Deterministic and Stochastic approaches to Relaxation to Equilibrium for Particle Systems

Josephine Angela Holly Evans

King's College

University of Cambridge

A thesis submitted for the degree of

Doctor of Philosophy

July 2018

Summary

This work is about convergence to equilibrium problems for equations coming from kinetic theory. The bulk of the work is about Hypocoercivity. Hypocoercivity is the phenomenon when a semi-group shows exponentially relaxation towards equilibrium without the corresponding coercivity (dissipativity) inequality on the Dirichlet form in the natural space, i.e. a lack of contractivity. In this work we look at showing hypocoercivity in weak measure distances, and using probabilistic techniques. First we review the history of convergence to equilibrium for kinetic equations, particularly for spatially inhomogeneous kinetic theory (Boltzmann and Fokker-Planck equations) which motivates hypocoercivity. We also review the existing work on showing hypocoercivity using probabilistic techniques.

We then present three different ways of showing hypocoercivity using stochastic tools. First we study the kinetic Fokker-Planck equation on the torus. We give two different coupling strategies to show convergence in Wasserstein distance, W_2. The first relies on explicitly solving the stochastic differential equation. In the second we couple the driving Brownian motions of two solutions with different initial data, in a well chosen way, to show convergence. Next we look at a classical tool to show convergence to equilibrium for Markov processes, Harris's theorem. We use this to show quantitative convergence to equilibrium for three Markov jump processes coming from kinetic theory: the linear relaxation/BGK equation, the linear Boltzmann equation, and a jump process which is similar to the kinetic Fokker-Planck equation. We show convergence to equilibrium for these equations in total variation or weighted total variation norms. Lastly, we revisit a version of Harris's theorem in Wasserstein distance due to Hairer and Mattingly and use this to show quantitative hypocoercivity for the kinetic Fokker-Planck equation with a confining potential via Malliavin calculus.

We also look at showing hypocoercivity in relative entropy. In his seminal work work on hypocoercivity Villani obtained results on hypocoercivity in relative entropy for the kinetic Fokker-Planck equation. We review this and subsequent work on hypocoercivity in relative entropy which is restricted to diffusions. We show entropic hypocoercivity for the linear relaxation Boltzmann equation on the torus which is a non-local collision equation. Here we can work around issues arising from the fact that the equation is not in the Hörmander sum of squares form used by Villani, by carefully modulating the entropy with hydrodynamical quantities. We also briefly review the work of others to show a similar result for a close to quadratic confining potential and then show hypocoercivity for the linear Boltzmann equation with close to quadratic confining potential using similar techniques.

We also look at convergence to equilibrium for Kac's model coupled to a non-equilibrium thermostat. Here the equation is directly coercive rather than hypocoercive. We show existence and uniqueness of a steady state for this model. We then show that the solution will converge exponentially fast towards this steady state both in the GTW metric (a weak measure distance based on Fourier transforms) and in W_2. We study how these metrics behave with the dimension of the state space in order to get rates of convergence for the first marginal which are uniform in the number of particles.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below. It is not substantially the same as any that I have submitted, or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

Chapter 1 is an introduction and contains my own review done under the guidance of Clément Mouhot. Chapter 2 is a joint work with Helge Dietert who was at the University of Cambridge at the time and is now at University Paris Diderot and Thomas Holding who was also at the university of Cambridge and is now at Imperial College London, the collaboration and problem was suggested by Clément Mouhot, this work will shortly be published in *Kinetic and Related Models*. Chapter 3 is a joint work with José Cañizo at the University of Granada, Cao Chuqi at University Paris Dauphine and Havva Yoldas at Basque Centre for Applied Mathematics the collaboration and problem was suggested by José Cañizo. Chapter 4 Is my own work done under the supervision and guidance of Clément Mouhot. Chapter 5 is my own work and has been submitted for publication it was done under the supervision and guidance of Clément Mouhot. Chapter 6 is my own work and has been published in the *Journal of Statistical Physics* [57] this was also done with the supervision and guidance of Clément Mouhot

Acknowledgements

I am very grateful to many people for their encouragement and support during my time as a PhD student. First, I would like to thank my supervisor Prof. Clément Mouhot. I knew nothing of PDEs before taking his course in part III by which I was immediately converted. During my PhD I have benefited enormously from his insight and almost countless questions! I would like to thank my collaborators. My first piece of work was in collaboration with two other students of Clément: Tom Holding and Helge Dietert. They were very kind a patient in introducing me to the world of research. I would like to thank José Cañizo and Havva Yoldas for welcoming to Granada and for the very exciting project! The kinetic group in Cambridge has been a great source of ideas, support and advice. I would like to thank Sara Merino-Aceituno, Tom Holding, Helge Dietert, Franca Hoffmann, Ludovic Cesbron, Davide Piazzoli, Megan Griffin-Pickering, Lisa Kreusser, Angeliki Menegaki, Amit Einav, Harsha Hutridurga, Arianne Trescases, Mikaela Iacobelli, Hong Xiao, Andre Ichim and Iván Moyano. Outside of Cambridge I would like to thank Jean Dolbeault, Francesco Salvarani and Alessandra Iacobucci for welcoming to Dauphine. I hope we can work on many projects together in the future. I would also like to thank Pierre Monmarché, Amit Einav, Megan Griffin-Pickering and Iván Moyano for reading parts of this work and their helpful comments. I would like to thank Jean Dolbeault and James Norris for agreeing to examine this thesis.

I would like to thank my parents and my friends in college, the department and elsewhere in the university including Benjamin Barrett, Lawrence Barrott, James Bell, Nigel Burke, Ellie Davies, Nicolas Dupré, Oliver Feng, Fiona Hamey, Amit Hazi, Oísin Huhn, James Kilbane, Markus Kunesch, James Munro, Alex Povey, Jack Smith and Benjamin Stokell. Finally I would like to thank Tom for everything.

Contents

Chapter 1

Introduction

This thesis is about proving quantitative rates of convergence to equilibrium for a range of equations coming from kinetic theory. In this introduction first we introduce kinetic theory, Boltzmann's equation and give Boltzmann's H-theorem as the main historical example related to convergence to equilibrium for kinetic equations. We introduce a range of equations which will be discussed throughout this work. After this we give a review of some of the existing work on convergence to equilibrium, briefly discussing the entropy-entropy production method, and then focusing on spatially inhomogeneous kinetic equations. We also discuss *hypocoercivity* particularly in this context. Lastly we briefly describe the contents of each of the chapters.

1.1 Kinetic Equations

Kinetic theory was developed in the 19^{th} century, most notably by Boltzmann and Maxwell, in the modelling of dilute gases. Kinetic equations model the evolution of a gas in an intermediate scale between the microscopic description which is given by Newton's laws and a macroscopic fluid descriptions of the observed behaviour.

If we have a system of N particles performing either deterministic or stochastic dynamics we can write their joint distribution at time t,

$$F_N(t, z_1, \dots, z_N).$$

In this case z_i is either v_i, the velocity of the i^{th} particle, or (x_i, v_i), the position and velocity of the i^{th} particle. We look at the situation where this equation models a large number of indistinguishable agents, for example gas particles. In this situation the dynamics are relatively simple, Newton's laws, but the equation is very high dimensional. Kinetic equations are derived by looking at the average behaviour of one particle as the total number of particles tends to infinity, i.e. if Π_1 is the marginal distribution of the first particle then

$$f(t, z) = \lim_{N \to \infty} \Pi_1[F_N](z).$$

Under appropriate assumptions on the scaling and dynamics it is then possible (at least formally) to write an equation for f and also show that f describes the average behaviour in the system in the sense that the empirical measure

$$f^N = \frac{1}{N}\sum_{i=1}^{N}\delta_{(x_i,v_i)}$$

will converge weakly towards f.

We study particles interacting in a gas by Newton's laws, then they will follow the equations

$$\begin{aligned}\dot{x}_i =& v_i,\\ \dot{v}_i =& \sum_j F_{j,i} + F,\end{aligned}$$

where $F_{j,i}$ is the force acting on particle i due to particle j and F is an external force. For collisional gases the equation derived in this process, via the Boltzmann-Grad scaling, is Boltzmann's equation

$$\partial_t f + v\cdot\nabla_x f = Q(f,f).$$

Here

$$Q(f,g) = \int_{\mathbb{R}^d}\int_{\mathbb{S}^{d-1}} B(|v-v_*|,(v-v_*)\cdot\sigma)\left(f(v')g(v'_*) - f(v)g(v_*)\right)\mathrm{d}\sigma\mathrm{d}v_*,$$
$$v' = \frac{v+v_*}{2} + \frac{|v-v_*|}{2}\sigma, \quad v'_* = \frac{v+v_*}{2} - \frac{|v-v_*|}{2}\sigma.$$

We also sometimes write this using the ω formulation where we reparametrise to have

$$v' = v - (v-v_*)\cdot\omega\omega, \quad v'_* = v_* + (v-v_*)\cdot\omega\omega.$$

Here

$$\sigma = \frac{(v-v_*)}{|v-v_*|} - 2\left(\frac{(v-v_*)}{|v-v_*|}\cdot\omega\right)\omega,$$

and this transformation is invertible. B is called the *collision kernel.* Here f represents the density in phase space of a single particle in the ensemble. We can observe here the general structure of a collisional kinetic equation

$$\partial_t f + v\cdot\nabla_x f = L(f),$$

where the $v\cdot\nabla_x$ operator comes from the transport term in Newton's law. The $L(f)$ operator acts only in v and is the result of collisions between particles or between particles and a background medium. We can derive macroscopic quantities from this density.

$$\rho(x) = \int f(x,v)\mathrm{d}v,$$

is the local density,

$$u(x) = \frac{1}{\rho(x)}\int v f(x,v)\mathrm{d}v,$$

is the local speed and

$$T(x) = \frac{1}{\rho(x)}\int |v-u(x)|^2 f(x,v)\mathrm{d}v,$$

is the local temperature. The steady state solution of Boltzmann's equation was derived by Maxwell and shown to be the unique asymptotic equilibrium by Boltzmann. It is known as the Maxwellian

$$\mathcal{M}(v) = \rho(2\pi T)^{-d/2} \exp\left(-\frac{1}{2T}|v|^2\right).$$

Here ρ, T are the spatial averages of the local quantities above. Since this is a steady state of the equation it becomes natural to ask whether the solution to Boltzmann's equations will eventually come close to the Maxwellian and if so how fast this will happen.

Boltzmann's H-theorem

The first and most celebrated work relating to convergence to equilibrium for a kinetic equation is Boltzmann's H-theorem, which can be found in [23, 109]. Boltzmann showed that the entropy of a solution to the Boltzmann equation will always increase. We will give a sketch of the proof in this section.

Definition 1.1. *If f is a probability density then we define the entropy of f by*

$$H(f) = \int_{\mathbb{R}^d} f(v) \log(f(v)) \mathrm{d}v.$$

In order to look at how entropy behaves along the flow of Boltzmann's equation it is helpful to look at a dual formulation of the Boltzmann equation.

Lemma 1.1. *Suppose $f(t,x,v)$ is a smooth solution to the spatially inhomogeneous Boltzmann equation then, if ϕ is a smooth function depending only on v, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} f(t,v)\phi(v)\mathrm{d}v = \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B f(v) f(v_*) \left(\phi(v') + \phi(v'_*) - \phi(v) - \phi(v_*)\right) \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v.$$

Proof. We use the fact that if we define k as $(v - v_*)/|v - v_*|$ then

$$(v, v_*, \sigma) \leftrightarrow (v', v'_*, k) \quad \text{and} \quad (v, v_*, \sigma) \leftrightarrow (v_*, v, -\sigma),$$

are invertible transformations with Jacobian 1 which leave B constant. Also a ϕ only depends on v the transport term vanishes. This gives,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} f(t,v)\phi(v)\mathrm{d}v &= \int_{\mathbb{R}^d} Q(f,f)\phi(v)\mathrm{d}v \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B \left(f(v')f(v'_*) - f(v)f(v_*)\right) \phi(v) \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B f(v) f(v_*) \phi(v') \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v \\
&\quad - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B f(v) f(v_*) \phi(v) \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B f(v) f(v_*) \phi(v'_*) \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v \\
&\quad - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B f(v) f(v_*) \phi(v_*) \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v \\
&= \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B f(v) f(v_*) \left(\phi(v') + \phi(v'_*) - \phi(v) - \phi(v_*)\right) \mathrm{d}\sigma \mathrm{d}v_* \mathrm{d}v.
\end{aligned}$$

Here to get the last expression we just average the previous two. □

We can use Lemma 1.1 to look at the derivative of $H(f)$.

Theorem 1.1 (Boltzmann's H-theorem)**.** *If $f(t,v)$ is a smooth solution to the spatially inhomogeneous Boltzmann equation with finite entropy then*

$$\frac{\mathrm{d}}{\mathrm{d}t}H(f) = -D(f) \leq 0$$

and

$$D(f) = 0 \Leftrightarrow f(v) = C(2\pi T)^{-d/2}\exp\left(-\frac{1}{2T}|v-u|^2\right),$$

for fixed C, T, u. The C, T, u possible are determined by the initial conditions.

Proof. We have that by Lemma 1.1 taking $\phi = \log(f)$ assuming $f > 0$ everywhere,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} f(v)\log(f(v))\mathrm{d}v =& \frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{S}^{d-1}} Bf(v)f(v_*)\log\left(\frac{f(v')f(v'_*)}{f(v)f(v_*)}\right)\mathrm{d}\sigma\mathrm{d}v_*\mathrm{d}v \\ =& \frac{1}{4}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{S}^{d-1}} B\left(f(v)f(v_*) - f(v')f(v'_*)\right)\log\left(\frac{f(v')f(v'_*)}{f(v)f(v_*)}\right)\mathrm{d}\sigma\mathrm{d}v_*\mathrm{d}v.\end{aligned}$$

We know that the function

$$(a-b)\log(a/b)$$

is positive. So it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}H(f) \leq 0.$$

From this we see that $H(f)$ will always be non-increasing. We can see that the dissipation is only zero when

$$f(v')f(v'_*) = f(v)f(v_*)$$

for all values of v, v_* and σ. We can integrate the left hand side in σ to see that the product $f(v)f(v_*)$ only depends on $|v-v_*|$ and $(v+v_*)/2$. i.e.

$$f(v)f(v_*) = \int_{\mathbb{S}^{d-1}} f\left(\frac{v+v_*}{2} + \frac{|v-v_*|}{2}\sigma\right) f\left(\frac{v+v_*}{2} - \frac{|v-v_*|}{2}\sigma\right)\mathrm{d}\sigma.$$

We can write this in terms of the total momentum

$$m = v + v_*,$$

and the total energy

$$e = \frac{|v|^2}{2} + \frac{|v_*|^2}{2}.$$

Therefore we have that

$$\log(f(v)) + \log(f(v_*)) = g(m,e).$$

We can differentiate this in v to give

$$(\nabla\log(f))(v) = \nabla_m g(m,e) + \partial_e g(m,e)v.$$

Symmetrically,

$$(\nabla \log(f))(v_*) = \nabla_m g(m,e) + \partial_e g(m,e) v_*.$$

Therefore

$$(\nabla \log(f))(v) - (\nabla \log(f))(v_*)$$

is parallel to $v - v_*$. Immediately we can write

$$(\nabla \log(f))(v) = (\nabla \log(f))(0) + \alpha(v)v,$$

for some scalar function α. Then we also have that for all v_*

$$\alpha(v)v - \alpha(v_*)v_* = \alpha(v)(v - v_*) + (\alpha(v) - \alpha(v_*))v_*,$$

is parallel to $(v - v_*)$ for all v and v_*. This shows that $\alpha(v) = \alpha(v_*)$ whenever v is not parallel to v_* therefore α must be constant. Therefore for some $\mu \in \mathbb{R}^d$ we have

$$(\nabla \log(f))(v) = \alpha v + \mu.$$

It follows from this that f is a Maxwellian. □

1.1.1 Equations

First we make a distinction between spatially inhomogeneous and spatially homogeneous kinetic equations. We call an equation spatially homogeneous if we do not track the transport part of the equation and look only at the velocity process. We can see that this is the equation we would follow if f did not depend on x for all t. The equation is the same just removing the transport operator. For the Boltzmann equation it is

$$\partial_t f = Q(f,f), \quad f = f(t,v), \quad t \in \mathbb{R}_+, v \in \mathbb{R}^d.$$

On the other hand we can look at a pure transport equation

$$\partial_t f + v \cdot \nabla_x f = 0.$$

Here, we can solve this explicitly to get

$$f(t,x,v) = f^{in}(x - vt, v).$$

We can see from this that the transport operator does not have a unique equilibrium state. However, when we combine this operator with a spatially homogeneous kinetic equation it often produces a global equilibrium state.

We can also look at equations where instead of just transport we have both transport and confinement. This allows us to have equilibrium states where the x variable is in the whole of $\mathbb{R}^d$. This represents an external confining force. We write this equation as

$$\partial_t f + v \cdot \nabla_x f - \nabla_x U(x) \cdot \nabla_v f = Q(f,f). \tag{1.1}$$

Throughout the rest of this thesis we use the notation $\mathcal{M}(v)$ to refer to the normalised Maxwellian with $u = 0$ and $T = 1$. That is

$$\mathcal{M}(v) = (2\pi)^{-d/2} \exp\left(-\frac{|v|^2}{2}\right).$$

The kinetic Fokker-Planck equation

The kinetic Fokker-Planck equation is one of the simplest equations in kinetic theory. It is also known as the Kramers-Fokker-Planck equation. It is a kinetic version of the Fokker-Planck equation which was developed by Fokker [44], and Planck [106]. The Fokker-Planck and kinetic Fokker-Planck equation were later independently derived by Kolmogorov [87]. It is a basic model for a solute being dispersed by a medium. It is derived from Langevin dynamics, where the time scale of observation is much larger than the correlation time of the solute-fluid interactions (see e.g. [118]). We write it

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U(x) \cdot \nabla_v f) = \nabla_v \cdot (\nabla_v f + v f). \tag{1.2}$$

Here the phase space is either $\mathbb{R}^d \times \mathbb{R}^d$ or $\mathbb{T}^d \times \mathbb{R}^d$. The equilibrium state is

$$e^{-U(x)}\mathcal{M}(v) = \exp\left(-\left(U(x) + \frac{1}{2}|v|^2\right)\right).$$

This is an equation for the law of a particle satisfying the stochastic differential equation.

$$\begin{cases} \mathrm{d}X_t = V_t \mathrm{d}t, \\ \mathrm{d}V_t = -V_t\mathrm{d}t - (\nabla_x U(X_t)\mathrm{d}t) + \sqrt{2}\mathrm{d}W_t. \end{cases} \tag{1.3}$$

The spatially homogeneous version of this equation is called the Fokker-Planck equation corresponding to the Orstein-Uhlenbeck process,

$$\partial_t f = \nabla_v \cdot (\nabla_v f + v f). \tag{1.4}$$

We look at long time behaviour of the kinetic Fokker-Planck equation in Chapters 2 and 4 and briefly mention it in Chapter 5.

The linear relaxation Boltzmann equation

The linear relaxation Boltzmann equation is the simplest example of a scattering equation from kinetic theory. It is also known as the linear BGK equation. We write it

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U(x) \cdot \nabla_v f) = \Pi_{\mathcal{M}} f - f, \quad \Pi_{\mathcal{M}} f = \left(\int_{\mathbb{R}^d} f(x,u)\mathrm{d}u\right)\mathcal{M}(v). \tag{1.5}$$

Here again the phase space is either $\mathbb{R}^d \times \mathbb{R}^d$ or $\mathbb{T}^d \times \mathbb{R}^d$. The equilibrium state is

$$e^{-U(x)}\mathcal{M}(v) = \exp\left(-\left(U(x) + \frac{1}{2}|v|^2\right)\right).$$

This is an equation for the law of a particle satisfying the stochastic differential equation, in integrated form,

$$\begin{cases} X_t = X_0 + \int_0^t V_s \mathrm{d}s \\ V_t = V_0 - \left(\int_0^t \nabla_x U(X_s)\mathrm{d}s\right) + \int_0^t \int_{\mathbb{R}^d} (w - V_{s^-})\, P(\mathrm{d}s, \mathrm{d}w). \end{cases} \tag{1.6}$$

Here P is a Poisson point process with intensity Leb $\times\, \gamma$ where $\mathrm{d}\gamma = \mathcal{M}(v)\mathrm{d}v$. The spatially homogeneous version of this equation is particularly simple. The solution has probability e^{-t} to be in its original state and probability $1 - e^{-t}$ of being in a normal distribution.

