(F(\cdot;x,t),M_{F(\cdot;x,t)})\rho(x,t)\mathrm{d}x\right)^4\\ &\ge CW_1^4(f(\cdot,\cdot,t),M_{f(\cdot,\cdot,t)})\end{aligned}\tag{8.6}$$

for a constant C depending only on the dimension d and the energy of the initial data. Note in the last line, the subscript in the expression $W_1(f(\cdot,\cdot,t),M_{f(\cdot,\cdot,t)})$, where W_1 denotes the 1–Wasserstein metric defined on the set of densities on $T^d\times I\!R^d$. To obtain (8.6) we have used Schwarz's inequality yielding that $W_1\le W_2$, and we have used the simple fact that since f and M_f have the same spatial marginal density $\rho(x,t)$, then

$$W_1(f(\cdot,\cdot,t),M_{f(\cdot,\cdot,t)}) \le \int_{T^d} W_1(F(\cdot,\cdot,t),M_{F(\cdot\cdot,t)})\rho(x,t)\mathrm{d}x\ .$$

It seems likely that

$$t\mapsto W_1(f(\cdot,\cdot,t),M_{f(\cdot,\cdot,t)})$$

is Hölder continuous in t. Indeed, given the bounds in Section 5, it would suffice to show that

$$W_1(M_{f(\cdot,\cdot,s)},M_{f(\cdot,\cdot,t)})\le C|t-s|^{1/2}$$

for some constant C independent of s and t. However, we do not know how to do this. For each $T>0$ large time, we then introduce $\mathcal{M}_T$, a subset of the set $\mathcal{M}$ of local Maxwellian densities, that depends on the initial density f_0 :

$$\mathcal{M}_T := \{M_{f(t,\cdot,\cdot)}:\ t\ge T\}.$$

For $f\in\mathcal{M}$, we define

$$\mathrm{dist}_T(f,\mathcal{M}) = \inf\{W_1(f,M)\mid M\in\mathcal{M}_T\}\ ,$$

through the W_1 Wasserstein metric.

Theorem 8.1: (Gbobal asymptotic behavior for large times) *Let f be a solution of (1.7) constructed as in Theorem 7.1, and let $T>0$. Then*

$$\lim_{t\to+\infty}\mathrm{dist}_T(f(\cdot,\cdot,t),\mathcal{M}_T)=0.$$

Proof: Define

$$B(t) = \mathrm{dist}_T(f(\cdot,\cdot,t), \mathcal{M}).$$

Schwarz's inequality that gives that $W_1 \leq W_2$. This, together with Theorem 7.1, yields that there exists $C > 0$ depending only on the entropy and the energy of the initial data f_0 such that

$$|B(t) - B(s)| \leq C\left(|t-s| + \sqrt{|t-s|}\right),$$

for all $s, t \in (0, +\infty)$. Since (5.11), (8.1), and (8.6) imply that $B^4 \in L^1(0,\infty)$, we conclude the proof of Theorem 8.1. ■

Note: The fact on which the proof of Theorem 8.1 turns is that an integrable uniformly Hölder continuous function on $I\!R$ tends to zero at infinity.

We also observe that as a consequence of (5.11) and (8.1) we have that

$$\int_0^\infty H\left(f(\cdot,\cdot,t)|M_{f(\cdot,\cdot,t)}\right) \mathrm{d}t < \infty\ .$$

Hence, as a consequence of the Csiszar–Kullback inequality,

$$\int_0^\infty \|f(\cdot,\cdot,t) - M_{f(\cdot,\cdot,t)}\|^2_{L^1(T^d \times I\!R^d)} \mathrm{d}t < \infty\ .$$

These estimates complement the information on asymptotic behavior that is provided by Theorem 8.1.

Appendix: Regularity of Bessel potentials

In this appendix, we prove regularity estimates on Bessel potentials that are used in the velocity averaging analysis. These results are probably in the literature, but we have not found a reference. Results comparable to those we need here appear in a lecture notes by Tartar [?]. For some reason most versions of the Sobolev inequality giving L^∞ bounds are stated for bounded domains $I\!R^d$. The version of Morrey's inequality stated in Brezis's book [6] is an exception, though it doesn't directly give exactly what we need. The following simple result does.