There are variants of this equation which we do not study but exhibit similar behaviour. First we can generalise the collision kernel

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U(x) \cdot \nabla_v f) = \int_{\mathbb{R}^d} k(v,u) f(x,u)\mathrm{d}u - f. \tag{1.7}$$

Here $k(v,u)$ represents the rate of jumping from velocity u to velocity v.

Second we can look at the equation when the rate of collision depends on space

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U(x) \cdot \nabla_v f) = \sigma(x) \left(\Pi_{\mathcal{M}} f - f\right). \tag{1.8}$$

We look at the linear relaxation Boltzmann equation in Chapters 3 and 5.

The linear Boltzmann equation

We also look at the linear Boltzmann equation. This is a scattering type equation which can be derived from microscopic dynamics with particles interacting with a heat bath. The equation is

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U(x) \cdot \nabla_v f) = Q(f, \mathcal{M}). \tag{1.9}$$

Here Q is the Boltzmann collision operator as given before. This equation is much simpler than the full Boltzmann equation. It is linear and is the equation of the density for a Markov process provided B is sufficiently nice. The spatially homogeneous linear Boltzmann equation

$$\partial_t f = Q(f, \mathcal{M}), \tag{1.10}$$

has been well studied. The convergence to equilibrium problem is studied in [115, 93, 17]. We look at the linear Boltzmann equation in Chapters 3 and 5.

Kac's Model

Kac's model is a probabilistic model meant to have many of the same qualitative behaviours as the particle system leading to the Boltzmann equation. Here the spatial variable is treated as hidden information. It was introduced by Marc Kac in [85]. The N-particle model is

$$\partial_t F_N = N(Q[F_N] - F_N). \tag{1.11}$$

Here,

$$Q[F_N] = \frac{1}{\binom{N}{2}} \sum_{i<j} \int_0^{2\pi} b(\theta) F_N(R^\theta_{i.j}(v))\mathrm{d}\theta,$$

and $R^\theta_{i,j}$ is a rotation in the i and j^{th} variables of angle θ. There is no spatially inhomogeneous version of Kac's master equation as the randomness can be thought of replacing the x-variable for the particle system. We study Kac's model in Chapter 6.

1.1.2 Entropy-entropy production inequalities

A major way of showing convergence to equilibrium with a given rate is to show an entropy-entropy production inequality. If we can prove such an inequality quantitatively this will give us quantitative rates. If f_t is the solution to an equation then we define the entropy production, D, by

$$\frac{\mathrm{d}}{\mathrm{d}t} H(f) = -\mathcal{D}(f). \tag{1.12}$$

Then an entropy-entropy production inequality relates the functional $\mathcal{D}(f)$ to $H(f)$. It is usually of the form

$$\mathcal{D}(f) \geq \Theta(H(f)), \tag{1.13}$$

where Θ is an increasing function. This is most useful when Θ is linear then we get that

$$\frac{\mathrm{d}}{\mathrm{d}t} H(f) \leq -CH(f),$$

for some constant C. This combined with Grönwall's inequality allows one to show exponential convergence towards equilibrium.

The Fokker-Planck equation

A first example of this is to look at the Fokker-Planck equation.

$$\partial_t f = \nabla_v \cdot (\nabla_v f + vf). \tag{1.14}$$

We can define $h = f/\mathcal{M}$ and get the equation

$$\partial_t h = (\nabla_v - v) \cdot \nabla_v h. \tag{1.15}$$

Then instead of looking at the absolute entropy we can look at the entropy relative to $\mathcal{M}$.

Definition 1.2. *The relative entropy of a measure ν to another μ is given by*

$$H(\nu|\mu) = \int \log \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right) \mathrm{d}\nu.$$

We also sometimes write this with the argument h where $\mathrm{d}\nu/\mathrm{d}\mu = h$ rather than with the argument ν

$$H_\mu(h) = \int h \log(h) \mathrm{d}\mu.$$

Then in the same way as for the Boltzmann equation we can look at the production of this entropy. We have that if h is a solution to (1.15) and if μ is the measure which has density $\mathcal{M}$

then

$$\frac{\mathrm{d}}{\mathrm{d}t}H(f|\mu) = \int (\partial_t h)\log(h)\mathrm{d}\mu$$
$$= \int ((\nabla_v - v)\cdot\nabla_v h)\log(h)\mathrm{d}\mu$$
$$= -\int \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu.$$

We define the relative Fisher information,

$$I(f|\mu) = \int \frac{|\nabla h|^2}{h}\mathrm{d}\mu.$$

Now we would like to have a relationship between $H(f|\mu)$ and its time derivative.

Definition 1.3. *We say that a measure μ satisfies a logarithmic Sobolev inequality if there exists a C such that for all h we have*

$$H(f|\mu) \leq CI(f|\mu).$$

Then we have due to [63] that μ, in our situation which is a standard Gaussian, will satisfy a logarithmic Sobolev inequality with $C = 1/2$. Using this inequality we have

$$\frac{\mathrm{d}}{\mathrm{d}t}H(f|\mu) \leq -2H(f|\mu).$$

Therefore, after Grönwall's inequality we have

$$H(f(t)|\mu) \leq e^{-2t}H(h(0)|\mu).$$

This shows that $H(f|\mu)$ converges exponentially fast towards zero. We also have the Csiszar-Kullback-Pinsker inequality which can be found in [117] in Chapter 22 remark 22.12. It was derived independently by Csiszár [43], Kullback [88] and Pinsker [105].

Lemma 1.2 (Csiszar-Kullback-Pinsker Inequality)**.** *If μ and ν are two probability measures then*

$$\|\nu - \mu\|_{TV} \leq \sqrt{\frac{1}{2}H(\nu|\mu)}.$$

This shows that the relative entropy controls the total variation distance and also that $H_\mu(\nu)$ can only be zero when $\mu = \nu$ almost everywhere.

The Boltzmann equation

Entropy-entropy production inequalities for the spatially homogeneous Boltzmann equation are more complicated. In 1999, Bobylev and Cercignani [18] showed that for physical collision kernels there is no hope of a linear entropy-entropy production inequality as we see in the Fokker-Planck case. In 1988-89 Desvillettes established an entropy-entropy production inequality using compactness tools [45] The first quantitative entropy-entropy production inequality was established by Carlen and Carvalho in 1992 [36] and looks like

$$D(f) \geq C(f_0)\Theta(H(f)),$$

where Θ is very flat around 0 and f_0 is the initial condition. It was then shown by Toscani and Villani in 1999 [111] that when B is sufficiently well behaved one can prove an entropy-entropy production inequality of the form

$$D(f) \geq C_\epsilon(f_0) H(f)^{1+\epsilon},$$

for any $\epsilon > 0$.

1.1.3 Other Metrics/Entropies

Entropy and relative entropy are not the only ways to measure the convergence of a solution towards equilibrium The entropy or relative entropy we define sits within a class of other entropies which we call Φ-entropies. These are introduced properly in chapter 4. In general they take the form

$$\int \Phi\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu.$$

The most important examples are the entropy from above where

$$\Phi(t) = t\log(t) - t + 1$$

and the squared entropy

$$\Phi(t) = (t-1)^2.$$

This second one also corresponds to the Sobolev space $L^2(\mu^{-1})$ when μ has a density. There are also the p-entropies which interpolate between them when $p \in (1,2]$ and

$$\Phi(t) = (t-1)^p.$$

Exponentially Weighted Sobolev spaces

Suppose that f_t is a density and a solution to an equation with equilibrium state μ then let us write for $\epsilon > 0$

$$f = \mu + \epsilon h.$$

Then we have that the relative entropy is

$$H_\mu(f) = \int (\mu + \epsilon h)\log(1 + \epsilon h/\mu).$$

Assuming that $\epsilon > 0$ is small this is approximately

$$H_\mu(f) \approx \epsilon \int h + \epsilon^2 \frac{1}{2}\int h^2\mu^{-1} + O(\epsilon^3) = \epsilon^2\frac{1}{2}\int h^2\mu^{-1} + O(\epsilon^3).$$

Therefore in a perturbative setting it makes sense to study the equation in $L^2(\mu^{-1})$. We see that if $H(f|\mu)$ is a Lyapunov function for a linear or non-linear flow then $L^2(\mu^{-1})$ will be a Lyapunov function for the flow or the linearised flow.

In many situations Sobolev spaces are the most natural way of studying convergence to equilibrium. We can look at entropy-entropy production inequalities in this situation. For example for the Fokker-Planck equation if we look at $g = f/\mu$ then the $L^2(\mu^{-1})$ norm of f is the $L^2(\mu)$ norm of g, and we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\|g_t\|^2_{L^2(\mu)} = -2\|\nabla_v g_t\|^2_{L^2(\mu)}.$$

Then instead of the logarithmic Sobolev inequality we can use the Poincaré inequality, which can be derived by linearising the logarithmic Sobolev inequality to get

$$\frac{\mathrm{d}}{\mathrm{d}t}\|g_t\|^2_{L^2(\mu)} \leq -2C\|g_t\|^2_{L^2(\mu)}.$$

Using $L^2(\mu^{-1})$ or $H^k(\mu^{-1})$ also gives us many other tools such as spectral theory, geometry of Hilbert spaces, and much more.

Probabilistic metrics

All the equations we look at model the time evolution of a probability density. In fact many of them also correspond to a stochastic process. It therefore makes sense to study the behaviour of the solution in probabilistic distances. This allows us to use tools from probability and also to look at weaker convergence of solutions. We can look at total variation

$$\|\mu - \nu\|_{TV} = \sup_A (\mu(A) - \nu(A) - \mu(A^c) + \nu(A^c)) = 2\sup_A |\mu(A) - \nu(A)|. \tag{1.16}$$

We also look at Wasserstein-p distances corresponding to a metric d. That is

$$\mathcal{W}(\mu,\nu) = \inf_\pi \left(\int d(x,y)^p \pi(\mathrm{d}x, \mathrm{d}y) \right)^{1/p}, \tag{1.17}$$

where the infimum is taken over all couplings π, where π is a measure on $X \times X$ if X is the state space and π has marginals μ and ν. Lastly, in the chapter about Kac's model we will look at the Gabetta-Toscani-Wennberg metric on probability measures with finite second moment and the same first moment [59]:

$$d_{GTW}(\mu,\nu) = \sup_{\xi \neq 0} \frac{|\hat{\mu}(\xi) - \hat{\nu}(\xi)|}{|\xi|^2}. \tag{1.18}$$

A simple example of how this can lead to different types of proof is looking at constructing a coupling of two solutions to show convergence to equilibrium in Wasserstein distance for the Fokker-Planck equation. We can generate two different solutions to the SDE with different initial conditions but the same driving Brownian motion (i.e. a synchronous coupling). We have that

$$\mathrm{d}(V_t^1 - V_t^2) = -(V_t^1 - V_t^2)\mathrm{d}t.$$

Therefore,

$$\mathbb{E}((V_t^1 - V_t^2)^2) \leq e^{-2t}\mathbb{E}((V_0^1 - V_0^2)^2).$$

We can use this to evolve any coupling of $\mu(0), \nu(0)$, the initial data, to a coupling of two solutions $\mu(t), \nu(t)$. This gives that

$$\mathcal{W}_2(\mu(t), \nu(t)) \leq e^{-t}\mathcal{W}_2(\mu(0), \nu(0)).$$

1.1.4 Spatially inhomogeneous equations

We will look at convergence to equilibrium problems for spatially inhomogeneous kinetic equations. Recall that is an equation of the form

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U(x) \cdot \nabla_v f) = Q_v(f).$$

Here $f(t,\cdot,\cdot)$ is a probability density on phase space. We have that $v \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$ or $\mathbb{T}^d$. Importantly Q_v acts only in the velocity variable. For simplicity of exposition, let us look at the common situation where

$$Q_v(\mathcal{M}) = 0,$$

where $\mathcal{M}(v)$ is the Maxwellian, $(2\pi)^{-d/2}\exp\left(-|v|^2/2\right)$. This means a global steady state will be the Gibb's state

$$\mu(x,v) = \exp(-U(x))\mathcal{M}(v) \text{ or } \mu(x,v) = \mathcal{M}(v).$$

Let us again assume this will be the unique global equilibrium. If we look at entropy dissipation L^2 or H^1 spaces weighted by μ^{-1} then we have that

$$\frac{\mathrm{d}}{\mathrm{d}t} H_\mu(f) = \int Q_v(f)\log(f/\mu)\mathrm{d}\mu$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\|f(t)\|^2_{L^2(\mu^{-1})} = 2\langle Q_v(f), f\rangle_{L^2(\mu^{-1})}.$$

The transport operator is anti-symmetric on these spaces so the terms coming from this will disappear. More explicitly for relative entropy the term coming from the transport part is

$$\int (v\cdot\nabla_x f - \nabla_x U(x)\cdot\nabla_v f)(\log(f/\mu)+\mu) = \int (v\cdot\nabla_x - \nabla_x U(x)\cdot\nabla_v)\left(f\log(f/\mu)\right).$$

Integrating by parts means that this term disappears. This means that the production of the entropy will vanish on functions of the form $\rho(x)\mathcal{M}(v)$. This shows that the equation cannot be coercive in these norms or indeed in any norms where the transport operator is antisymmetric. This situation is typical for hypocoercivity. We see convergence in the velocity variable but not the spatial direction.

In 1974 Ukai showed the existence of a spectral gap for the spatially inhomogeneous Boltzmann equation for hard spheres [114]. This paper is non-quantitative as it relies on compactness through Weyl's theorem. Exponential convergence was shown for the Vlasov-Poisson-Boltzmann equation [66] and the Landau equation [65] via non-constructive estimates on the average of the collision operator over small times, this approach was applied to the Boltzmann equation in a perturbative setting [67]. This approach was then made constructive in order to show exponential convergence for the Vlasov-Maxwell-Boltzmann system in a perturbative setting [68], the Boltzmann equation in a perturbative setting in [69] and to show almost exponential convergence for various spatially inhomogeneous kinetic equations in a perturbative setting in [110]. The work [91] also shows polynomial convergence to equilibrium for the Boltzmann equation in a perturbative setting.

1.1.5 Desvillettes-Villani method for inhomogeneous equations

Prior to hypocoercivity and the works of Guo and Liu, Yang and Yu quoted above the only quantitative tool for dealing with such equations is the method of Desvillettes and Villani [46, 47] which shows quantitative convergence to equilibrium like $O(t^{-\infty})$ i.e. faster than any inverse power of t. This method is conditional on existence and regularity of solutions and works in a non-perturbative setting. This method was used by Desvillettes and Villani to show convergence to equilibrium for the kinetic Fokker-Planck equation and the full Boltzmann equation (not in a perturbative setting). It has also been adapted to work for linear relaxation equations in [33]. The result for the kinetic Fokker-Planck equation is

Theorem 1.2 (Desvillettes-Vilani 2001 [46])**.** *If f is a solution to the kinetic Fokker-Planck equation (1.2), μ the equilibrium. Then if there exists a, A such that*

$$a\mu \leq f_0 \leq A\mu,$$

and the confining potential U satisfies

$$U(x) = \omega_0^2 \frac{|x|^2}{2} + \Psi(x) + U_0,$$

with ω_0, U_0 constants and Ψ a smooth function which goes to zero as $|x| \to \infty$, and U_0 chosen so that $\exp(-U)$ integrates to one, then for every $\epsilon > 0$ there is a constant $C_\epsilon(f_0) > 0$ which is explicitly computable and only depends on U, f_0, ϵ such that

$$\|f(t) - \mu\|_1 \leq C_\epsilon(f_0) t^{-1/\epsilon}.$$

For the Boltzmann equation the situation is much more complicated. They achieve convergence rates like $O(t^{-\infty})$ in polynomially weighted Sobolev spaces conditional on a priori smoothness and lower bounds.

We very briefly sketch the proof for the kinetic Fokker-Planck equation. Now suppose that μ is the global equilibrium state, and the local equilibrium state we write as $\rho\mathcal{M}$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t} H(f|\mu) \leq -2H(f|\rho\mathcal{M}). \tag{1.19}$$

This encodes the fact that the dissipation of $H(f|\mu)$ depends on how far f is from the set of local equilibria It is then natural to look at how $H(f|\rho\mathcal{M})$ behaves in time. They then show that for any $\epsilon \in (0,1)$ that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} H(f|\rho\mathcal{M}) \geq \frac{K}{2} H(f|\mu) - C_\epsilon(f_0) H(f|\rho\mathcal{M})^{1-\epsilon}. \tag{1.20}$$

They then show that any solution to this pair of equations has $H(f|\mu)$ converging to 0 like $1/t^{1-1/\epsilon}$.

1.2 Hypocoercivity

The name hypocoercivity was first used by Villani in his mémoire *Hypocoercivity* [116]. He credits the name to Thierry Gallay to emphasise the link with hypoellipticity. Let us begin by giving a definition of hypocoercivity

Definition 1.4. *If we have an equation*

$$\partial_t f + Lf = 0,$$

and we write $f(t)$ to be the solution to this at time t, then we say the equation is hypocoercive in the norm $\|\cdot\|$, if there exists constants C, λ such that for all initial data $f(0)$ we have

$$\|f(t)\| \leq Ce^{-\lambda t}\|f(0)\|.$$

This is not necessarily a very helpful definition since this concept pre-dates hypocoercivity significantly. This inequality on the semigroup is equivalent to the generator L having a spectral gap in the norm $\|\cdot\|$. Hypocoercivity is generally a name for an extension of the entropy-entropy production method. We therefore wish to prove convergence of the above form by proving *functional inequalities*. These functional inequalities must necessarily be not on the entropy and entropy production but on different functionals which we can relate to entropy. However, hypocoercivity has become a name used for theorems which show exponential convergence for spatially inhomogeneous kinetic equations and equations with similar types of degeneracies even if they do not work via functional inequalities.

The operator is called coercive if this inequality is true with $C = 1$. In this case, if $\|\cdot\|$ is a Hilbert space norm, it is more normal to say an operator L is coercive if it satisfies the functional inequality

$$\langle (L + L^*)f/2, f\rangle \geq \lambda\|f\|.$$

This is equivalent to the earlier definition. We can see that if $\langle Lf, f\rangle \geq \|f\|^2$ then we have formally

$$\frac{\mathrm{d}}{\mathrm{d}t}\|f(t)\|^2 = -\langle (L + L^*)f(t), f(t)\rangle \leq -2\lambda\|f(t)\|^2.$$

Therefore by Grönwall's lemma we have

$$\|f(t)\|^2 \leq e^{-2\lambda t}\|f(0)\|^2.$$

On the other hand if we know that the equation is coercive we have that at $t = 0$,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \|f(t)\|^2 \leq -2\lambda\|f(0)\|^2.$$

Therefore we have

$$\langle (L + L^*)f(0)/2, f(0)\rangle \geq \lambda\|f(0)\|^2.$$

Since we can choose $f(0)$ freely this gives equivalence. On the other hand we cannot differentiate the hypocoercivity inequality with $C = 0$ to get information on the operator L. Hypocoercivity is usually used in the situation where

$$\langle Lg, g\rangle = 0,$$

for some class of functions which is larger than the set of global equilibria.

A key aspect of hypocoercivity is to try and prove constructive theorems which give explicit forms for C and λ. In particular constructive estimates for C which are important because it allows us to know the time after which the convergence effects shown by the inequality will act. An inequality of the form shown in hypocoercivity does not give any convergence until $t \geq \log(C)/\lambda$, so if C is unknown and potentially very large the result may not hold in the time frame for which the model is valid. This means it is not sufficient to know the spectral gap for a non symmetric operator. However, even the spectral gap is not computed when using many methods based on compactness.

Before hypocoercivity, there were many influential works showing convergence to equilibrium. We have discussed [66, 65, 67, 68, 69, 91, 110]. Another very influential paper was [80] which studies convergence to equilibrium for the kinetic Fokker-Planck equation with a confining potential in L^2 norm. This paper shows what we would now call hypocoercivity as well as hypoellipticity for the kinetic Fokker-Planck equation with a confining potential. This is a first example of what we will call L^2-hypocoercivity which is theorems which show hypocoercivity directly in weighted L^2 distances. This paper was one of the influences for Villani's seminal mémoire *Hypocoercivity*, [116]. In this work Villani named and formalised the study of hypocoercivity. He proved a more general result for hypoelliptic type operator in both H^1 and entropy distance, as well as reformulating the work in [102] and discussing this in the context of his earlier work with Desvillettes [46, 47]. In this section we review in more detail the proofs of hypocoercivity in H^1, relative entropy and directly in L^2. We then also review the proofs of convergence for kinetic equations using probabilistic tools.

1.2.1 H^1 Hypocoercivity

The key works in H^1 hypocoercivity are [102, 116]. We present a version of the proof in [116] in a simple setting. First we review equations in what Villani calls Hörmander sum of squares form. That is

$$\partial_t f + \sum_i A^* A f + Bf = 0,$$

where A^* is the conjugate in $L^2(\mu)$ for some probability measure μ and $B^* = -B$. We review the results in a simple 2D setting where the space is spanned taking only one order of commutators and we make several more simplifying assumptions on the commutators. We also look at the case where A and B are differentials. It does not materially change the calculations. A key example of an equation in this sum of squares form is the kinetic Fokker-Planck equation.

$$\partial_t f + v \cdot \nabla_x f = \nabla_v \cdot (\nabla_v f + vf).$$

Showing convergence to equilibrium for the kinetic Fokker-Planck equation is the main goal of the methods based on Hörmander sum of squares form. The abstract form introduced in [116] allows us to see the importance of the commutator brackets.

Let us look at an equation of the form

$$\partial_t f + Bf + A^* A f = 0, f = f(z), z \in \mathbb{R}^2 \tag{1.21}$$

where A, B are first order differential operators, with $A = a(z) \cdot \nabla, B = b(z) \cdot \nabla$. We have a probability distribution μ with A^* the conjugate in $L^2(\mu)$, $B^* = -B$ and $A\mu = 0$. let us write $C = [A, B]$.

Theorem 1.3 (Theorem 18 in section 1.4 of [116] in a simplified setting)**.** *Let $f(t, z)$ be a solution of (1.21) with initial data $f(0, z)$. Suppose that μ satisfies a Poincaré inequality. Suppose that*

$C = c(z) \cdot \nabla$ and $aa^T + cc^T \geq \lambda I$, and $[A^*, A] = 0, [C^*, C] = 0, [A, C] = 0$ and $\|[C, B]f\| \leq c_1\|Af\| + c_2\|Cf\|$. Then we will have

$$\|f(t)\|_{H^1(\mu)} \leq Ke^{-\lambda t}\|f(0)\|_{H^1(\mu)}.$$

Proof. We have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\|f\|^2_{L^2(\mu)} = -\langle Bf, f\rangle_{L^2(\mu)} - \langle Af, Af\rangle_{L^2(\mu)} = -\langle Af, Af\rangle_{L^2(\mu)} = -\int \nabla f^T \left(a(z)a(z)^T\right) \nabla f \mathrm{d}\mu.$$

Since a is one dimensional the matrix aa^T has rank one so the dissipation will vanish when ∇f is perpendicular to a at each point. We need to be able to see dissipation in the other direction. Let us look at

$$\mathscr{F}(f) = \|f\|_{L^2(\mu)} + \alpha\eta\|Af\|^2_{L^2(\mu)} + 2\beta\eta\langle Af, Cf\rangle_{L^2(\mu)} + \gamma\eta\|Cf\|^2_{L^2(\mu)},$$

with $\beta^2 < \alpha\gamma$ so that this is equivalent to H^1. We have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}(f) =& -2\|Af\|^2 - 2\alpha\eta\langle Cf, Af\rangle - 2\beta\eta\|Cf\|^2 - 2\gamma\eta\langle [C, B]f, Cf\rangle \\
& - 2\alpha\eta\|AAf\|^2 - 2\beta\eta\langle AAf, ACf\rangle - 2\gamma\eta\|ACf\|^2 \\
\leq& -2\|Af\|^2 - 2\alpha\eta\|Cf\|\|Af\| - 2\beta\eta\|Cf\|^2 + 2\gamma\eta c_1\|Af\|\|Cf\| + 2\gamma\eta c_2\|Cf\|^2 \\
=& -2\|Af\|^2 - 2\eta(\beta - \gamma c_2)\|Cf\|^2 + 2\eta(\alpha + \gamma c_1)\|Af\|\|Cf\| \\
\leq& -(2 - \eta(\alpha + \gamma c_1)/\epsilon)\|Af\|^2 - \eta(2(\beta - \gamma c_2) - \epsilon(\alpha + \gamma c_1))\|Cf\|^2.
\end{aligned}$$

Now we can choose ϵ small enough so that the coefficient in front of $\|Cf\|$ is negative and then η small enough so that the coefficient in front of $\|Af\|$ is negative so that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}(f) \leq& -\delta\left(\|Af\|^2 + \|Cf\|^2\right) \\
\leq& -\delta\lambda\|\nabla f\|^2 \\
\leq& -\frac{\delta\lambda}{2}\|\nabla f\|^2 - \frac{\delta\lambda C_P}{2}\|f\|^2 \\
\leq& -\Lambda\mathscr{F}(f).
\end{aligned}$$

In the last line we use the equivalence between $\mathscr{F}(\cdot)$ and the H^1 norm. By Grönwall's lemma this gives us that

$$\mathscr{F}(f(t)) \leq e^{-\Lambda t}\mathscr{F}(f(0)),$$

then the equivalence between $\mathscr{F}(\cdot)$ and H^1 gives

$$\|f(t)\|^2_{H^1(\mu)} \leq Ke^{-\Lambda t}\|f(0)\|^2_{H^1(\mu)}.$$

□

Remark. *Here one of the key properties of the equation that we used is that, although the operator A did not show diffusion in all the directions, it mixed with B to give C which we could use to generate convergence in the other direction. This is very similar to the kind of mechanism used to prove hypoellipticity.*

Now let us look at proving hypoellipticity in a similar way. This is an adaptation of some of the material in [79], which was originally shown in in [78] and similar results appear in the second appendix of [116].

Theorem 1.4. *With the same assumptions as above we can show that for some constant K we have*

$$\|\nabla f(t)\| \leq \frac{K}{t^{3/2}}\|f(0)\|.$$

Proof.

$$\mathscr{F}(t,f) = \|f\|^2 + \alpha(t)\|Af\|^2 + 2\beta(t)\langle Af, Cf\rangle + \gamma(t)\|Cf\|^2.$$

Then we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F} \leq& -2\|Af\|^2 + \dot{\alpha}\|Af\|^2 - 2\alpha\langle Cf, Af\rangle + \dot{\beta}\langle Af, Cf\rangle - 2\beta\|Cf\|^2 \\
&+ \dot{\gamma}\|Cf\|^2 - 2\gamma\langle [C,B]f, Cf\rangle \\
\leq& -(2-\dot{\alpha})\|Af\|^2 + (2\alpha+\dot{\beta})\|Af\|\|Cf\| - (2\beta - \dot{\gamma})\|Cf\|^2 \\
&+ 2\gamma c_1\|Af\|\|Cf\| + 2\gamma c_2\|Cf\|^2 \\
\leq& -(2-\dot{\alpha})\|Af\|^2 - (2\beta - \dot{\gamma} - 2\gamma c_2)\|Cf\|^2 + (2\alpha + \dot{\beta} + 2\gamma c_1)\|Af\|\|Cf\|.
\end{aligned}$$

Let us set $\alpha = \tilde{\alpha}\epsilon t, \beta = \tilde{\beta}\epsilon t^2, \gamma = \tilde{\gamma}\epsilon t^3$. This gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F} \leq& -(2-\epsilon\tilde{\alpha})\|Af\|^2 - t^2\epsilon(2\tilde{\beta} - 3\tilde{\gamma} - 2\tilde{\gamma}c_2 t)\|Cf\|^2 + t\epsilon(2\tilde{\alpha} + 2\tilde{\beta} + 2t^2\tilde{\gamma}c_1)\|Af\|\|Cf\| \\
\leq& -\left(2 - \epsilon\tilde{\alpha} - \frac{\epsilon\eta}{2}(2\tilde{\alpha} + 2\tilde{\beta} + 2t^2\tilde{\gamma}c_1)\right)\|Af\|^2 \\
&- t^2\epsilon\left(2\tilde{\beta} - 3\tilde{\gamma} - 2t^2\tilde{\gamma}c_2 - \frac{1}{2\eta}(3\tilde{\alpha} + 2\tilde{\beta} + 2t^2\tilde{\gamma}c_1)\right)\|Cf\|^2.
\end{aligned}$$

Now for $t < 1$ (for example) we can choose $\tilde{\alpha} = 2(2c_2+3), \tilde{\beta} = 1, \tilde{\gamma} = 1/(2c_2+3)$. This gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F} \leq -\left(2 - \epsilon 2(2c_2+3) - \frac{\epsilon\eta}{2}\delta\right)\|Af\|^2 - t^2\epsilon\left(1 - \frac{1}{2\eta}\delta\right)\|Cf\|^2,$$

where

$$\delta = 6(2c_2+3) + 2 + 2c_1/(2c_2+3).$$

So we choose $\eta = \delta$. Then we need to make ϵ small so that

$$\epsilon(2(2c_2 + 3 - \delta^2/2) = 1.$$

With this ϵ we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F} \leq -\epsilon t^2\left(\|Af\|^2 + \|Cf\|^2\right).$$

Therefore, for $t < 1$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F} \leq 0.$$

Still for $t < 1$ we have

$$\mathscr{F}(f(t)) \leq \mathscr{F}(f(0)) = \|f(0)\|^2.$$

And $\mathscr{F}$ bounds some multiple of

$$t\left(\|Af\|^2+t^2\|Cf\|^2\right)\geq t^3\left(\|Af\|^2+\|Cf\|^2\right).$$

Therefore for $t<1$ we have

$$\|Af\|^2+\|Cf\|^2\leq \frac{K}{t^3}\|f(0)\|^2.$$

□

In fact we can prove both things simultaneously, as inspired from [41].

Theorem 1.5. *With the same assumptions as above we can show that*

$$\|f(t)\|^2+\delta(t)^3\|\nabla f(t)\|^2\leq De^{-\lambda t}\|f(0)\|^2.$$

Here $\delta(t)$ is a function of t which looks like t for small t and 1 for large t ($\delta(t)=1-e^{-t}$).

Proof. Let us set $\delta(t)=(1-e^{-t})$

$$\mathscr{F}=\|f\|^2+\epsilon\left(\alpha\delta\|Af\|^2+2\beta\delta^2\langle Af,Cf\rangle+\gamma\delta^3\|Cf\|^2\right)$$

We have from before that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}\leq-(2-\epsilon\alpha\dot{\delta})\|Af\|^2-\epsilon\left(2\beta\delta^2-3\gamma\dot{\delta}\delta^2-2\gamma c_2\delta^3\right)\|Cf\|^2+\epsilon\left(2\alpha\delta+2\beta\dot{\delta}\delta+2\gamma c_1\delta^3\right)\|Af\|\|Cf\|.$$

Using that $\delta\leq 1,\dot{\delta}\leq 1$ we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}\leq&-\left(2-\epsilon\left(\alpha+\frac{1}{2\eta}\left(2\alpha+2\beta+2\gamma c_1\right)\right)\right)\|Af\|^2\\&-\epsilon\delta^2\left(2\beta-3\gamma-2\gamma c_2-\frac{\eta}{2}(2\alpha+2\beta+2\gamma c_1)\right)\|Cf\|^2.\end{aligned}$$

As before we choose $\beta=1$ and γ so that

$$3\gamma+2\gamma c_2=1.$$

Then we fix α large enough to make sure $\alpha\gamma>1$. This means we can choose η small so that the coefficient of $\|Cf\|^2$ is negative then, given this, choose ϵ small enough so that the term in front of $\|Af\|^2$ is negative. So after this we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}\leq-K\delta^2\left(\|Af\|^2+\|Cf\|^2\right).$$

As before this gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}\leq-K\delta^2\mathscr{F}.$$

Therefore,

$$\mathscr{F}(t,f(t))\leq\exp\left(-\lambda\int_0^t\delta^2(s)\mathrm{d}s\right)\mathscr{F}(0,f(0)).$$

Integrating δ^2 gives that the exponential term is bounded by $De^{-\lambda t}$ for some constant D. Then, working as before, we have for some other constant, which we also call D that

$$\|f(t)\|^2 + \delta(t)^3\|\nabla f(t)\|^2 \leq De^{-\lambda t}\|f(0)\|^2.$$

$\square$

These theorems work well for degenerate diffusion equations. We would like to also be able to deal with integro-differential equation which are an important class of kinetic equations. In particular we look at equations like the linear relaxation Boltzmann equation on the torus. Hypocoercivity was first shown for these equations in [102]. This is related to the earlier works of Guo and Liu, Yang and Yu [66, 65, 67, 68, 69, 91, 110].

We recall that the linear relaxation Boltzmann equation is

$$\partial_t f + v \cdot \nabla_x f = \Pi_{\mathcal{M}} f - f, \;\; f = f(t,x,v), (x,v) \in \mathbb{T}^d \times \mathbb{R}^d, \tag{1.22}$$

where

$$\Pi_{\mathcal{M}} f = \int_{\mathbb{R}^d} f(x,u)\mathrm{d}u\mathcal{M}(v), \quad \mathcal{M}(v) := (2\pi)^{-d/2}e^{-|v|^2/2}.$$

Here it is possible to write the equation in the form

$$A^*A + B,$$

however we can see that we will not fulfil the criteria. In this case $A = I - \Pi_{\mathcal{M}}$ and $B = v \cdot \nabla_x$.

$$[A,B]f = -\Pi_{\mathcal{M}}(v \cdot \nabla_x f) + v \cdot \nabla_x \Pi_{\mathcal{M}}(f) = \nabla_x \cdot \left(\int_{\mathbb{R}^d} (v-u)f(x,u)\mathrm{d}u \right) \mathcal{M}(v).$$

In the earlier proof we we had that $\|f\|^2$ was controlled by $\|Af\|^2 + \|Cf\|^2$. This isn't the case here. We need to use a different strategy. This is from [102] and is what Villani calls the auxiliary operator method.

Theorem 1.6. *Suppose f is a solution to (1.22). Then there exists some explicitly computable C such that*

$$\|\nabla h_t\|^2 \leq Ce^{-t}\|\nabla h_0\|^2.$$

Proof. The idea is to use the same entropy functional as for the kinetic Fokker-Planck equation. In this case the entropy functional would be

$$\|h\|^2 + A_1\|\nabla_x h\|^2 + 2A_2\langle \nabla_x h, \nabla_v h\rangle + A_3\|\nabla_v h\|^2,$$

for well chosen A_1, A_2, A_3. A key element which helps in this proof is that the collision operator acts only in the v-variable in a way which means we can use the dissipation of $\|\nabla_x h\|$ to help control mixed terms. Here $f = \mu + h\mu^{1/2}$, where μ is the equilibrium measure $\mathrm{d}\mu = \mathcal{M}(v)\mathrm{d}x\mathrm{d}v$, and the inner product is

$$\langle h, g\rangle = \int_{\mathbb{T}^d \times \mathbb{R}^d} h(x,v)g(x,v)\mathrm{d}x\mathrm{d}v.$$

We have that

$$\partial_t h + v \cdot \nabla_x h = \mu^{1/2} \int h(x,u)\mu^{1/2}(u)\mathrm{d}u - h(x,v) = \Pi h - h.$$

Differentiating along the flow we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\|h\|^2 =& -2\|(I-\Pi)h\|^2, \\
\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla_x h\|^2 =& 2\int_{\mathbb{R}^d\times\mathbb{T}^d}\nabla_x(\Pi h - h)\cdot\nabla_x h \mathrm{d}x\mathrm{d}v = -2\|\nabla_x(I-\Pi)h\|^2, \\
\frac{\mathrm{d}}{\mathrm{d}t}\langle\nabla_x h, \nabla_v h\rangle =& -\|\nabla_x h\|^2 - \int_{\mathbb{R}^d\times\mathbb{T}^d}\nabla_x(h-\Pi h)\cdot\nabla_v h \mathrm{d}x\mathrm{d}v \\
&- \int_{\mathbb{R}^d\times\mathbb{T}^d}\nabla_x h\cdot\nabla_v(h-\Pi h)\mathrm{d}x\mathrm{d}v \\
=& -\|\nabla_x h\|^2 - 2\langle\nabla_x(h-\Pi h), \nabla_v h\rangle, \\
\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla_v h\|^2 =& -2\langle\nabla_x h, \nabla_v h\rangle - 2\langle\nabla_v(h-\Pi h), \nabla_v h\rangle.
\end{aligned}$$

We have that

$$\nabla_v(h-\Pi h) = \nabla_v h + \frac{1}{2}v\Pi h,$$

$$\langle\nabla_v\Pi h, \nabla_v h\rangle = \frac{1}{2}\langle v\Pi h, \nabla_v h\rangle \le \frac{1}{2}\|\Pi h\|\|\nabla_v h\|.$$

We also have that thanks to the Poincaré inequality on the torus

$$\|\Pi h\|^2 \le C\|\nabla_x\Pi h\|^2 \le C\|\nabla_x h\|^2.$$

Now let us look at a functional of the form

$$\mathscr{F}(h) = \alpha\|\nabla_x h\|^2 + 2\langle\nabla_x h, \nabla_v h\rangle + \gamma\|\nabla_v h\|^2.$$

Differentiating this gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}(h) \le& -(2\alpha - 1/\epsilon_1)\|\nabla_x(\Pi h - h)\|^2 - (2-\gamma)\|\nabla_x h\|^2 \\
&- (2\gamma - \epsilon_1 - \gamma/2 - \gamma)\|\nabla_v h\|^2 + \gamma/2\|\Pi h\|^2 \\
\le& -(2\alpha - 1/\epsilon_1)\|\nabla_x(\Pi h - h)\|^2 - (2-\gamma-C\gamma/2)\|\nabla_x h\|^2 - (\gamma/2-\epsilon_1)\|\nabla_v h\|^2.
\end{aligned}$$

Now set $\epsilon_1 = \gamma/4$ and $\gamma = 1/(1+C/2)$ This gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}(h) \le -(2\alpha - (4+2C))\|\nabla_x(\Pi h - h)\|^2 - \|\nabla_x h\|^2 - \frac{1}{4+2C}\|\nabla_v h\|^2,$$

so we can set $\alpha = 2 + C$. We have that

$$\frac{2}{6+3C}\|\nabla h\|^2 \le \mathscr{F}(h) \le (4+2C)\|\nabla h\|^2.$$

Then we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}(h) \le -\frac{1}{4+2C}\|\nabla h\|^2 \le -\mathscr{F}(h).$$

Which then implies

$$\|\nabla h_t\|^2 \le 3(2+C)^2 e^{-t}\|\nabla h_0\|^2.$$

□

Remark. *Some of the theorems in [102] require use of a $\|h\|^2$ term in the entropy as well. This theorem only works for kinetic equations on the torus.*

1.2.2 Entropic Hypocoercivity

Hypocoercivity in relative entropy was introduced in section 6 of part 1 of [116] alongside the H^1 theory. Here Villani shows that very similar calculations and results hold when the H^1 norm is replaced by a combination of relative entropy and Fisher information. Working in relative entropy and Fisher information increases the space of possible starting conditions. As mentioned before the $L^2(\mu^{-1})$ norm is strong and requires that the initial data f_0 decays faster than a Gaussian at infinity. However requiring relative entropy and Fisher information to be finite imposes much weaker conditions on the tail of the distribution. We show Villani's result from [116] for the kinetic Fokker-Planck equation in the situation mirroring the result of the last section which combines both regularisation and long time convergence from [41].