First, let us recall that for any $g \in L^p(I\!R^d)$ and $\alpha > 0$,

$$(1-\Delta)^{-\alpha/2} g(v) = \int_{I\!R^d} G_\alpha(v-w) g(w) \mathrm{d}w \tag{9.1}$$

where G_α is a Bessel potential given by

$$G_\alpha(v) = C_\alpha \int_0^\infty e^{-|v|^2/4t} e^{-t} t^{(-d+\alpha)/2} \frac{\mathrm{d}t}{t} \tag{9.2}$$

and C_α is a constant depending only on α and d. The constant are such that $\int_{I\!R^d} G_\alpha(v)\mathrm{d}v = 1$, and so the convolution operation in (9.1) is a contraction on each $L^p(I\!R^d)$, $1 \leq p \leq \infty$. For these facts, as well as the value of C_α, see [31]. There, one also finds the following: First since $|v|^2/4t + t \geq |v|$,

$$e^{-|v|^2/4t} e^{-t} \leq e^{-|v|/2} e^{-|v|^2/8t} e^{-t/2}\ .$$

Therefore, for $|v| > 1$,

$$G_\alpha(v) \le e^{-|v|/2} C_\alpha \int_0^\infty e^{-1/8t} e^{-t/2} t^{(-d+\alpha)/2} \frac{\mathrm{d}t}{t} \ .$$

Hence there is a constant C depending only on α and d so that

$$G_\alpha(v) \le C e^{-|v|/2} \qquad \text{for all} \qquad |v| \ge 1 \ .$$

Also, deleting the factor of e^{-t} from the integral (9.2),

$$G_\alpha(v) \le C_\alpha \int_0^\infty e^{-|v|^2/4t} t^{(-d+\alpha)/2} \frac{\mathrm{d}t}{t} = C|v|^{(-d+\alpha)}$$

where again, C is a finite constant depending only on α and d. To avoid frequent repetition of this phrase, we fix for the rest of this appendix the convention that C denotes such a constant. It follows that for $p(\alpha - d) > -d$,

$$G_\alpha \in L^p(I\!R^d) \ . \tag{9.3}$$

The result proven here is the following:

Theorem: (Hölder Continuity of Bessel Potentials) *For $0 \le \alpha < d$, $1 \le p < \infty$, and and $p(\alpha - d) > -d$, the map*

$$a \mapsto G_\alpha(\cdot - a)$$

is Hölder continuous into $L^p(I\!R^d)$ with

$$\|G_\alpha(\cdot) - G_\alpha(\cdot - a)\|_p \le C|a|^{\beta/(1+\beta)} \tag{9.4}$$

where $\beta = (d + p(\alpha - d))/(2p)$.

An immediate consequence of (9.3) is that for $0 \le \alpha < d$ and g in $L^{p'}(I\!R^d)$, $p' > d/\alpha$,

$$\|G_\alpha * g\|_\infty \le C\|g\|_{p'}.$$

Moreover, (9.4) gives that

$$\|G_\alpha * g(v) - G_\alpha * g(w)| \le C\|g\|_{p'}|v - w|^{\beta/(1+\beta)} \ .$$

Proof: For any $T > 0$, define

$$H_{\alpha,T}(v) = C_\alpha \int_0^T e^{-|v|^2/4t} e^{-t} t^{(-d+\alpha)/2} \frac{\mathrm{d}t}{t}$$

and

$$K_{\alpha,T} = G_\alpha - H_{\alpha,T} \ .$$

Then

$$\|G_\alpha(\cdot) - G_\alpha(\cdot - a)\|_p \le 2\|H_{\alpha,T}\|_p + \|K_{\alpha,T}(\cdot) - K_{\alpha,T}(\cdot - a)\|_p \ . \tag{9.5}$$

We now show that

$$\|H_{\alpha,T}\|_p \le C T^{(d-p(d-\alpha))/2p} \ . \tag{9.6}$$

Indeed,

$$\begin{aligned}\int_0^T e^{-|v|^2/4t}e^{-t}t^{(-d+\alpha)/2}\frac{\mathrm{d}t}{t} &= \left(\int_0^{T/|v|^2} e^{-1/4s}s^{(-d+\alpha)/2}e^{-s|v|^2}\frac{\mathrm{d}s}{s}\right)|v|^{\alpha-d}\\ &\le \left(\int_0^{\infty}\left(\max_{0\le s\le T/|v|^2} e^{-1/8s}\right)e^{-1/8s}s^{(-d+\alpha)/2}\frac{\mathrm{d}s}{s}\right)|v|^{\alpha-d}\\ &= \left(\int_0^{\infty} e^{-1/8s}s^{(-n+\alpha)/2}\frac{\mathrm{d}s}{s}\right)e^{-|v|^2/8T}|v|^{\alpha-n}\end{aligned}$$

The integral is finite, and hence

$$H_{\alpha,T}(v) \le Ce^{-|v|^2/8T}|v|^{\alpha-d}$$

from which (9.6) follows immediately.