Theorem 1.7. *If f is a solution to (1.2) with U having bounded Hessian then there exists some C and λ that we can calculate so that*

$$H_\mu(h(t)) + \delta^3(t) I_\mu(h(t)) \leq Ce^{-\lambda t} H_\mu(h(t)).$$

Proof. As before we look at a twisted functional using the components of Fisher information. We can calculate that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int h\log(h)\mathrm{d}\mu &= -\int \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu,\\
\frac{\mathrm{d}}{\mathrm{d}t}\int \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu &= 2\int \frac{\nabla_x h \mathrm{Hess}(U)\nabla_v h}{h}\mathrm{d}\mu - 2\int \frac{|\nabla_x\nabla_v h|^2}{h}\mathrm{d}\mu,\\
\frac{\mathrm{d}}{\mathrm{d}t}\int \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu &= -\int \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu + \int \frac{\nabla_v h\mathrm{Hess}(U)\nabla_v h}{h}\mathrm{d}\mu\\
&\quad -\int \frac{\nabla_x\cdot\nabla_v h}{h}\mathrm{d}\mu - 2\int\frac{\nabla_x\nabla_v h : \nabla_v\nabla_v h}{h}\mathrm{d}\mu,\\
\frac{\mathrm{d}}{\mathrm{d}t}\int \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu &= -2\int \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu - 2\int \frac{|\nabla_v\nabla_v h|^2}{h}\mathrm{d}\mu\\
&\quad -2\int \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu.
\end{aligned}$$

Therefore let us take a functional $\mathscr{F}$ of the form

$$\mathscr{F}(t,h) = H_\mu(h) + \epsilon\left(A_1\delta(t)^3 I^X + \delta(t)^2 I^M + A_3\delta(t) I^V\right).$$

For constants A_1, A_2, A_3, ϵ all positive. Here as before $\delta(t) = (1-e^{-t})$ and I^X is the component of Fisher information with derivatives only in x, I^M is the component with mixed derivatives and I^V is the component with v-derivatives. Suppose $\mathrm{Hess}(U)\leq M$ then we have

$$\begin{aligned}
\dot{\mathscr{F}} \leq &-(1+2\epsilon\delta A_3 - M\epsilon\delta^2 - \epsilon\dot{\delta}A_3)I^V\\
&-\epsilon\delta^2(1-3\dot{\delta}A_1)I^X\\
&+\epsilon\delta\left(2A_1M\delta^2 + 2\dot{\delta} + 2 + A_3\delta\right)\sqrt{I^X I^V}.
\end{aligned}$$

Letting $A_1 = 1/6, A_3 = 6$ we have

$$\begin{aligned}\dot{\mathscr{F}} \leq & -(1+12\epsilon\delta - M\epsilon\delta^2 - 6\epsilon\dot{\delta})I^V \\ & -\epsilon\delta^2(1-\dot{\delta}/2)I^X \\ & +\epsilon\delta(M\delta^2/3+2\dot{\delta}+2+6\delta)\sqrt{I^X I^V} \\ \leq & -\left(1+12\epsilon\delta - \epsilon(M\delta^2+6\dot{\delta}+(M\delta^2/3+2\dot{\delta}+2+6\delta)/2\eta\delta)\right)I^V \\ & -\epsilon\delta^2\left(1-\dot{\delta}/2-\eta(M\delta^2/3+2\dot{\delta}+2+6\delta)/2\right)I^X,\end{aligned}$$

for any $\eta > 0$. So we choose η sufficiently small so that the coefficient of X is negative. Then for this η we can choose ϵ small enough so that the coefficient of I^V is positive. Therefore for some C we have

$$\dot{\mathscr{F}} \leq -C\delta^2(I^X+I^V).$$

Then as before this leads to

$$H_\mu(h(t)) + \delta^3(t)I_\mu(h(t)) \leq Ce^{-\lambda t}H_\mu(h(t)).$$

□

1.2.3 L^2 Hypocoercivity

L^2 hypocoercivity was developed in [77] to show hypocoercivity for the linear relaxation Boltzmann equation. It was then generalised in [50] to give a strategy for showing hypocoercivity for a range of kinetic equations with one conservation law. We briefly describe the results of [50]. Here we write an abstract kinetic equation

$$\partial_t f + Tf = Lf. \tag{1.23}$$

Here the theorem gives abstract conditions which T and L should fulfil but we essentially imagine T to be a transport operator

$$T = v\cdot\nabla_x f - \nabla_x U(x)\cdot\nabla_v f,$$

and L to be a collision operator which acts multiplicatively on functions which only depend on x. We write Π to be the projection on the null space of L. Generally this will be the set of local equilibria $\rho(x)\mathcal{M}(v)$ for some function ρ. So in the case of the kinetic Fokker-Planck and the linear relaxation Boltzmann equation we have

$$\Pi f = \left(\int_{\mathbb{R}^d} f(x,u)\mathrm{d}u\right)\mathcal{M}(v).$$

The idea of this theorem is that hypocoercivity can be seen as the combination of two effects.

- *Microscopic coercivity* which is coercivity on the kinetic level. There exists $\lambda_m > 0$ such that

$$-\langle Lf, f\rangle_{L^2(\mu^{-1})} \geq \lambda_m\|(I-\Pi)f\|_{L^2(\mu^{-1})}.$$

  i.e. the equation pushes the solution toward the set of local equilibria.

- *Macroscopic coercivity* which is coercivity on the level of the hydrodynamic limit equation. This is seen through coercivity of the operator T on the set of local equilibria, there exists

$\lambda_M > 0$ such that

$$\|T\Pi f\|^2_{L^2(\mu^{-1})} \geq \lambda_M \|\Pi f\|^2_{L^2(\mu^{-1})}.$$

In the situation where Π and T are as above we have that

$$\|T\Pi f\|^2_{L^2(\mu^{-1})} = \int \left(v\mathcal{M} \cdot \nabla_x \left(\frac{\rho(x)}{e^{-U(x)}} \right) e^{-U(x)} \right)^2 \mathcal{M}(v)^{-1} e^{U(x)} \mathrm{d}v\mathrm{d}x$$

$$= \int \left| \nabla_x \left(\frac{\rho(x)}{e^{-U(x)}} \right) \right|^2 e^{-U(x)} \mathrm{d}x,$$

and

$$\|\Pi f\|^2_{L^2(\mu^{-1})} = \int \left(\frac{\rho(x)}{e^{-U(x)}} \right)^2 e^{-U(x)} \mathrm{d}x.$$

Theorem 1.8 (Dolbeault-Mouhot-Schmeiser '15)**.** *Suppose that T, L satisfy the microscopic and macroscopic coercivity assumptions. Suppose further that $\Pi T \Pi = 0$ and various auxiliary operators are bounded. Then, there exists constants C, λ such that*

$$\|e^{t(L-T)} f\|_{L^2(\mu^{-1})} \leq C e^{-\lambda t} \|f\|_{L^2(\mu^{-1})}.$$

Remark. *The bounding of the auxiliary operators mentioned in the statement is via an elliptic regularity type estimate.*

Like in H^1 hypocoercivity the proof proceeds by showing an entropy-entropy production inequality for a functional which is equivalent to our desired distance. In this case the functional has a very different form. The proof of hypocoercivity in L^2 then begins by constructing the new norm

$$H(f) = \frac{1}{2}\|f\|^2_{L^2(\mu^{-1})} + \epsilon \langle Af, f \rangle_{L^2(\mu^{-1})},$$

where

$$A = (1 + (T\Pi)^*(T\Pi))^{-1}(T\Pi)^*.$$

In [80, 77] the new entropy constructed has a similar form. The main disadvantage of this approach is that it can currently only deal with equations with one conservation law. In general H^1 hypocoercivity methods do not work for equations with a confinement potential which are also not a diffusion. It can also be extended to work with equations where the equilibrium measure is not explicit and so no Poincaré inequality is known as in [27, 82].

1.2.4 Probabilistic Hypocoercivity methods

Many hypocoercive equations are the Kolmogorov backwards equations of the laws of SDEs. Both kinetic Fokker-Planck style diffusions and linear scattering equations have an interpretation as a Markov process. Consequently, tools from probability can be fruitfully used to study the long time behaviour of these equations.

Coupling methods for hypocoercivity

To the best of my knowledge direct coupling methods have been mainly used to show convergence for diffusion equations. In [22], the authors study the kinetic Fokker-Planck equation when the confining potential is close to quadratic. They then extend this to dealing with a weakly non-linear case. If we look at a kinetic Fokker-Planck equation with quadratic confinement we can write it as

$$\begin{aligned} \mathrm{d}X_t =& V_t \mathrm{d}t, \\ \mathrm{d}V_t =& -V_t \mathrm{d}t - \lambda X_t \mathrm{d}t + \mathrm{d}W_t. \end{aligned}$$

Suppose we generate two solutions to this SDE, $(X_t^1, V_t^1), (X_t^2, V_t^2)$ and give them the same driving Brownian motion. Then if we write $Y_t = X_t^1 - X_t^2, P_t = V_t^1 - V_t^2$ we have

$$\begin{aligned} \mathrm{d}Y_t =& P_t \mathrm{d}t, \\ \mathrm{d}P_t =& -P_t \mathrm{d}t - \lambda Y_t \mathrm{d}t. \end{aligned}$$

Then we have that

$$|Y_t|^2 + |P_t|^2 \leq C e^{-\gamma t}(|Y_0|^2 + |P_0|^2),$$

where we can calculate C and γ explicitly. We can reach close to quadratic confinement by a perturbation of this result.

However, this result does not use the diffusivity of the solution at all and a similar method cannot be expected to work when the confining potential is not strictly convex. We can see if we have the kinetic Fokker-Planck on the torus without confining potential this still converges to equilibrium but if we try a similar procedure we get

$$\begin{aligned} \mathrm{d}Y_t =& P_t \mathrm{d}t, \\ \mathrm{d}U_t =& -P_t \mathrm{d}t. \end{aligned}$$

We can solve this explicitly to get that $P_t = e^{-t} P_0$ and $Y_t = Y_0 + (1 - e^{-t}) P_0$. So we have that Y_t does not converge towards zero. Here, however, we did not need to couple the processes synchronously for all time in order to get convergence towards equilibrium We would like to add some randomness to P_t for short times in order to compensate for $Y_0 + P_0$. This is the idea behind the work in chapter 2. Suppose we couple the two processes independently up to a stopping time T and then we couple them synchronously We get

$$P_t = e^{-(t-T)} P_T, \;\; Y_t = Y_T + (1 - e^{-(t-T)}) P_T, \;\; Y_T = Y_0 + (1 - e^{-T}) P_0 + \int_0^T \int_0^s e^{-(s-r)} \mathrm{d}(W_r^1 - W_r^2) \mathrm{d}s.$$

Then Y_t behaves sufficiently randomly for $t \leq T$ that we can set T to be the first time $Y_t + P_t$ hits 0.

The same problem occurs in a much more complex setting when we do have a confining potential but it is only strictly convex at infinity. A coupling strategy to show convergence in this situation is given in [52]. Here the idea is to combine reflection coupling. Suppose we have the SDE

$$\begin{aligned} \mathrm{d}X_t =& V_t \mathrm{d}t, \\ \mathrm{d}V_t =& -V_t \mathrm{d}t - \nabla_x U(X_t) \mathrm{d}t + \mathrm{d}W_t. \end{aligned}$$

Then they define two solutions with driving Brownian motions W^1 and W^2 and use two different ways of coupling W^1 and W^2. One is a well chosen reflection coupling which means that the noise acts to cancel out the difference in initial data for the X processes and one is a synchronous coupling which brings the V processes closer together when the X processes are already close together. They also need to combine these couplings with a Lyapunov condition which shows that the dynamics return to the centre of the phase space sufficiently often.

Hypocoercivity via Bakry-Emery style methods

Following on from the works of Bakry and Emery [10, 4] methods have been developed to prove functional inequalities and then rates of convergence for diffusions. Suppose we have the equation

$$\partial_t f + Lf = 0,$$

then we can define the carré du champ by

$$\Gamma(f, g) = \frac{1}{2}(L(fg) - fLg - gLf).$$

We can then iterate this and define

$$\Gamma_2(f, g) = \frac{1}{2}(L\Gamma(f, g) - \Gamma(f, Lg) - \Gamma(Lf, g)).$$

The celebrated *Bakry-Emery criterion* curvature dimension criterion $CD(\rho, n)$ is

$$\Gamma_2(f) \geq \rho\Gamma(f) + \frac{1}{n}(Lf)^2.$$

If this condition is satisfied then one is able to show local logarithmic Sobolev and Poincaré inequalities which can be used to show convergence to equilibrium results.

If we try to follow this procedure for the kinetic Fokker-Planck equation we get that

$$\Gamma(f, g) = \nabla_v f \cdot \nabla_v g,$$

and

$$\Gamma_2(f) = |\text{Hess}_v(f)|^2 + |\nabla_v f|^2 - \nabla_x f \cdot \nabla_v f.$$

We see that this cannot possibly satisfy a curvature dimension criterion since that would involve bounding $\nabla_x f \cdot \nabla_v f$ by terms involving only $\nabla_v f$.

However, the Bakry-Emery method has been extended in [12, 98] to help give results for hypocoercive operators including the kinetic Fokker-Planck equation. Here the role of the twisted norm is replaced by altering Γ so it does not depend so directly on L. They replace Γ by a new quadratic form which in the case of the kinetic Fokker-Planck equation is written

$$\tilde{\Gamma}(f, g) = a\nabla_x f \cdot \nabla_x g + b\nabla_x f \cdot \nabla_v g + b\nabla_v f \cdot \nabla_x g + c\nabla_v f \cdot \nabla_v g,$$

then define

$$\tilde{\Gamma}_2(f, g) = \frac{1}{2}(L\tilde{\Gamma}(f) - 2\tilde{\Gamma}(Lf, f)).$$

This does not allow one to prove Poincaré or log Sobolev inequalities but a new curvature dimension criterion combined with local inequalities gives point wise convergence results for semigroup generated by L. These point wise results can be integrated to show existing results and in [11] they show that these point wise results also imply convergence in Wasserstein distance.

1.3 List of the works in this thesis

Hypocoercivity via coupling for the kinetic Fokker-Planck equation on the torus

This chapter is the paper [48] which was written in collaboration with Helge Dietert and Thomas Holding. The paper is accepted for publication in *Kinetic and related models.*

It is an open question given at the end of part 1 in [116] to find direct coupling strategies to show convergence to equilibrium for hypocoercive equations coming from SDEs such as the kinetic Fokker-Planck equation. In this paper we look at the kinetic Fokker-Planck equation on the torus with no confining potential and give two coupling strategies to show convergence to equilibrium in Wasserstein-2. First we give a straightforward strategy for coupling based on an explicit solution to the SDE which is possible only in the unconfined case. Using this we can show that for two solutions to the kFP we have the following theorem

Theorem 1.9. *If μ_t and ν_t are two solutions to the kinetic Fokker-Planck equation (1.2), then we have*

$$\mathcal{W}_2(\mu_t, \nu_t) \leq \left(e^{-\lambda t} + c\, e^{-t/2\lambda^2 L^2}\right) \mathcal{W}_2(\mu_0, \nu_0)$$

for a constant c only depending on L, where we place the equation on the torus of length L.

This strategy is unusual in that it is not adapted to the filtration generated by the driving Brownian motions defining the solution to the SDE. It is also a strategy which is only possible after having an explicit solution to the SDE. We therefore develop a new method of coupling two solutions which relies on switching between synchronous and asynchronous couplings of the driving Brownian motions of two solutions. Using this coupling strategy we have the following result

Theorem 1.10. *Given initial distributions μ_0 and ν_0, then we have a coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ such that*

$$\begin{aligned}\mathcal{W}_2(\mu_t, \nu_t) &\leq \left(\mathbb{E}\left[|X_t^1 - X_t^2|_{\mathbb{T}}^2 + (V_t^1 - V_t^2)^2\right]\right)^{1/2} \\ &\leq C\zeta(t)(\sqrt{\mathcal{W}_2(\mu_0, \nu_0)} + \mathcal{W}_2(\mu_0, \nu_0)),\end{aligned}$$

where

$$\zeta(t) = \begin{cases} e^{-\min(2\lambda, 1/(2\lambda^2 L^2))t} & 4L^2\lambda^3 \neq 1 \\ e^{-2\lambda t}(1+t) & 4L^2\lambda^3 = 1 \end{cases}$$

and C is a constant that depends only on λ and L.

We show that this loss in the dependence on the initial data is necessary in the set of all co adapted couplings. We do not show that our rates our sharp but we believe that they are close to being optimal since it is the minimum of the rate of convergence corresponding to the Orstein-Uhlenbeck process in velocity and the rate of convergence for a diffusion on the torus. After this work, as discussed earlier, Eberle, Guillin and Zimmer [52] have shown hypocoercivity in weighted Wasserstein distances for the kinetic Fokker-Planck equation with confinement using a combination of reflection and synchronous coupling along with Lyapunov structure of the equation. This uses a similar but much more intricate strategy as the second strategy we give.

Perspectives

As we have mentioned the paper [52] shows how to use coupling techniques to show hypocoercivity for a kinetic Fokker-Planck equation with fairly general conditions on the confining potential. A natural next step might be to investigate showing hypocoercivity via coupling strategies for other commonly studied kinetic equations. The work in Chapters 3 and 4 could be reinterpreted to give a coupling strategy to show that the equations studied there converge. In these we take a very different direction in constructing couplings. It would be interesting to see if the work [52] could be extended to the linear relaxation equation to give convergence in Wasserstein distances for that equation.

Hypocoercivity via Harris's theorem for kinetic equations with jumps

This work is done in collaboration with José Cañizo, Cao Chuqi and Havva Yoldas. The work is not intended to be published in its current state. We intend to submit it for publication in the near future.

Harris's theorem [74, 96, 73] is a result from the theory of Markov processes. Reproving and writing Harris's theorem in a PDE context is a subject of an ongoing work from José Cañizo and Stéphane Mischler. Harris's theorem shows quantitative rates of convergence to equilibrium for processes satisfying two assumptions. First we need a Lyapunov condition which says that there is a function V s.t.

$$\frac{\mathrm{d}}{\mathrm{d}t}\int f(t,x,v)V(x,v)\mathrm{d}x\mathrm{d}v \leq -\lambda \int f(t,x,v)V(x,v)\mathrm{d}x\mathrm{d}v + C\int f(t,x,v)\mathrm{d}x\mathrm{d}v.$$

This shows that the majority of the mass concentrates in the set where V is small, so if $V \to \infty$ as $(x,v) \to \infty$ then this shows that the mass of $f(t)$ concentrates in the centre of the phase space.

The next assumption is a generation of a lower bound on the centre of the phase space. It says that for any z in the centre of the space if $f^z(t)$ is the solution with initial condition δ_z then uniformly in z there is a constant $\alpha \in (0,1)$ and a probability measure ν such that

$$f^z(t) \geq \alpha\nu.$$

If you have these two assumptions then we can define the weighted TV distance by

$$\|\mu_1 - \mu_2\| = \int (1+V(z))|\mu^1 - \mu^2|(\mathrm{d}z).$$

And we have constants A, γ depending in an explicit way on V, α, λ, C such that

$$\|f^1(t) - f^2(t)\| \leq Ae^{-\gamma t}\|f^1(0) - f^2(0)\|.$$

Harris's theorem has been used to show convergence to equilibrium for kinetic equations. In [94] they show convergence for kinetic Fokker-Planck equations using Harris's theorem. They do not verify the minorisation condition with a quantitative method. Therefore the end result is not quantitative. In [14] they show convergence to equilibrium using Harris's theorem for various scattering equations that includes equations similar to the ones studied in these chapters. Again it gives rates which are not quantitative. In [38] the authors use Doeblin's theorem to prove quantitative rates of convergence for some non-linear kinetic equations on the torus with a non-equilibrium steady state.

In this chapter we study three different kinetic equations with jumps. The linear relaxation Boltzmann equation both with a confining potential and on the torus, the linear Boltzmann equation on the torus and the kinetic non-local diffusion equation which is a non-local diffusion equation approximating the kinetic Fokker-Planck equation. We show convergence to equilibrium in a weighted total variation distance with quantitative rates. The weighting is comparable to $U(x) + |x|^2 + |v|^2$ where $U(x)$ is the confining potential. A similar method using Harris's theorem to get quantitative rates for jump equation has been used in [60, 32] to show convergence to equilibrium for equations modelling biological processes.

The precise results are the following.

Theorem 1.11. *The solutions to equation (1.5) on the flat torus without confining potential converge exponentially fast to equilibrium in total variation distance. This rate is explicitly calculable. i.e. There exists some $\lambda > 0$ and $C > 0$ such that*

$$\|f(t) - \mu\|_{TV} \leq Ce^{-\lambda t}\|f(0) - \mu\|_{TV},$$

where $f(t)$ is a solution to (1.5) at time t.

Theorem 1.12. *Suppose that $U(x)$ is a function satisfying*

$$|\nabla_x U(x)| \leq CU(x)^\eta$$

for some $\eta \in (0, 1)$ and

$$x \cdot \nabla_x U(x) \geq \gamma_1 |x|^2 + \gamma_2 U(x) - A$$

for strictly positive constants C, γ_1, γ_2, A and $\gamma_1 \leq 1$. Then the solution to (1.5) converges exponentially fast to equilibrium in a weighted total variation norm. More specifically there exists $C > 0$ and $\lambda > 0$ which we can calculate explicitly such that

$$\rho(\mathcal{P}_t \mu_1, \mathcal{P}_t \mu_2) \leq Ce^{-\lambda t}\rho(\mu_1, \mu_2),$$

where

$$\rho(\mu_1, \mu_2) = \int (1 + U(x) + \frac{1}{2}|v|^2 + \frac{1}{4}x \cdot v + \frac{1}{8}|x|^2)|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Furthermore as U is super quadratic at infinity ρ is equivalent to the distance weighted by the Hamiltonian

$$\tilde{\rho}(\mu_1, \mu_2) = \int (1 + H(x, v))|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Remark. *We have a tentative proof which allows us to remove the condition that*

$$|\nabla_x U(x)| \leq CU(x)^\eta$$

for some $\eta \in (0, 1)$ and replace it with the assumption that U is C^2.

Theorem 1.13. *Suppose that $U(x)$ is a function satisfying*

$$|\nabla_x U(x)| \leq U(x)^\eta, \quad x \cdot \nabla_x U(x) \geq \gamma_1 \langle x \rangle^\beta + \gamma_2 U(x) - A.$$

Where

$$\langle x \rangle = 1 + |x|^2,$$

and $\beta \in (0,1)$*. Then the solution to* (1.5) *converges to equilibrium in a weighted total variation norm in the following way. We define the function* M *by*

$$M(x,y) = U(x) + \frac{1}{2}|v|^2 + \frac{1}{4}x \cdot v + \frac{1}{8}|x|^2.$$

Then there exits a constant $C > 0$ *such that*

$$\|\mathcal{P}_t\delta_{z_1} - \mathcal{P}_t\delta_{z_2}\|_{TV} \leq C(M(z_1) + M(z_2))(1+t)^{-1/(1-\beta)},$$

and

$$\|\mathcal{P}_t\delta_z - \mu\|_{TV} \leq CM(z)(1+t)^{-1/(1-\beta)} + C(1+t)^{-\beta/(1-\beta)}.$$

Theorem 1.14. *If* $f(t)$ *is the solution to the linear Boltzmann equation,* (1.9)*, for Maxwell molecules with cut off and* b *bounded below then there exists* $C > 0$ *and* $\lambda > 0$*, which we can compute explicitly, such that*

$$\rho(\mathcal{P}_t\mu_1, \mathcal{P}_t\mu_2) \leq Ce^{-\lambda t}\rho(\mu_1,\mu_2),$$

where

$$\rho(\mu_1,\mu_2) = \int (1+|v|^2)|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Remark. *Our hope is to extend this to the confining potential case before submitting the paper.*

We also look at what we call the kinetic non-local diffusion equation

$$\partial_t f + v \cdot \nabla_x f - \nabla_x U \cdot \nabla_v f = K * f - f + \nabla_v \cdot (vf). \tag{1.24}$$

Here K is smooth radial and compactly supported.

Theorem 1.15. *The solution to* (1.24) *on the torus converges exponentially fast in a weighted TV distance. Specifically there exists* $C > 0$ *and* $\lambda > 0$*, which we can compute explicitly, such that*

$$\rho(\mathcal{P}_t\mu_1, \mathcal{P}_t\mu_2) \leq Ce^{-\lambda t}\rho(\mu_1,\mu_2),$$

where

$$\rho(\mu_1,\mu_2) = \int (1+|v|^2)|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Theorem 1.16. *Suppose that* $U(x)$ *is a function satisfying*

$$|\nabla_x U(x)| \leq CU(x)^\eta$$

for some $\eta \in (0,1)$ *and*

$$x \cdot \nabla_x U(x) \geq \gamma_1|x|^2 + \gamma_2 U(x) - A$$

for positive constants and $\gamma_1 \leq 1$*. Then the solution to* (1.24) *converges exponentially fast to equilibrium in a weighted total variation norm. More specifically there exists* $C > 0$ *and* $\lambda > 0$*,* *which we can calculate explicitly, such that*

$$\rho(\mathcal{P}_t\mu_1, \mathcal{P}_t\mu_2) \leq Ce^{-\lambda t}\rho(\mu_1, \mu_2),$$

where

$$\rho(\mu_1, \mu_2) = \int \left(1 + U(x) + \frac{1}{2}|v|^2 + \frac{1}{2}x \cdot v + \frac{1}{4}|x|^2\right) |\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Furthermore if V is super quadratic at infinity (which is implied by earlier assumptions) then ρ is equivalent to the distance weighted by the Hamiltonian

$$\tilde{\rho}(\mu_1, \mu_2) = \int (1 + H(x, v))|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Theorem 1.17. *Suppose that $U(x)$ is a function satisfying for some $\eta \in (0, 1)$*

$$|\nabla_x U(x)| \leq U(x)^\eta, \quad x \cdot \nabla_x U(x) \geq \gamma_1 \langle x\rangle^\beta + \gamma_2 U(x) - A,$$

where

$$\langle x\rangle = 1 + |x|^2,$$

and $\beta \in (0, 1)$. Then the solution to the non-local diffusion equation converges to equilibrium in a weighted total variation norm in the following way. We define the function M by

$$M(x, y) = U(x) + \frac{1}{2}|v|^2 + \frac{1}{2}x \cdot v + \frac{1}{4}|x|^2.$$

Then there exits a constant $C > 0$, explicitly computable, such that

$$\|\mathcal{P}_t\delta_{z_1} - \mathcal{P}_t\delta_{z_2}\|_{TV} \leq C(M(z_1) + M(z_2))(1 + t)^{-1/(1-\beta)},$$

and

$$\|\mathcal{P}_t\delta_z - \mu\|_{TV} \leq CM(z)(1 + t)^{-1/(1-\beta)} + C(1 + t)^{-\beta/(1-\beta)}.$$

Perspectives

This method works very well for kinetic equations with jumps. It seems hopeful that it could be applied to many more models. It seems likely that these ideas could be applied to relaxation equations with spatially inhomogeneous jump rates. For example it is an open problem to show quantitative rates of convergence to equilibrium for the equations

$$\partial_t f_v \cdot \nabla_x f = \sigma(x) \left(\int f(t, x, u)\mathrm{d}u - f\right), \quad x \in \mathbb{T}^d, a \leq |v| \leq A$$

Where σ is a function which vanishes at more than isolated points. It is known [16, 15] that this equation will converge exponentially fast to equilibrium in L^1 provided σ satisfies the *geometric control condition* but there is no quantitative rate.

Another possible application would be to look at the linear Boltzmann equation with soft potentials. In this chapter we look at the hard potential case and with cut off. There is a version of Harris's theorem which will give sub geometric rates of convergence given a weaker Lyapunov condition. It seems likely that we could apply this to the linear Boltzmann equation with cut off and soft potentials to see rates of convergence.

It would be interesting to try and extend this to diffusion equations via a limiting process. The kinetic non-local diffusion equation is

$$\partial_t f + v \cdot \nabla_x f = \lambda(K * f - f) + \nabla_v \cdot (vf).$$

We can choose a sequence of constants λ_n and smooth functions K_n such that solutions to this equations converge to solutions of the kinetic Fokker-Planck equation. If we can apply our method with constants that don't depend on n then we could use this to get rates of convergence for the kinetic Fokker-Planck equation. In the chapter after this one we give another way of applying these ideas to the kinetic Fokker-Planck equation.

Hypocoercivity for the kinetic Fokker-Planck equation with a confining potential via Hairer and Mattingly's Wasserstein-2 Harris theorem and Malliavin calculus

This chapter studies convergence to equilibrium for the kinetic Fokker-Planck equation with a confining potential. This work is not in collaboration and has not been submitted for publication.

There are strong mathematical connections between hypocoercivity and hypoellipticity because of these the original goal of this project was to see if it is possible to understand anything about hypocoercivity for SDEs using Malliavin calculus [92, 103, 104]. In some sense the Malliavin derivative can be thought of as a way of differentiating the solution to an SDE in terms of the driving Brownian motion. We write the Malliavin derivative of Z as $\mathcal{D}Z$. $\mathcal{D}Z$ is a function so we write its value at time s as $\mathcal{D}_s Z$. The Clarke-Ocone formula is

$$Z_t = \mathbb{E}(Z_t) + \int_0^t \mathbb{E}(\mathcal{D}_s Z_t | \mathscr{F}_s) \mathrm{d}W_s.$$

This is a Malliavin calculus version of the fundamental theorem of calculus. Here $\mathscr{F}$ is the filtration or the Brownian motion. Now if $Z_t = \mathbb{E}(Z_t) + G_t$ where G_t is a (possibly degenerate) Gaussian, if $\mathbb{E}(Z_t)$ is sufficiently well behaved then we can show a uniform minorisation iff G_t is non-degenerate. We can think of this as saying that uniform minorisation would happen in this situation if and only if the law of Z_t is spreading out in all direction. Now if Z satisfies the SDE

$$\mathrm{d}Z_t = B(Z_t)\mathrm{d}t + A(Z_t)\mathrm{d}W_t,$$

we can show that $\mathcal{D}_s X_t$ satisfies an SDE (differentiating in s) where the commutators between A, B appear in the coefficients. In the context of the kinetic Fokker-Planck equation this means that we can Taylor expand to get that

$$\mathcal{D}_s \begin{pmatrix} X_t \\ V_t \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - (t-s) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + E_{s,t}.$$

Here $E_{s,t}$ is an error which is order $(t-s)^2$. Combining this with the Clarke-Ocone formula we can see that (X_t, V_t) behaves like a Gaussian centred around its expectation up to a small error, and this Gaussian is non-degenerate. However, whilst the error is small in the sense that

$$\mathbb{E}\left(\left(\int_0^t \mathbb{E}(E_{s,t} | \mathscr{F}_s) \mathrm{d}W_s \right)^2 \right) \leq Ct^2,$$

as $t \to 0$ this bound is to weak too get a minorisation condition from this expansion. This means that instead of using Harris's theorem we use a version of Harris's theorem due to Hairer and Mattingly [72] which gives convergence in Wasserstein distance. This theorem combines a weaker minorisation condition with a Lyapunov structure. Our strategy allows us to verify the minorisation condition. However there is another assumption which is a point wise gradient bound on the semigroup. We verify this using hypocoercive type twisted norms in place of the Γ functional in Bakry-Emery style calculus.

Precisely we have the following theorem.

Theorem 1.18. *Suppose that $\mathcal{P}_t$ is a semigroup corresponding to the solution to the kinetic Fokker-Plank (1.2) with the confining potential V being a smooth function satisfying*

$$\mathit{Hess}(U)(x) \leq M, \quad x \cdot \nabla_x U(x) \geq c_1 U(x) + c_2 x^2 - c_3$$

for some constants M, c_1, c_2, c_3. Then we can choose constants a_* and k depending on these other constants to define the function

$$L(x, v) = \exp\left(a_*\left(v^2 + 2U(x) + 2kx^2 + kxv\right)\right).$$

We define ρ corresponding to L with

$$\rho(z_1, z_2) = \inf_{\gamma \in \Gamma} \int_0^1 L(\gamma(t)) \|\dot{\gamma}(t)\| \mathrm{d}t.$$

Here Γ is the set of all C^1 paths between z_1 and z_2. Then if $\mathcal{W}_\rho$ is the Wasserstein-1 distance associated to ρ we have constants $C > 0$ and $\lambda > 0$, which we can compute explicitly, such that

$$\mathcal{W}_\rho(\mathcal{P}_t \mu, \mathcal{P}_t \nu) \leq C e^{-\lambda t} \mathcal{W}_\rho(\mu, \nu).$$

Perspectives

The most natural next step from this proof would be to try and extend it to a wider class of kinetic equations. The first step would be to try and remove the assumption that the Hessian of the confining potential is bounded. The proof relies very strongly on the fact that the first two vector fields appearing in the Taylor expansion of the Malliavin derivative are constant. Removing this assumption and the bounded Hessian assumption would be the first step towards finding general sufficient conditions on A, B for an SDE

$$\mathrm{d}Z_t = B(Z_t)\mathrm{d}t + A(Z_t)\mathrm{d}W_t$$

to be hypocoercive in Wasserstein. In particular such a theorem might allow us to look at anharmonic chains of oscillators or similar systems.

Hypocoercivity in Φ-entropy for the linear relaxation Boltzmann equation

This chapter is adapted from the paper [58]. This paper is submitted for publication.

As we have seen relative entropy is a major tool for showing convergence to equilibrium for kinetic equations. Hypocoercivity in relative entropy was first shown by Villani in section 6, part

1 of [116]. This result holds for operators of the form

$$\sum_i A_i^* A_i + B,$$

where the A_i are first order derivations and the conjugate is taken in $L^2(\mu)$ for some probability measure μ. Therefore the result holds for degenerate diffusion type equations. In the context of kinetic equations the main example is the kinetic Fokker-Planck equation. Whilst most hypocoercivity theory has been done in $L^2(\mu^{-1}), H^1(\mu^{-1})$ there are several motivations to try and push the theory in the context of relative entropy.

- We can enlarge the space of initial data for which we can show exponentially fast convergence to equilibrium. If we show a result for $f \in L^2(\mu^{-1})$ then we are constrained to work with initial data in $L^2(\mu^{-1})$. This means that f_0 must decay very fast at infinity. However, if $\mu = \exp(-|v|^2/2 + U(x))$ then we have

  $$H_\mu(f) = \int f \log(f/\mu) \mathrm{d}x\mathrm{d}v = \int f \log(f) \mathrm{d}x\mathrm{d}v + \int f(|v|^2/2 + U(x)) \mathrm{d}x\mathrm{d}v.$$

  Similarly, for Fisher information we have

  $$I_\mu(f) \leq I(f) + \int f|\nabla(|v|^2/2 + U(x))|^2 \mathrm{d}v\mathrm{d}x.$$

  So these quantities will be finite provided we have some moment bounds (depending on $U(x)$) and finite entropy and Fisher information. This is true for many distributions which decay only polynomially at infinity.

- If we want to eventually study non-linear equations then it is often the case that strong spaces like $L^2(\mu^{-1})$ will not be a natural space for the equation. For initial data which is neither small nor close to the Maxwellian there is no well posedness theory for the Boltzmann equation in Hilbert spaces weighted against the equilibrium This problem is solved in the context of the Boltzmann equation by combining linearised theory with enlarging the space of solutions [64] and Desvillettes-Villani results to show when the solution will enter the linearised regime.

- The relative entropy and relative Fisher information functionals behave well with respect to the dimension of the phase space that the equation is set in. Furthermore, constants in this equation often depend on constants in the logarithmic Sobolev inequality which can often be shown to behave well with dimension. More specifically, suppose that $F_N = f^{\otimes N}$ then we have

  $$H(F_N) = \int f^{\otimes N}(z) \sum_i \log(f(z_i)) \mathrm{d}z = \sum_i \int f(z_i) \log(f(z_i)) \mathrm{d}z_i = NH(f).$$

  We can also show that if $\Pi_1(F_N)$ is its first marginal, and the particles are indistinguishable then

  $$H(\Pi_1(F_N)) \leq \frac{1}{N} H(F_N).$$

Therefore, if we know that for all N that

$$H(F_N(t)) \leq Ce^{-\lambda t} H(F_N(0)),$$

then we have that

$$H(\Pi_1(F_N(t))) \leq \frac{C}{N} e^{-\lambda t} H(F_N(0)).$$

Furthermore if $F_N(0)$ is a tensor product or similar we will have

$$H(\Pi_1(F_N(t))) \leq Ce^{-\lambda t}$$

where C does not depend on N. Therefore the rates of convergence to equilibrium are uniform in N. On the other hand for L^2 the distance $\|F_N\|_2$ behaves like $\|\Pi_1 F_N\|_2^N$. So if we try the same computation we get that

$$\|\Pi_1 F_N(t)\|_2 \leq Ce^{-\lambda t/N}.$$

This effect becomes particularly important if one wishes to study particle systems and derive convergence results which are uniform in the number of particles. Entropic hypocoercivity has been used in [90] to show convergence to the limit equation for oscillator chains.

Hypocoercivity in entropy has been studied by several authors. There is a series of works looking at sharp rates for diffusions with linear drifts. This looks at functionals which are just Fisher information with no entropy term and exploits some nice cancellations between mixed terms which can be seen in [7, 3, 98, 6]. There is a work [41] which extends Villani's proof of hypocoercivity for the kinetic Fokker-Planck equation to a much wider class of confinement potential. Another is [49] which looks at the kinetic Fokker-Planck equation in p-entropies, which interpolate between L^2 and relative entropy and shows improved rates in these distances. There are also several works which look at point wise bounds on the semigroups rather than integrated estimates using similar calculations to those of H^1 hypocoercivity. These then give results in relative entropy as well as other entropies and Wasserstein-2 distance [12, 98, 11]. All these works look at diffusion equations which can be written in the form

$$A^* A + B$$

as in [116]. We study the linear relaxation Boltzmann equation on the torus

$$\partial_t f + v \cdot \nabla_x f = \Pi_{\mathcal{M}} f - f.$$

Theorem 1.19. *If f is a solution to the equation above with initial data f_0 such that*

$$I(f_0|\mu) < \infty, \qquad f_0 \in W^{1,1}(\mu),$$

then there exist constants $\Lambda > 0$ and $\alpha > 0, \beta > 0$, which we can compute explicitly, depending on λ but not on the dimension such that

$$I(f_t|\mu) + \beta H(\Pi_{\mathcal{M}} f_t|\mu) \leq \exp(-\Lambda t)\left(\alpha I(f_0|\mu) + 2\beta H(\Pi_{\mathcal{M}} f_0|\mu)\right).$$

This implies that for some γ,

$$H(f_t|\mu) \leq \exp(-\Lambda t)\left(\gamma I(f_0|\mu)\right).$$

Hypocoercivity for the linear Boltzmann equation has already been shown in [102] and [77]. The L^2 result implies the decay of relative entropy as $L^2(\mu^{-1})$ controls relative entropy. However, this theorem holds only for initial data in $L^2(\mu^{-1})$ which is a distinct set to having finite relative entropy and Fisher information. Also, this proof is directly in this non-linear distances. We also show that this result holds when relative entropy is replaced with p-entropies. Here we also show new functional inequalities which allow us to prove this convergence behaviour.

We have added some sections after the original submitted paper. We give an alternative way of calculating the dissipation of Fisher information which is simpler and allows us to extend our proof to a wider class of Φ-entropy. We then briefly review the paper [99] which extends the results of the paper to the unconfined setting with close to quadratic force. We then show similar calculations allow us to give a result for the linear Boltzmann equation with cut off, Maxwell molecules and close to quadratic confinement.

Perspectives

The natural next step would be to try and show entropic hypocoercivity for the linear relaxation Boltzmann equation with more general confining potentials. This is potentially quite challenging as the proof here resembles the proof in [102] which is in the H^1 setting. This type of proof does not allow us to deal with confining potentials. Hypocoercivity in the confining potential case is shown using L^2 hypocoercivity theorems which are more structurally different to the proof given here than [102]. It would also be interesting to try and extend the proof on the torus to other operators for which Fisher Information is a Lyapunov functional in the spatially homogeneous case such as the linear Boltzmann equation for Maxwell molecules.