We now show that

$$\|K_{\alpha,T}(\cdot) - K_{\alpha,T}(\cdot - a)\|_p \le C\left(\frac{|a|}{T}\right) . \tag{9.7}$$

Let $w = v - a$. Then

$$||v|^2 - |w|^2| \le |a|(2|v| + |a|) ,$$

and hence

$$\begin{aligned}\left|e^{-|v|^2} - e^{-|w|^2}\right| &\le e^{-|v|^2}\left(e^{|a|(2|v|+|a|)} - 1\right)\\ &\le |a|(2|v| + |a|)e^{|a|(2|v|+|a|)}e^{-|v|^2}\\ &\le |a|(2|v| + |a|)e^{|a|(2|v|+|a|)-|v|^2/2}e^{-|v|^2/2}\\ &\le |a|(2|v| + |a|)e^{5|a|^2}e^{-|v|^2/2} .\end{aligned}$$

Therefore,

$$\begin{aligned}|K_{\alpha,T}(v) - K_{\alpha,T}(v-a)| &\le C_\alpha\int_T^\infty \frac{|a|(2|v|+|a|)}{4t}e^{5a^2/4t}e^{-|v|^2/8t}e^{-t}t^{(-d+\alpha)/2}\frac{\mathrm{d}t}{t}\\ &\le C_\alpha\frac{|a|(2|v|+|a|)}{4T}e^{5a^2/4T}\int_T^\infty e^{-|v|^2/8t}e^{-t}t^{(-d+\alpha)/2}\frac{\mathrm{d}t}{t}\\ &\le \frac{|a|(2|v|+|a|)}{4T}e^{5a^2/4T}CG_\alpha(v/\sqrt{2}) .\end{aligned}$$

Now both $G_\alpha(v)$ and $|v|G_\alpha(v)$ belong to $L^p(I\!R^d)$, and so it follows from this that

$$\|K_{\alpha,T}(\cdot) - K_{\alpha,T}(\cdot - a)\|_p \le C\left(\frac{|a| + |a|^2}{T}\right)e^{5|a|^2/T} . \tag{9.8}$$

This is not quite (9.7), but since no matter how large $|a|$ is, $\|K_{\alpha,T}(\cdot) - K_{\alpha,T}(\cdot - a)\|_p \le 2\|G_\alpha\|_p$, we can increase C so that (9.7) holds.

Combining (9.4), (9.6) and (9.7), we have

$$\|G_\alpha(\cdot) - G_\alpha(\cdot - a)\|_p \le C(T^\beta + |a|/T) .$$

The optimal choice of T is $T = |a|^{1/(1+\beta)}$, which yields (9.4) ■

References

[1] R. Adams. *Sobolev Spaces*, Academic Prss, New York. 1975.

[2] M. Bézard. *Régularité précisse des moyennes dans les équations de transport.* Bull. Soc. Math. France **122**, 29–76, 1994.

[3] F. Bouchut, F. Golse and M. Pulvirenti. *Kinetic equations and asymptotic theories*, 2001.

[4] L. Boltzmann. *Weitere Studien über das Wäregleichgewicht unter Gasmolekülen.* Sitzungberichte Akad. Wiss. , Vienna **66** 275–370 1872.

[5] Y. Brenier. *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math., **44** 375–417, 1991.

[6] H. Brezis. *Analyse Fonctionelle, Théorie et Applications.* Masson, Paris, 1983.

[7] L.A. Caffarelli. *The regularity of mappings with a convex potential.* J. Amer. Math. Soc., **5** 99–104, 1992.

[8] L.A. Caffarelli. *Boundary regularity of maps with convex potentials II.* Ann. of Math., **144** 453–496, 1996.

[9] L.A. Caffarelli. *Monotonicity properties of optimal transportation and the FKG and related inequalities.* To appear in Comm. Math. Phys., 2001.

[10] E. A. Carlen and W. Gangbo. *Constrained Steepest Descent in the 2–Wasserstein Metric* (preprint).

[11] C. Cercignani. *Boltzmann Equation.* Springer 1996.

[12] C. Cercignani, R. Illner and M. Pulvirenti. *The Mathematical Theory of Dilute Gasses.* Applied Mathematical Science, Vol 106, Springer Verlag, Berlin, 1994.

[13] L.A. DiPerna, P.L. Lions. *On the Cauchy Problem for Boltzmann Equations.* Annals of Math., **130**, 321–366, 1989.

[14] L.A. DiPerna, P.L. Lions. *Global solutions of Boltzmann's Equation and the entropy inequality.* Arch. Rational. Mech. Anal., **114**, 47–55, 1991.

[15] L.A. DiPerna, P.L. Lions and Y. Meyer. *L^p regularity of velocity averages*, Ann. I.H.P. Analyse non-linéaire **8**, 271–287, 1991.

[16] L.C. Evans. *Partial differential equations and Monge-Kantorovich mass transfer.* Int. Press, Current Dev. Math, **26**, 26–78, 1997.

[17] L.C. Evans and W. Gangbo. *Differential equations methods for the Monge-Kantorovich mass transfer problem.* Memoirs of AMS., no 653, vol. 137, 1–66, 1999.

[18] W. Gangbo. *An elementary proof of the polar factorization of vector-valued functions.* Arch. Rational Mech. Anal., **128** 381–399, 1994.

[19] W. Gangbo and R.J. McCann. *The geometry of optimal transportation.* Acta Math., **177** 113–161,

1996.

[20] Jordan, Kinderleher and Otto. *The variational formulation of the Fokker–Planck equation.* SIAM Jour. Math Anal., **29** 1–17, 1998.

[21] L. Kantorovich. *On the translocation of masses.* C.R. (Doklady) Acad. Sci. USSR (N.S.), **37** 199–201, 1942.

[22] L.V. Kantorovich and G.S. Rubinstein. *On a functional space and certain extremum problems.* Dokl. Akad. Nauk USSR (N.S.), **115** 1058–1061, 1957.

[23] E. Lieb and M. Loss. *Analysis.* Graduate Studies in Math, Volume 14, Amer. Math Soc., Providence , 1997.

[24] J.C. Maxwell. *On the Dynamical Theory of Gasses.* Phil. Trans. Royal Soc. London **157** 49–88, 1867.

[25] R.J. McCann. *Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J., **80** 309–323, 1995.

[26] R.J. McCann. *A convexity principle for interacting gases.* Adv. Math. 128 153–179, 1997.

[27] G.Monge. *Mémoire sur la théorie des déblais et de remblais.* Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, pages 666–704 (1781).

[28] F. Otto. *Doubly degenerate diffusion equations as steepest descent.* Preprint no 480 of the SFB 256, University of Bonn (1996). See also http://riemann.iam.uni-bonn.de/ otto/publications.html.

[29] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems*; Vol. I & II. Probability and its Applications, Springer Verlag, Heidleberg 1998.

[30] S.T. Rachev. *The Monge-Kantorovich mass transference problem and its stochastic applications.* Theory Prob. Appl., **29**, 647–676, 1984.

[31] E. Stein. *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, 1970.

[32] E. Stein. *The characterization of of functions arising as potentials I.* Bull. Amer. Math. Soc., **67**, 1961, 102–104, "II" (*ibid*), **68**, 1962, 577-582.

[33] M. Talgrand. *Transportation cost for Gaussian and other Product Measures.* Geom. Func. Anal., **6**, 587–600, 1996.

[34] L. Tartar. *An introduction to Sobolev spaces and interpolation spaces.* http://www.math.cmu.edu/cna/publications.html

[35] C. Villlani. *Topics in Mass Transportation* Lecture notes from a course of the author. Private communication.

[36] C. Villlani. *Generalization of an inequality by Talagrand and its links with the logarithmic Sobolev inequality.* Jour. Func. Anal., **173**, 361–400, 2000.

[37] L.N. Wasserstein. *Markov processes over denumerable products of spaces describing large systems of automata*. Problems of Information Transmission, **5**, 47–52, 1969.