More speculatively, as discussed already, showing entropic hypocoercivity is most relevant when trying to study equations when we want control uniformly in the dimension of the phase space. An example of this is oscillator chains. Here entropic hypocoercivity techniques have been used to show hydrodynamic limits in [90]. Entropic hypocoercivity in this context is more challenging because these systems have non-equilibrium steady states about which not much is known. Performing hypocoercivity for these systems requires one to prove bounds on the derivative of the steady state and functional inequalities for this state.

For some non-linear equations entropy-entropy production inequalities only hold for certain Φ-entropies and not in L^2. In this case the only hope of proving hypocoercivity for the fully non-linear equation is to try and use Φ-entropic hypocoercivity.

Non-equilibrium steady states in Kac's model coupled to a thermostat

This is the only work in this thesis not directly related to hypocoercivity. In this chapter we study convergence to equilibrium for the Kac Master equation coupled to a thermostat. The chapter is adapted from the paper [57]. This work has been published in *The Journal of Statistical Physics.*

The model is

$$\partial_t F_n = -\lambda N(I-Q)[F_N] - \mu \sum_{j=1}^{N}(I-R_j)[F_N] = \mathcal{L}[F_N], \tag{1.25}$$

where

$$Q[F_N] = \frac{1}{\binom{N}{2}} \sum_{i<j} \fint_0^{2\pi} F_N(v_{i,j}(\theta))\mathrm{d}\theta,$$

and

$$R_j[F_N] = \int \mathrm{d}w \fint_0^{2\pi} \mathrm{d}\theta g(w_j^*)F_N(v_j(w,\theta)).$$

In these

$$\begin{aligned} v_{ij}(\theta) &= (v_1, \dots, v_i\cos(\theta) + v_j\sin(\theta), \dots, -v_i\sin(\theta) + v_j\cos(\theta), \dots, v_N), \\ v_j(w,\theta) &= (v_1, \dots, v_j\cos(\theta) + w\sin(\theta), \dots, v_N), \\ w_j^* &= w\cos(\theta) - v_j\sin(\theta). \end{aligned}$$

This corresponds to a simple 1D model of N gas particles colliding with each other and particles in an infinite, unchanging heat bath. Each particle in the heat bath has velocity distributed according to the density g. This work is fundamentally motivated by two others. Firstly [26] which studies the Kac master equation coupled to a Gaussian thermostat. In this paper they show convergence to Gaussian equilibrium in relative entropy. It is part of a series of papers where they aim to study Kac's model when a large proportion of the particles are already at equilibrium. The other paper is [39] where they study the existence and convergence to non-equilibrium steady states for the spatially homogeneous Boltzmann equation coupled to a non-equilibrium thermostat. We study Kac's model with the same non-equilibrium coupling. A 'physical' system this might correspond to is if the gas could interact with a number of different thermostats at different temperatures. We show existence and convergence to a non-equilibrium steady state. We work in both the Gabetta-Toscani-Wennberg distance which is what is used in [39] and Wasserstein-2 using a simple coupling strategy. This allows us to give both a 'deterministic' and a 'probabilistic' proof of convergence to equilibrium for the system. We then study the behaviour of the system as the number of particles goes to infinity. We show that both distances behave well with respect to the ambient dimension and this allows us to get uniform convergence rates for the first marginal. Lastly, we add a short section to the original paper which shows how the result of Hauray in [75] can be translated into a result with Fourier distances.

The precise results for the fixed N model are the following.

Theorem 1.20. *A steady state for the master equation exists, is unique and has the same moments up to order 2 as $g^{\otimes N}$.*

Furthermore if we start with initial data F_N^0 and H_N^0 which are probability distributions on $\mathbb{R}^N$ with finite first and second moments then we have the following possible situations:

1. If F^0 and H^0 have the same mean initially then the GTW distance between the solutions is finite for all time and we get the exponential convergence:

$$d_{GTW,N}(F_N(t), H_N(t)) \le e^{-\mu t/2} d_{GTW,N}(F_N^0, H_N^0).$$

2. If F^0 and H^0 have different means then we can construct an altered distance in which the solutions still converge exponentially fast towards each other with rate $\mu/2$. We also have the estimate

$$d_{T1,N}(F_N(t), H_N(t)) \leq e^{-\mu t/4} d_{T1,N}(F_N^0, H_N^0).$$

Theorem 1.21. *If μ_N and ν_N are two solutions to the master equation with finite second moments then*

$$\mathcal{W}_2(\mu_N(t), \nu_N(t)) \leq e^{-\mu t/2} \mathcal{W}_2(\mu_N(0), \nu_N(0)).$$

As we let $N \to \infty$ we have the following theorems.

Theorem 1.22. *Let F_N^0 and H_N^0 be respectively f and h chaotic families where the GTW distance between F_N^0 and $f^{\otimes N}$ (resp. for H_N^0 and $h^{\otimes N}$) is bounded uniformly in N. Furthermore if f and h are probability densities with finite first and second moments and differentiable Fourier transforms, then we can choose a family of functions χ (one for each N) to construct an altered distance $\tilde{d}$ so that*

$$\tilde{d}\,(\Pi_1[F_N], \Pi_1[H_N])) \leq (C_1 + (C_2 + C_3)\sqrt{N} + \tilde{d}(f, h))e^{-\frac{\mu}{2}t}.$$

We also look at the $T1$ distance which is very similar to the GTW distance it is

$$d_{T1}(\mu, \nu) = \sup_{\xi \neq 0} \frac{|\hat{\mu}(\xi) - \hat{\nu}(\xi)|}{|\xi|}.$$

Theorem 1.23. *Suppose that f and h are probability densities on $\mathbb{R}$ with finite mean. Suppose $(F_N(0, v))_{N \geq 2}$ and $(H_N(0, v))_{N \geq 2}$ are respectively f, h-chaotic families with respect to the $T1$ metric, and the $T1$ distance between $F_N(0, \cdot)$ and $f^{\otimes N}$, and between $H_N(0, \cdot)$ and $h^{\otimes N}$ are bounded uniformly in N. Furthermore, let F_N, H_N be the solution to the N-particle coupled Kac's master equation with this initial data. Then there exists a C (the bound between the initial data and the tensorised form) independent of N such that*

$$d_{T1,1}(\Pi_1[F_N](t), \Pi_1[H_N](t)) \leq (C + \sqrt{N} d_{T1,1}(f, h))e^{-\mu t/4}.$$

Theorem 1.24. *Suppose that $\mu_N(t)$ and $\nu_N(t)$ are solutions to the master equation at time t, with initial data $\mu_0^{\otimes N}$ and $\nu_0^{\otimes N}$. Then we have that for any N,*

$$\mathcal{W}_{2,1}(\Pi_1(\mu_N(t)), \Pi_1(\nu_N(t))) \leq e^{-\mu t/2} \mathcal{W}_{2,1}(\mu_0, \nu_0).$$

Perspectives

Since this work was completed, other works have been done on Kac's model coupled to a thermostat. This work is a route to studying Kac's model when most of the system is in equilibrium [113, 25]. It has also been shown that the Kac semigroup is not contractive in GTW [112]. These papers all deal with the case when the thermostat is Maxwellian and you have equilibrium steady states. It would be interesting to see if the last section of this chapter which translates the results of [75] into a Fourier distance can be extended to the multidimensional Kac model.

Chapter 2

Hypocoercivity via coupling for the kinetic Fokker-Planck equation on the torus

2.1 Introduction

In this chapter we study the kinetic Fokker-Planck equation on the torus. We prove contraction properties of the spatially periodic kinetic Fokker-Planck equation in the Wasserstein metric, and show to what extent the probabilistic technique of coupling can be used in such situations. This is of interest, both intrinsically, and in the broader context of analytic and probabilistic methods of proving convergence to equilibrium and contraction properties of Fokker-Planck equations which we summarise in the paragraphs below. Since this paper was originally written the paper [52] appeared which shows a similar result in a similar but more challenging setting. They deal with the kinetic Fokker-Planck equation with a confining potential. It seems likely their techniques would adapt in a straightforward way to the situation studied here. They perform a similar change of variables as to the one given in the Markovian section of this chapter and use a combination of reflection and synchronisation couplings as is also used here. The Monge-Kantorovich-Wasserstein (MKW) distance comes from optimal transport and is defined as

$$\mathcal{W}_2(\mu,\nu) = \inf_{\pi \in \Pi_{\mu,\nu}} \left(\int |x-y|^2 \mathrm{d}\pi(x,y) \right)^{1/2},$$

where $\Pi_{\mu,\nu}$ is the set of all couplings between μ and ν.

A common analytic technique to show contraction or convergence to equilibrium of Fokker-Planck equations is to work in a L^2 space weighted by the reciprocal of the equilibrium measure. Here, in the spatially homogeneous setting, contractivity is established by showing that the generator of the Fokker-Planck semi-group is *coercive* on this L^2 space, which implies that the generator has a spectral gap. In the spatially inhomogeneous setting, which is common in kinetic theory, the generator is, however, not *coercive* in this space and this method fails.

The kinetic Fokker-Planck equation in particular has received much attention [80, 61, 102] both in the case of a spatial confining potential and in, the analytically simpler, case of spatial periodicity. The paper [61] considers exactly our equation and finds explicitly the optimal rates of convergence in weighted L^2 space. The motivation is similar to that of this paper, which is to study a simple toy model on which more explicit calculations can be performed in order to explore alternative methods for proving hypocoercivity. These works, however, do not address the question of contraction in the Wasserstein metric $\mathcal{W}_2$, as this distance is currently inaccessible from these analytic tools; the closest result to this being [97] where $\mathcal{W}_1$ results are obtained by duality. Using interpolation estimates and convergence results in other spaces, one can conclude exponential decay in the Wasserstein $\mathcal{W}_2$ distance. However, then the control in terms of the initial data only holds for a power strictly less than one.

Another viewpoint, strongly related to the first, comes from the theory of gradient flows [84], in which the Fokker-Planck equation is identified with the steepest descent flow of an entropy functional in the Wasserstein space $\mathcal{W}_2$. However, the theory does not cover the considered model due to the kinetic structure. Dissipation in the Wasserstein distance can also be shown for non-gradient drifts in the homogeneous setting using analytic methods [21].

A common probabilistic technique to show contraction or convergence is to construct a *coupling* between two copies of the stochastic process that realises the desired bound on the metric between the laws. In the spatially homogeneous Fokker-Planck equation, the *synchronisation* coupling, where the infinitesimal motions of the noise are coupled together, gives contraction in Wasserstein metrics when the velocity potential is strongly convex. In the spatially inhomogeneous case with a confining potential, such a straightforward coupling only establishes contraction if the confining potentials are quadratic (or a small perturbation thereof) see for example [22]. Establishing contraction in the Wasserstein metric for more general confining potentials is an open problem. In the spatially periodic case results are even more limited. In this case the synchronisation coupling does not cause the spatial distance on the torus to decay. Thus the spatially periodic case is more difficult in the probabilistic case. This is in contrast to the analytic setting, where having the spatial variable on the torus means hypocoercivity can be shown by a very similar, and in fact slightly simpler, computation to that in part 1 section 7 of [116] will show hypocoercivity.

In this work we study the contraction properties in the Wasserstein metric of the kinetic Fokker-Planck equation with spatial variable on the torus and a quadratic velocity potential. Despite the simplicity of this equation, to the authors' knowledge this question has not been answered in the literature, and a second goal of this manuscript it to understand what difficulties might explain this.

This kinetic Fokker-Planck equation describes the law of a particle moving in the phase space $\mathbb{T} \times \mathbb{R}$ whose location in the phase space is (X_t, V_t) and evolves as

$$\begin{cases} \mathrm{d}X_t = V_t \mathrm{d}t, \\ \mathrm{d}V_t = -\lambda V_t \mathrm{d}t + \mathrm{d}W_t, \end{cases} \tag{2.1}$$

where W_t is a Brownian motion and the spatial variable is in the torus $\mathbb{T} = \mathbb{R}/(2\pi L\mathbb{Z})$ of length $2\pi L$.

The corresponding law μ_t on $\mathbb{T} \times \mathbb{R}$ evolves as

$$\partial_t \mu_t + v \partial_x \mu_t = \partial_v [\lambda v \mu_t + \frac{1}{2} \partial_v \mu_t], \tag{2.2}$$

where this equation is considered in the weak sense. The equilibrium state for this equation is

$$\frac{1}{2\pi L} Leb \otimes \sqrt{\frac{\lambda}{\pi}} \exp\left(-\frac{1}{4\lambda} v^2\right) Leb.$$

Solving the stochastic evolution, we show exponential decay of the distance between two solutions.

Theorem 2.1. *If μ_t and ν_t are two solutions to the kinetic Fokker-Planck equation (2.2), then we have*

$$\mathcal{W}_2(\mu_t, \nu_t) \leq \left(e^{-\lambda t} + c\, e^{-t/(2\lambda^2 L^2)}\right) \mathcal{W}_2(\mu_0, \nu_0)$$

for a constant c only depending on L.

Remark. *We are not aware of any paper showing optimal rate of convergence for this process in $\mathcal{W}_2$. The paper [61] shows that for large times this is the optimal rate of convergence in a weighted L^2 space. Also, we show later that we can split the process in components which are broadly an Orstein-Uhlenbeck process with rate λ and a Brownian motion with diffusivity $1/\lambda$ on the torus. One would expect the optimal rate of convergence for an O-U process in any reasonable distance to be λ and the optimal rate of convergence for the diffusion process to be $1/(2\lambda^2 L^2)$. Therefore it seems likely that our rates are optimal.*

The key idea is that, after conditioning on the final velocity, the spatial variable has enough randomness left to allow such a coupling. This approach is not based on a functional inequality which is integrated over time.

In fact the evolution is not a contraction semigroup in the considered distance which we can show directly in a straightforward way using the explicit solution to the SDE. Precisely,

Proposition 2.1. *The kinetic Fokker-Planck operator is not coercive in the MKW distance. The inequality*

$$\mathcal{W}_2(\mu_t, \nu_t) \leq e^{-\gamma t} \mathcal{W}_2(\mu_0, \nu_0), \;\; \forall \mu_0, \nu_0$$

cannot hold for any $\gamma > 0$.

In order to construct a coupling showing convergence in the MKW distance, random variables (X_t^i, V_t^i) are constructed for $t \in \mathbb{R}^+$ and $i = 1, 2$ such that (X_t^1, V_t^1) has law μ_t and (X_t^2, V_t^2) has law ν_t. Then for $t \in \mathbb{R}^+$ the coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ gives an upper bound of the MKW distance $\mathcal{W}_2(\mu_t, \nu_t)$.

The fact that (2.1) is an evolution equation means that it could be considered more natural from a probabilistic viewpoint to consider couplings that evolve along the flow of the equation. This motivates us to look at couplings where (X_t^i, V_t^i) are continuous Markov processes with initial distribution μ_0 and ν_0, respectively, and whose transition semigroup is determined by (2.1). For such couplings we can consider a more restrictive class of couplings.

Definition 2.1 (co-adapted coupling)**.** *The coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ is co-adapted if, for $i = 1, 2$, under the filtration $\mathcal{F}$ that is generated by the coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$, the process (X_t^i, V_t^i) is a continuous Markov process whose transition semigroup is determined by (2.1).*

This is an important subclass of couplings, which contains many natural couplings, and an even more restrictive subclass is the class of Markovian couplings, where additionally the coupling itself is imposed to be Markovian. The existence and obtainable convergence behaviour under this restriction has already been studied in different cases, e.g. [89, 30, 42]. Note that the co-adapted coupling is equivalent to the condition that the filtration generated by (X_t^i, V_t^i) is immersed in the filtration generated by the coupling, which motivates Kendall [86] to call such couplings *immersed couplings*.

By adapting the reflection/synchronisation coupling, we can still obtain exponential convergence but with a loss in dependence on the initial data.

Theorem 2.2. *Given initial distributions μ_0 and ν_0, then there exists a co-adapted coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ such that*

$$\begin{aligned}\mathcal{W}_2(\mu_t, \nu_t) &\leq \left(\mathbb{E}\left[|X_t^1 - X_t^2|_{\mathbb{T}}^2 + (V_t^1 - V_t^2)^2\right]\right)^{1/2} \\ &\leq C\zeta(t)(\sqrt{\mathcal{W}_2(\mu_0, \nu_0)} + \mathcal{W}_2(\mu_0, \nu_0)),\end{aligned}$$

where

$$\zeta(t) = \begin{cases} e^{-\min(2\lambda, 1/(2\lambda^2 L^2))t} & 4L^2\lambda^3 \neq 1 \\ e^{-2\lambda t}(1+t) & 4L^2\lambda^3 = 1 \end{cases}$$

and C is a constant that depends only on λ and L.

Here we used the notation $|X_t^1 - X_t^2|_{\mathbb{T}}$ to emphasis that this is the distance on the torus $\mathbb{T}$. In fact the filtrations generated by (X^1, V^1) and (X^2, V^2) agree which Kendall [86] calls an equi-filtration coupling.

Remark. *This achieves the same exponential decay rate as the non-Markovian argument, except for the case $4L^2\lambda^3 = 1$, when the spatial and velocity decay rates coincide and we have an addition polynomial factor multiplied with the exponential.*

In general the loss in the dependence is necessary.

Theorem 2.3. *Suppose there exists a function $\alpha : \mathbb{R}^+ \mapsto \mathbb{R}^+$ and a constant $\gamma > 0$ such that for all initial distributions μ_0 and ν_0 there exists a co-adapted coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ such that*

$$\left(\mathbb{E}\left[|X_t^1 - X_t^2|_{\mathbb{T}}^2 + (V_t^1 - V_t^2)^2\right]\right)^{1/2} \leq \alpha(\mathcal{W}_2(\mu_0, \nu_0))e^{-\gamma t}.$$

Then there exists a constant C such that for $z \in (0, \pi L]$ we have the following lower bound on the dependence on the initial distance

$$\alpha(z) \geq C\sqrt{z}.$$

The idea is to focus on a drift-corrected position on the torus, which evolves as a Brownian motion. By stopping the Brownian motion at a large distance we can then prove the claimed lower bound.

This shows that a simple hypocoercivity argument on a Markovian coupling cannot work. Precisely, there cannot exist a semigroup P on the probability measures over $(\mathbb{T} \times \mathbb{R})^{\times 2}$, whose marginals behave like the solution of (2.1) and which satisfies $H(P_t(\pi)) \leq cH(\pi)e^{-\gamma t}$ for $H^2(\pi) = \int[(X^1 - X^2)^2 + (V^1 - V^2)^2]\mathrm{d}\pi(X^1, V^1, X^2, V^2)$. Otherwise, the Markov process associated to P would be a coupling contradicting 2.3.

2.2 Set up

The stochastic differential equation (2.1) has an explicit solution, when posed in $\mathbb{R}^2$. For clarity, when we are considering X to be in $\mathbb{R}$ rather than the torus we will denote it $\hat{X}$. The explicit solution is

$$\begin{aligned} \hat{X}_t &= \hat{X}_0 + \frac{1}{\lambda}(1-e^{-\lambda t})V_0 + \int_0^t \frac{1}{\lambda}(1-e^{-\lambda(t-s)})\mathrm{d}W_s, \\ V_t &= e^{-\lambda t}V_0 + \int_0^t e^{-\lambda(t-s)}\mathrm{d}W_s, \end{aligned} \tag{2.3}$$

where W_t is the common Brownian motion. In this we separate the stochastic driving as (A_t, B_t) given by the stochastic integrals

$$\begin{aligned} A_t &= \int_0^t \frac{1}{\lambda}(1-e^{-\lambda(t-s)})\mathrm{d}W_s, \\ B_t &= \int_0^t e^{-\lambda(t-s)}\mathrm{d}W_s, \end{aligned}$$

which evolve as a vector in $\mathbb{R}^2$ with the common Brownian motion W_t. By Itō's isometry (A_t, B_t) is a Gaussian random variable with covariance matrix $\Sigma(t)$ given by

$$\Sigma_{AA}(t) = \frac{1}{\lambda^2}\left[t - \frac{2}{\lambda}(1-e^{-\lambda t}) + \frac{1}{2\lambda}(1-e^{-2\lambda t})\right], \tag{2.4}$$

$$\Sigma_{AB}(t) = \frac{1}{\lambda^2}\left[(1-e^{-\lambda t}) - \frac{1}{2}(1-e^{-2\lambda t})\right], \tag{2.5}$$

$$\Sigma_{BB}(t) = \frac{1}{2\lambda}(1-e^{-2\lambda t}). \tag{2.6}$$

From this we calculate that the conditional distribution of A_t given B_t is a Gaussian with variance $\Sigma_{AA}(t) - \Sigma_{AB}^2(t)\Sigma_{BB}^{-1}(t)$ and mean given by

$$\mu_{A|B}(t,b) = \Sigma_{AB}(t)\Sigma_{BB}^{-1}(t)b.$$

We write $g_{A|B}$ for the conditional density of A given B and g_B for the marginal density of B. Hence

$$g(t,a,b) = g_{A|B}(t,a,b)g_B(t,b) \tag{2.7}$$

is the joint density of A and B.

The last part of the set up is the change of variables we will need for the Markovian coupling. We define new coordinates (Y, V) in $\mathbb{T} \times \mathbb{R}$ by taking the drift away

$$\begin{cases} Y = X + \dfrac{1}{\lambda}V, \\ V = V. \end{cases} \tag{2.8}$$

The motivation for this change is the explicit formulas found in (2.3) from which we see that Y is the limit as $t \to \infty$ of X_t without additional noise. In the new variables, (2.1) becomes

$$\begin{cases} \mathrm{d}Y_t = \dfrac{1}{\lambda}\mathrm{d}W_t, \\ \mathrm{d}V_t = -\lambda V_t \mathrm{d}t + \mathrm{d}W_t, \end{cases}$$

for the common Brownian motion W_t. Note that the motion of Y_t does not depend explicitly upon V_t and is a Brownian motion on the torus.

It remains to show that these new coordinates define an equivalent norm on $\mathbb{T}\times\mathbb{R}$. This follows from the triangle inequality and we have

$$|X^1 - X^2|_{\mathbb{T}} + |V^1 - V^2| \leq |Y^1 - Y^2|_{\mathbb{T}} + \left(1 + \frac{1}{\lambda}\right)|V^1 - V^2|$$

and the other direction is similar. Thus, the two norms are equivalent up to a constant factor that depends only on λ.

2.3 Non-Markovian Coupling

We wish to estimate how much the spatial variable will spread out over time. We will then use this to construct a coupling at a fixed time t which exploits the fact that a proportion of the spatial density is distributed uniformly. In order to do this we give a lemma on the spreading of a Gaussian density wrapped on the torus.

Lemma 2.1. *For $\sigma^2 > 2L^2\log(3)$ consider the Gaussian density h on $\mathbb{R}$ given by*

$$h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$$

and wrap it onto the torus $\mathbb{T}$, i.e. define the density Qh on $\mathbb{T}$ by

$$(Qh)(x) = \sum_{n\in\mathbb{Z}} h(x + 2\pi L n). \tag{2.9}$$

We have the following estimate on the spatial spreading

$$Qh(x) \geq \frac{\beta}{2\pi L}$$

where

$$1 - \beta = \frac{2e^{-\sigma^2/2L^2}}{1 - e^{-\sigma^2/2L^2}} \in (0, 1).$$

Proof. We define the Fourier transform of a function on $\mathbb{T}$ to be

$$(\mathcal{F}g)(k) = \int_{\mathbb{T}} e^{ikx/L} g(x)\mathrm{d}x,$$

where

$$\int_{\mathbb{T}} g(x)\mathrm{d}x = \int_0^{2\pi L} g(x)\mathrm{d}x.$$

By the definition of Q, the Fourier transform of Qh is given by

$$(\mathcal{F}Qh)(k) = \int_{\mathbb{T}} \sum_{n\in\mathbb{Z}} h(x + 2\pi L n)e^{ikx/L}\mathrm{d}x \qquad = \int_{\mathbb{R}} h(x)e^{ikx/L}\mathrm{d}x = \exp\left(-\frac{k^2\sigma^2}{2L^2}\right)$$

where we have used the well-known Fourier transformation of a Gaussian.

Writing Qh in terms of its Fourier series and subtracting the $k=0$ term, we have for any $x \in \mathbb{T}$

$$Qh(x) - \frac{\beta}{2\pi L} = \frac{1}{2\pi L} \sum_{|k|\geq 1} e^{-k^2\sigma^2/2L^2 - ikx/L} + \frac{1-\beta}{2\pi L}.$$

We want this to be positive. Therefore it is sufficient to show that

$$\left| \sum_{|k|\geq 1} e^{-k^2\sigma^2/2L^2 - ikx/L} \right| \leq 1-\beta.$$

We estimate the left hand side by

$$\left| \sum_{|k|\geq 1} e^{-k^2\sigma^2/2L^2 - ikx/L} \right| \leq 2 \sum_{k\geq 1} e^{-k\sigma^2/2L^2} = 1-\beta$$

where the final equality follows from summing the geometric series. □

We can now use this to construct a coupling at time t. We will use this coupling to prove exponential decrease in the Wasserstein distance.

Lemma 2.2. *Let $t \geq 0$, be large enough so the variance of $g_{A|B}$ is greater than $2L\log(3)$, and β be such that*

$$(Qg_{A|B})(t,a,b) \geq \frac{\beta}{2\pi L},$$

where $g_{A|B}$ is defined by (2.7) above. Let μ_t resp. ν_t be the distribution of the solution to the Fokker-Planck equation (2.2) with deterministic initial data $\mu_0 = \delta_{x_0^1, v_0^1}$ and $\nu_0 = \delta_{x_0^2, v_0^2}$ respectively, at time t. Then there exists a coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ between μ_t and ν_t satisfying

$$\mathbb{E}\left[(V_t^1 - V_t^2)^2\right] = e^{-2\lambda t}\left[(v_0^1 - v_0^2)^2\right]$$

and

$$\mathbb{E}\left[|X_t^1 - X_t^2|_{\mathbb{T}}^2\right] \leq 2(1-\beta)\left[|x_0^1 - x_0^2|_{\mathbb{T}}^2 + \frac{1}{\lambda^2}(v_0^1 - v_0^2)^2\right].$$

Proof. Let us construct such a coupling. Since we have seen that $g_{A|B}$ is Gaussian density with variance $\sigma^2 = \Sigma_{AA}(t) - \Sigma_{AB}^2(t)\Sigma_{BB}^{-1}(t)$, we can use 2.1 to split the distribution $Qg_{A|B}$ as

$$Qg_{A|B}(t,a,b) = \frac{\beta}{2\pi L} + (1-\beta)s(t,a,b).$$

Then by assumption s is again a probability density for the variable a on the torus $\mathbb{T}$. We now consider the torus as a subset of $\mathbb{R}$ and then $Qg_{A|B}$ and $1/2\pi L$ are probability density functions. Therefore, s is also probability density functions supported on $[0, 2\pi L]$. Let B be an independent random variable with density $g_B(t,b)$, let Z be an independent uniform random variable over $[0,1]$ and let U be an independent uniform random variable over the torus. Finally let S be a random variable on $\mathbb{R}$ with density $s(t,\cdot,B)$, viewed as a density function on $\mathbb{R}$, only depending on B.

With this define the random parts A^1, A^2 of X_t^1, X_t^2 as

$$\begin{aligned} A^1 =& 1_{Z\leq\beta}\left[U - x_0^1 - \frac{1}{\lambda}(1-e^{-\lambda t})v_0^1\right] + 1_{\beta>Z}S, \\ A^2 =& 1_{Z\leq\beta}\left[U - x_0^2 - \frac{1}{\lambda}(1-e^{-\lambda t})v_0^2\right] + 1_{\beta>Z}S. \end{aligned}$$

We then construct $(\hat{X}_t^1, V_t^1)$ defined by

$$\begin{aligned} \hat{X}_t^1 &= x_0^1 + \frac{1}{\lambda}(1-e^{-\lambda t})v_0^1 + A^1, \\ V_t^1 &= e^{-\lambda t}v_0^1 + B, \end{aligned}$$

and $(\hat{X}_t^2, V_t^2)$ defined by

$$\begin{aligned} \hat{X}_t^2 &= x_0^2 + \frac{1}{\lambda}(1-e^{-\lambda t})v_0^2 + A^2, \\ V_t^2 &= e^{-\lambda t}v_0^2 + B. \end{aligned}$$

We then construct X_t^i by wrapping $\hat{X}_t^i$ onto the torus (i.e. $X_t^i \in [0, 2\pi L)$ and $X_t^i \equiv \hat{X}_t^i \mod 2\pi L$). By construction the pairs (X^i, V^i) have the right laws so they form a valid coupling.

We find

$$\mathbb{E}\left[(V_t^1 - V_t^2)^2\right] = e^{-2\lambda t}\left[(v_0^1 - v_0^2)^2\right]$$

and

$$\mathbb{E}\left[|X_t^1 - X_t^2|_{\mathbb{T}}^2\right] = (1-\beta)\left[\left|x_0^1 - x_0^2 + \frac{1}{\lambda}(1-e^{-\lambda t})(v_0^1 - v_0^2)\right|_{\mathbb{T}}^2\right]$$

and we can use Young's inequality to find the claimed control. □

We now put these two lemmas together to prove 2.1, which states exponential convergence in the MKW $\mathcal{W}_2$ distance.

Proof of 2.1. We first show that we can reduce to working with deterministic initial conditions. We denote $\mu_t^{x,v}$ to be the law of the solution to the SDE initialized at (x, v). Suppose we know that

$$\mathcal{W}_2(\mu_t^{x_1,v_1}, \mu_t^{x_1,v_2}) \leq \omega(t) d((x_1, v_1), (x_1, v_2)).$$

Since, μ_t, ν_t are the laws of Markov processes we know that,

$$\mu_t(\phi) = \int\int \phi(y, u) \mathrm{d}\mu_t^{(x,v)}(y, u) \mathrm{d}\mu_0(x, v).$$

Hence given, π a coupling of μ_0, ν_0 we can construct a coupling of μ_t, ν_t by

$$\pi_t(\psi) =$$

$$\int \left(\int \psi((y_1, u_1), (y_2, u_2)) \mathrm{d}\mu_t^{(x_1,v_1)}(y_1, u_1) \mathrm{d}\nu_t^{(x_2,v_2)}(y_2, u_2)\right) \mathrm{d}\pi((x_1, v_1), (x_2, v_2)).$$

The couplings of this form are a subset of all the couplings of μ_t, ν_t therefore we can take the infimum over these couplings in order to bound the Wasserstein distance. Then given any coupling

π of initial measures μ_0, ν_0 we have

$$\begin{aligned}\mathcal{W}_2(\mu_t,\nu_t)^2 &\le \int_{(\mathbb{T}\times\mathbb{R})^2} \mathcal{W}_2(\mu_t^{x_1,v_1},\mu_t^{x_2,v_2})^2 \mathrm{d}\pi((x_1,v_1),(x_2,v_2)) \\ &\le \omega(t)^2 \int_{(\mathbb{T}\times\mathbb{R})^2} d((x_1,v_1),(x_2,v_2))^2 \mathrm{d}\pi((x_1,v_1),(x_2,v_2)).\end{aligned}$$

Then taking an infimum over π shows that this implies

$$\mathcal{W}_2(\mu_t,\nu_t) \le \omega(t)\mathcal{W}_2(\mu_0,\nu_0).$$

Given any initial points $((x_0^1,v_0^1),(x_0^2,v_0^2))$, we can use 2.2 to construct a coupling $((X_t^1,V_t^1),(X_t^2,V_t^2))$ of μ_t and ν_t. From explicitly calculating the variance of the distribution of $A|B$ using (2.4), (2.5), (2.6), we see that the variance grows asymptotically as t/λ^2. Hence by 2.1 we can choose β so that $1-\beta \to 0$ exponentially fast with rate $1/2\lambda^2 L^2$. This, combined with the control from the second lemma, shows that

$$\mathbb{E}\left[(V_t^1-V_t^2)^2 + |X_t^1-X_t^2|_{\mathbb{T}}^2\right] \le \left(e^{-2\lambda t} + ce^{-t/2\lambda^2L^2}\right)\left[(v_0^1-v_0^2)^2 + |x_0^1-x_0^2|_{\mathbb{T}}^2\right]. \qquad \square$$

The explicit solution also allows to prove that the evolution is not a contraction semigroup.

Proof of 2.1. We will prove the theorem by contradiction. Suppose $\gamma > 0$ and let $a \neq b$ be two distinct points on the torus. Consider the initial measures

$$\mu_0 = \delta_{x=a}\delta_{v=0}$$

and

$$\nu_0 = \delta_{x=b}\delta_{v=0}.$$

Then the distance is $\mathcal{W}_2(\mu_0,\nu_0) = |a-b|_{\mathbb{T}}$.

At time t the spatial distribution of μ_t and ν_t, interpreted in $\mathbb{R}$, is a Gaussian with variance Σ_{AA} which by the explicit formula 2.4 can be bounded as

$$\Sigma_{AA}(t) \le C_A t^2$$

for a constant C_A and $t \le 1$.

Hence for $d > 0$ and $t \le 1$ the spatial spreading is controlled as

$$\begin{aligned}\mu_t((\mathbb{T}\setminus[a-d,a+d])\times\mathbb{R}) &\le \frac{2\Sigma_{AA}(t)}{d\sqrt{2\pi}}\exp\left(\frac{-d^2}{2\Sigma_{AA}^2(t)}\right) \\ &\le C_1\frac{t^2}{d}\exp\left(-C_2\frac{d^2}{t^4}\right)\end{aligned}$$

for positive constants C_1 and C_2, where we have used the standard tail bound for the Gaussian distribution (see e.g. [100, Lemma 12.9]).

For any $d > 0$ small enough that $a \pm d$ and $b \pm d$ do not wrap around the torus, any coupling between μ_t and ν_t must transfer at least the mass

$$1 - \mu_t((\mathbb{T}\setminus[a-d,a+d])\times\mathbb{R}) - \nu_t((\mathbb{T}\setminus[b-d,b+d])\times\mathbb{R})$$

between $[a-d, a+d]$ and $[b-d, b+d]$.

Hence the Wasserstein distance is bounded by

$$\mathcal{W}_2^2(\mu_t, \nu_t) \geq (|a-b|_{\mathbb{T}} - 2d)^2 \left(1 - 2C_1 \frac{t^2}{d} \exp\left(-C_2 \frac{d^2}{t^4}\right)\right).$$

Taking $d = |a-b|_{\mathbb{T}} t^{3/2}$ for t sufficiently small, this shows that

$$\mathcal{W}_2^2(\mu_t, \nu_t) \geq |a-b|_{\mathbb{T}}^2 (1 - 2t^{3/2})^2 \left(1 - \frac{2C_1}{|a-b|_{\mathbb{T}}} \sqrt{t} \exp\left(-\frac{C_2 |a-b|_{\mathbb{T}}^2}{t}\right)\right).$$

However, for all small enough positive t, we have

$$(1 - 2t^{3/2})^2 > e^{-\gamma t/2}$$

and

$$\left(1 - \frac{2C_1}{|a-b|_{\mathbb{T}}} \sqrt{t} \exp\left(-\frac{C_2 |a-b|_{\mathbb{T}}^2}{t}\right)\right) > e^{-\gamma t/2}$$

contradicting the assumed contraction. For the second estimate we use $\exp(-c/t) \leq (1 + c/t)^{-1} = t/(c+t)$. □

2.4 Co-adapted couplings

2.4.1 Existence

For 2.2 we construct a reflection/synchronisation coupling using the drift-corrected positions Y_t^i. As the positions are on the torus we can use a reflection coupling until Y_t^1 and Y_t^2 agree. Afterwards, we use a synchronisation coupling which keeps $Y_t^1 = Y_t^2$ and reduces the velocity distance.

For a formal definition let $((X_0^1, V_0^1), (X_0^2, V_0^2))$ be a coupling between μ and ν obtaining the MKW distance (the existence of such a coupling is a standard result, see e.g. [117, Theorem 4.1.]).

We then define the evolution of this coupling in two stages. First, define (X_t^1, V_t^1) and (X_t^3, V_t^3) to be strong solutions to (2.1) with initial conditions $((X_0^1, V_0^1)$ and (X_0^2, V_0^2) respectively and driving Brownian motion W_t^1. Then we recall the definition of Y^i from (2.8), and define the stopping time $T := \inf\{t \geq 0 : Y_t^1 = Y_t^3\}$. Then we define a new process W_t^2 by

$$W_t^2 = \begin{cases} -W_t^1 & t \leq T, \\ W_t^1 - 2W_T^1 & t > T. \end{cases}$$

By the reflection principle, W^2 is a Brownian motion. We use this to define a new solution (X_t^2, V_t^2) to be the strong solution to (2.1) with driving Brownian motion W^2 and initial condition (X_0^2, V_0^2). Note now that $T = \inf\{t \geq 0 : Y_t^1 = Y_t^2\}$.

For the analysis we introduce the notation

$$M_t = Y_t^1 - Y_t^2,$$
$$Z_t = V_t^1 - V_t^2.$$

Then by the construction the evolution is given by

$$\mathrm{d}M_t = \frac{2}{\lambda} 1_{t\le T} \mathrm{d}W_t^1, \tag{2.10}$$
$$\mathrm{d}Z_t = -\lambda Z_t \mathrm{d}t + 2 \cdot 1_{t\le T} \mathrm{d}W_t^1, \tag{2.11}$$

where M_t evolves on the torus $\mathbb{T}$.

As a first step we introduce a bound for T.

Lemma 2.3. *The stopping time T satisfies*

$$\mathbb{P}(T > t|M_0) = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{2k+1} \exp\left(-\frac{(2k+1)^2}{2\lambda^2 L^2} t\right) \sin\left(\frac{(2k+1)|M_0|_{\mathbb{T}}}{2L}\right). \tag{2.12}$$

Proof. As M_t evolves on the torus, T is the first exit time of a Brownian motion starting at M_0 from the interval $(0, 2\pi L)$. See [100, (7.14-7.15)], from which the claim follow after rescaling to incorporate the $2/\lambda$ factor. □

Remark. *The second expression in (2.12) is obtained by solving the heat equation on $[0, 2\pi L]$ with Dirichlet boundary conditions and initial condition δ_{M_0}.*

Lemma 2.4. *There exists a constant C such that for any $t > 0$ the following holds*

$$\mathbb{P}(T > t|M_0) \le C|M_0|_{\mathbb{T}}(1 + t^{-1/2})e^{-t/(2\lambda^2 L^2)}. \tag{2.13}$$

Proof. Using (2.12) and the inequality $\sin(x) \le x$ for $x \ge 0$, we have

$$\begin{aligned}
\mathbb{P}(T > t|M_0) &\le \frac{4}{\pi} e^{-t/(2\lambda^2 L^2)} \sum_{k=0}^{\infty} \frac{|M_0|_{\mathbb{T}}}{2L} \frac{2k+1}{2k+1} e^{-4k^2 t/(2\lambda^2 L^2)} \\
&\le \frac{2}{\pi L} |M_0|_{\mathbb{T}} e^{-t/(2\lambda^2 L^2)} \left(1 + \int_0^{\infty} e^{-4u^2 t/(2\lambda^2 L^2)} \mathrm{d}u\right) \\
&= \frac{2}{\pi L} |M_0|_{\mathbb{T}} e^{-t/(2\lambda^2 L^2)} \left(1 + \sqrt{\frac{\pi}{8t/(\lambda^2 L^2)}}\right) \\
&\le C|M_0|_{\mathbb{T}}(1 + t^{-1/2})e^{-t/(2\lambda^2 L^2)}
\end{aligned}$$

where on the second line we have bounded the sum by an integral. □

Using these simple estimates, we now study the convergence rate of the coupling.

Lemma 2.5. *There exists a constants C such that for any $t \ge 0$ we have the bound*

$$\mathbb{E}\left[|M_t|_{\mathbb{T}}^2 + |Z_t|^2 \middle| (Z_0, M_0)\right] \le |Z_0|^2 e^{-2\lambda t} + \begin{cases} C|M_0|_{\mathbb{T}} e^{-2\lambda t} & 2\lambda < 1/(2\lambda^2 L^2) \\ C|M_0|_{\mathbb{T}}(1+t)e^{-2\lambda t} & 2\lambda = 1/(2\lambda^2 L^2) \\ C|M_0|_{\mathbb{T}} e^{-t/(2\lambda^2 L^2)} & 2\lambda > 1/(2\lambda^2 L^2). \end{cases}$$

Proof. Without loss of generality we may assume that Z_0 and M_0 are deterministic in order to avoid writing the conditional expectation.

Applying Itō's lemma, we find from (2.11) that

$$\mathrm{d}|Z_t|^2 = -2\lambda|Z_t|^2 \mathrm{d}t + 4 \cdot 1_{t\le T} Z_t \mathrm{d}W_t^1 + 2 \cdot 1_{t\le T} \mathrm{d}t.$$

After taking expectations we see that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}|Z_t|^2 = -2\lambda\mathbb{E}|Z_t|^2 + 2\mathbb{P}(t \le T). \tag{2.14}$$

By explicitly solving (2.14) and using 2.4, we obtain

$$\begin{aligned}\mathbb{E}|Z_t|^2 &= |Z_0|^2 e^{-2\lambda t} + 2e^{-2\lambda t}\int_0^t e^{2\lambda s}\mathbb{P}(s \le T)\,\mathrm{d}s \\ &\le |Z_0|^2 e^{-2\lambda t} + C|M_0|_{\mathbb{T}} e^{-2\lambda t}\underbrace{\int_0^t e^{(2\lambda - 1/(2\lambda^2 L^2))s}(1+s^{-1/2})\,\mathrm{d}s}_{=:I_t}.\end{aligned}$$

Let us bound I_t. As the integrand is locally integrable, we have for a constant C

$$I_t \le C\left(1 + \int_0^t e^{(2\lambda - 1/(2\lambda^2 L^2))s}\,\mathrm{d}s\right).$$

Here the $s^{-1/2}$ term can be bounded by 1 for $s > 1$ and for $s \le 1$ the additional contribution can be absorbed into the constant. To bound the remaining integral we consider three cases:

- $2\lambda < 1/(2\lambda^2 L^2)$: The integral (and I_t) are uniformly bounded, $I_t \le C$.
- $2\lambda = 1/(2\lambda^2 L^2)$: The integrand is equal to 1 and $I_t \le C(1+t)$.
- $2\lambda > 1/(2\lambda^2 L^2)$: The integrand grows and $I_t \le C(1 + e^{(2\lambda - 1/(2\lambda^2 L^2))t})$.

In each case we multiply I_t by $e^{-2\lambda t}$ to obtain the decay rate. In the first two cases this gives the dominant term with $|M_0|_{\mathbb{T}}$ (as opposed to $|Z_0|$) dependence, while in the last case it is lower order than the $e^{-t/(2\lambda^2 L^2)}$ decay we obtain from $\mathbb{E}|M_t|_{\mathbb{T}}^2$ below.

Next let us consider $\mathbb{E}|M_t|_{\mathbb{T}}^2$. Using the finite diameter of the torus we have the simple estimate

$$\mathbb{E}|M_t|_{\mathbb{T}}^2 \le \pi^2 L^2 \mathbb{P}(T > t).$$

For $t \ge 1$ (say), we can use 2.4, to obtain

$$\mathbb{E}|M_t|_{\mathbb{T}}^2 \le C|M_0|_{\mathbb{T}} e^{-t/(2\lambda^2 L^2)} \quad \text{for } t \ge 1.$$

This leaves the case when $t \le 1$ where (2.13) blows up. We instead use the martingale property of M_t. Without loss of generality we may assume that $M_0 \in [0, \pi L]$. Then as M_t is stopped at T we know that $M_t \in [0, 2\pi L]$ for all $t \ge 0$. Hence, for any $t \ge 0$,

$$\mathbb{E}|M_t|_{\mathbb{T}}^2 \le \mathbb{E}|M_t|^2 \le 2\pi L\mathbb{E}M_t = 2\pi L M_0 = 2\pi L|M_0|_{\mathbb{T}}$$

by the martingale property. Combining the $t \le 1$ and $t \ge 1$ estimates we have

$$\mathbb{E}|M_t|_{\mathbb{T}}^2 \le C|M_0|_{\mathbb{T}} e^{-t/(2\lambda^2 L^2)} \quad \text{for } t \ge 0.$$

This together with the bound for $\mathbb{E}|Z_t|^2$ provides the claimed bounds of the lemma and completes its proof. $\square$

Proof of 2.2. By the equivalence of the norms from (X, V) and (Y, V), we see that

$$\begin{aligned}
&\mathbb{E}\left(|X_t^1 - X_t^2|_{\mathbb{T}}^2 + |V_t^1 - V_t^2|^2\right) \leq \left(1 + \frac{1}{\lambda}\right)\mathbb{E}\left(|M_t|_{\mathbb{T}}^2 + |Z_t|^2\right) \\
&\leq C'\zeta(t)\mathbb{E}(|M_0|_{\mathbb{T}} + |Z_0|^2) \\
&\leq C\zeta(t)\mathbb{E}\left(\left(|X_0^1 - X_0^2|_{\mathbb{T}}^2 + |V_0^1 - V_0^2|^2\right)^{1/2} + \left(|X_0^1 - X_0^2|_{\mathbb{T}}^2 + |V_0^1 - V_0^2|^2\right)\right).
\end{aligned}$$

Here we used 2.5 to go between the first and second line, and to find the exponentially decreasing term ζ. The constants C and C' come from the constants in equivalence of norms. □

2.4.2 Optimality

In order to show 2.3, we focus on the drift-corrected positions Y_t^1 and Y_t^2 which behave like time-rescaled Brownian motion on the torus. For their quadratic distance we prove the following decay bound.

Proposition 2.2. *Suppose there exist functions $\alpha : (0, \pi L] \mapsto \mathbb{R}^+$ and $\zeta : [0, \infty) \mapsto \mathbb{R}^+$ with $\zeta \in L^1([0, \infty))$, such that, for any $z \in (0, \pi L]$ there exist two standard Brownian motions W_t^1 and W_t^2 taking values on the torus $\mathbb{T} = \mathbb{R}/(2\pi L\mathbb{Z})$ and both adapted to a common filtration such that $|W_0^1 - W_0^2| = z$, and for $t \in \mathbb{R}^+$ it holds that*

$$\mathbb{E}[|W_t^1 - W_t^2|_{\mathbb{T}}^2] \leq (\alpha(z))^2\zeta(t).$$

Then with a constant c only depending on L, the function α satisfies the bound

$$\alpha(z) \geq c\,\|\zeta\|_{L^1([0,\infty))}^{-1/2}\sqrt{z}.$$

From this 2.3 follows easily.

Proof of 2.3. Fix $z \in (0, \pi L]$ and consider the initial distributions $\mu = \delta_{X=0}\delta_{V=0}$ and $\nu = \delta_{X=z}\delta_{V=0}$. Between μ and ν, there is only one coupling and $\mathcal{W}_2(\mu, \nu) = z$.

If there exists a co-adapted coupling $((X_t^1, V_t^1), (X_t^2, V_t^2))$ satisfying the bound, then Y_{t/λ^2}^1 and Y_{t/λ^2}^2 are Brownian motions on the torus with a common filtration. Moreover,

$$\mathbb{E}[|Y_t^1 - Y_t^2|_{\mathbb{T}}^2] \leq C\,\mathbb{E}[|X_t^1 - X_t^2|_{\mathbb{T}}^2 + |V_t^1 - V_t^2|^2]$$

for a constant C only depending on λ. Hence we can apply 2.2 to find the claimed lower bound for α. □

For the proof of 2.2, we first prove the following lemma.

Lemma 2.6. *Given two Brownian motions W_t^1 and W_t^2 on the torus adapted to the same filtration, then there exists a numerical constant c such that*

$$\mathbb{E}[|W_t^1 - W_t^2|_{\mathbb{T}}^2] \geq c\,e^{-2t/L^2}\mathbb{E}[|W_0^1 - W_0^2|_{\mathbb{T}}^2].$$

Proof. The natural (squared) metric $|x-y|_{\mathbb{T}}^2$ on the torus is not a global smooth function of $x, y \in \mathbb{R}$

as it takes $x, y \bmod 2\pi L$. Therefore we introduce the equivalent metric

$$d_{\mathbb{T}}^2(x,y) = L^2 \sin^2\left(\frac{x-y}{2L}\right),$$

which is a smooth function of $x, y \in \mathbb{R}$. Moreover, the constants of equivalence are independent of L, i.e. there exist numerical constants c_1 and c_2 such that

$$c_1 |x-y|_{\mathbb{T}}^2 \leq d_{\mathbb{T}}^2(x,y) \leq c_2 |x-y|_{\mathbb{T}}^2.$$

Now consider H_t defined by

$$H_t = L \sin\left(\frac{W_t^1 - W_t^2}{2L}\right) \exp\left(\frac{[W^1 - W^2]_t}{4L^2}\right).$$

As W_t^1 and W_t^2 are Brownian motions, their quadratic variation is controlled as $[W^1 - W^2]_t \leq 4t$. By Itō's lemma

$$\mathrm{d}H_t = \frac{1}{2} \cos\left(\frac{W_t^1 - W_t^2}{2L}\right) \exp\left(\frac{[W^1 - W^2]_t}{4L^2}\right) \mathrm{d}(W^1 - W^2)_t.$$

Therefore we may bound the quadratic variation of H by

$$\begin{aligned}
[H]_t &= \int_0^t \frac{1}{4} \cos^2\left(\frac{W_t^1 - W_t^2}{2L}\right) \exp\left(\frac{[W^1 - W^2]_t}{2L^2}\right) \mathrm{d}[W^1 - W^2]_t \\
&\leq \int_0^t \exp\left(\frac{2t}{L^2}\right) \mathrm{d}t \\
&< \infty.
\end{aligned}$$

Therefore, as also $|H_0| \leq L$, the local martingale H_t is a true martingale and by Jensen's inequality

$$\mathbb{E}[|H_t|^2] \geq \mathbb{E}[|H_0|^2].$$

Using the equivalence of two metrics, we thus find the required bound

$$\begin{aligned}
\mathbb{E}[|W_t^1 - W_t^2|_{\mathbb{T}}^2] &\geq c_2^{-1} \mathbb{E}\left[|H_t|^2 \exp\left(-\frac{[W^1 - W^2]_t}{2L^2}\right)\right] \\
&\geq c_2^{-1} \mathbb{E}\left[|H_0|^2\right] \exp\left(-\frac{2t}{L^2}\right) \\
&\geq c_1 c_2^{-1} \mathbb{E}[|W_0^1 - W_0^2|_{\mathbb{T}}^2] \exp\left(-\frac{2t}{L^2}\right). \square
\end{aligned}$$

With this we approach the final proof.

Proof of 2.2. Fix $a \in (0,1)$, let $z \in (0, \pi L]$ be given, and by symmetry assume without loss of generality that $W_0^1 - W_0^2 = |W_0^1 - W_0^2| = z$. Then define the stopping time

$$T = \inf\{t \geq 0 : W_t^1 - W_t^2 \notin (az, \pi L)\}.$$

The distance can be directly bounded as

$$\mathbb{E}[|W_t^1 - W_t^2|_{\mathbb{T}}^2] \geq \mathbb{P}[T \geq t](az)^2.$$

As ζ is integrable, it must decay along a subsequence of times and thus T must be almost surely finite.

As W_t^1 and W_t^2, considered on $\mathbb{R}$, are continuous martingales, their difference is also a continuous martingale. By the construction of the stopping time, the stopped martingale $(W^1 - W^2)_{t\wedge T}$ is bounded by πL and the optional stopping theorem implies

$$\mathbb{P}[W_T^1 - W_T^2 = \pi L] = \frac{z - az}{\pi L - az}.$$

Since Brownian motions satisfy the strong Markov property, we find together with 2.6

$$\begin{aligned}
\mathbb{E}\int_0^\infty |W_t^1 - W_t^2|_{\mathbb{T}}^2 \mathrm{d}t &\geq \mathbb{E}\int_T^\infty |W_t^1 - W_t^2|_{\mathbb{T}}^2 \mathrm{d}t \\
&\geq \mathbb{P}[W_T^1 - W_T^2 = \pi L]\mathbb{E}\left[\int_T^\infty |W_t^1 - W_t^2|_{\mathbb{T}}^2 \,\mathrm{d}t \;\middle|\; W_T^1 - W_T^2 = \pi\right] \\
&\geq \mathbb{P}[W_T^1 - W_T^2 = \pi L] c\,(\pi L)^2 \int_0^\infty e^{-2t/L^2} \mathrm{d}t \\
&\geq \frac{z - az}{\pi L - az} c\,(\pi L)^2 \frac{L^2}{2} \\
&\geq C_a z
\end{aligned}$$

for a constant C_a only depending on a and L, where the strong Markov property and then the lemma are applied on the second line.

On the other hand, integrating the assumed bound gives

$$\mathbb{E}\int_0^\infty |W_t^1 - W_t^2|_{\mathbb{T}}^2 \mathrm{d}t \leq (\alpha(z))^2 \int_0^\infty \zeta(t)\mathrm{d}t \leq (\alpha(z))^2 \|\zeta\|_{L^1([0,\infty))}.$$

Hence

$$C_a z \leq (\alpha(z))^2 \|\zeta\|_{L^1([0,\infty))}$$

which is the claimed result. ∎

Chapter 3

Hypocoercivity via Harris's theorem for Kinetic Equations with jumps

3.1 Introduction

The goal of this chapter is to give quantitative rates of convergence to equilibrium for some kinetic equations with jumps via Harris's theorem. Harris's theorem [74, 96, 73] is a classical theorem in Markov processes. It is originates in the paper [74] where Harris gave conditions for existence and uniqueness of a steady state for Markov processes. It was then pushed forward by Meyn and Tweedie [96] to show exponential convergence to equilibrium. The paper, [73], gives an efficient way of getting quantitative rates for convergence to equilibrium once you have quantitatively verified the assumptions. Harris's theorem says, broadly speaking, that if you have a good confining property and some uniform mixing property in the centre of the state space then you have exponentially fast convergence to equilibrium in a weighted total variation norm. We give the precise statement in the next section. Harris's theorem has already been used to show convergence to equilibrium for a kinetic equation in [94]. Here the authors show convergence to equilibrium for the kinetic Fokker-Planck equation with non-quantitative rates. In [14] the authors use a strategy based on Harris's theorem to show non-quantitative rates for convergence to equilibrium for scattering equations similar to those we study. In [38] the authors show convergence to a non-equilibrium steady state for some non-linear kinetic equations on the torus using Doeblin's theorem.

We will show quantitative rates of convergence for the following equations.

- The linear relaxation Boltzmann equation

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U \cdot \nabla_v f) = \Pi f - f, \tag{3.1}$$

  where

$$\Pi f = \left(\int f(t,x,u)\mathrm{d}u \right) \mathcal{M}(v).$$

  This simple equation is well studied in kinetic theory and can be thought of as a toy model with similar properties to either the non-linear BGK equation or linear Boltzmann equation. Its also one of the simplest examples of a hypocoercive equation.

- The linear Boltzmann equation

$$\partial_t f + v \cdot \nabla_x f = Q(f, \mathcal{M}), \tag{3.2}$$

where $\mathcal{M}$ is the Maxwellian with temperature 1 and mean 0, and

$$Q(f,g) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(v, v_*, \sigma) \left(f(v')g(v'_*) - f(v)g(v_*) \right) \mathrm{d}\sigma \mathrm{d}v_*,$$

$$v' = \frac{v+v_*}{2} + \frac{|v-v_*|}{2}\sigma, \quad v'_* = \frac{v+v_*}{2} - \frac{|v-v_*|}{2}\sigma.$$

This equation models gas particles interacting with a background medium which is already in equilibrium.

- The kinetic non-local diffusion equation

$$\partial_t f + v \cdot \nabla_x f - (\nabla_x U \cdot \nabla_v f) = K * f - f + \nabla_v(vf), \tag{3.3}$$

where K is a smooth, radial, compactly supported function. This is a non-local equation so a jump process which behaves in a similar way to a kinetic Fokker-Planck or fractional kinetic Fokker-Planck equation.

These are all spatially inhomogeneous equations where we would expect to see hypocoercivity [102, 50]. Hypocoercivity is proved using well chosen changes of norms and a combination of functional inequalities. Here we present a different strategy of proof. As an example let us look at the linear relaxation Boltzmann equation on the torus as a stochastic process. We can write an SDE in an integrated form

$$\begin{aligned} X_t =& X_0 + \int_0^t V_s \mathrm{d}s, \\ V_t =& V_0 + \int_0^t \int_{\mathbb{R}^d} (w - V_{s^-}) P(\mathrm{d}s, \mathrm{d}w). \end{aligned}$$

Here P is a Poisson point process on $[0, \infty) \times \mathbb{R}^d$ with intensity measure the tensor product of Lebesgue in time and a Gaussian in velocity space. This is a piecewise deterministic Markov process, at each jump time noise enters the system at the level of velocity. This noise is then transferred into the spatial variable by the transport operator. We will write this in more detail in the later sections but here we give a flavour of the strategy. Consider initial data which has the law $\delta_{(x_1,v_1)}$. After one jump the noise is the velocity variable but not the spatial variable, the spatial variable is deterministic but the velocity is now random. The transport operator mixes between the velocity and the spatial variable so after one jump and some transport we are still supported on a d-dimensional subset of the $2d$ dimensional space but we can view this as having noise in the spatial variable and the velocity variable taking one of a discrete set of values conditional on the spatial variable. If we jump again after this we inject noise in again at the level of velocity so now the support of the solution is $2d$ dimensional. Following this carefully will allow us to bound the law at time t, $f(t)$, below by

$$\alpha 1_{B(0,R_1)}(x) 1_{B(0,R_2)}(v),$$

for some strictly positive constants α, R_1, R_2. This effect is in some ways reminiscent of transferring the coercivity to the other directions given by higher order Hörmander brackets as in part one of [116].

Using Harris's theorem gives an alternative and very different strategy to the methods following on from Villani in [116] for proving quantitative hypocoercivity theorems. It allows us to look at hypocoercive effects on the level of SDEs and to look at specific trajectories which might allow one to produce quantitative theorems based on more trajectorial intuition. Another difference is that the confining behaviour is shown here by exploiting good behaviour of moments rather than a Poincaré inequality. These are often equivalent [9, 40] and have advantages and disadvantages. However, the condition on the moments used here might be much easier to verify in the case where the equilibrium state cannot be made explicit. Harris's theorem has a restriction which is that we can only consider linear Markov processes which does not include some important spatially inhomogeneous kinetic equations.

We also look at Harris type theorems with weaker controls on moments to give analogues of all our theorems when the confining potential is weaker and give algebraic rates of convergence with rates depending on the assumption we make of the confining potential. Subgeometric convergence for kinetic Fokker-Planck equations with weak confinement has been shown in [51, 9, 34]. To our knowledge this is the only work showing this type of convergence in a quantitative way for the equations we present.

The plan of the chapter is as follows. We introduce Harris's theorem. Then we have a section on each equation describing how to verify the assumptions of Harris's theorem.

In all our theorems it is possible to compute the constants in the final exponential convergence to equilibrium result. The proofs are constructive and do not rely on contradiction or compactness techniques. When the equations are set on the flat torus these constants depend on the collision operator but not on any a priori information about the solution. When we look at equations with confining potentials the constants depend both on the collision operator and on the confining potential. The dependence on the confining potential can be made explicit in terms of the various bounds we will assume on this potential. There is no mathematical difficulty in actually calculating the constants but they will be extremely cumbersome and complicated. They are also extremely unlikely to be sharp.

3.2 Harris's Theorem

Now let us be more specific about Harris's theorem. We give the theorems and assumption as in the setting of [73] where they make it clear how the rates depend on those in the assumptions. Let us define our setting. Rather than looking at these equations as PDEs we can consider the Markov semigroup $\mathcal{P}_t$. This is the continuous space version of a Markov transition matrix. $\mathcal{P}_t\mu$ is the weak solution to the PDE with initial data μ. Therefore $\mathcal{P}_t = e^{tG}$ where G is the generator. Saying that $\mathcal{P}_t$ is a Markov semigroup means that if $M(S)$ is the space of measure on phase space S then we have that $\mathcal{P}_t$ is a *linear* map given by a measurable kernel. This means that, if we consider $\mathcal{P}_t$ acting on density functions, then $\mathcal{P}_t$ will be *linear, mass preserving* and *positivity preserving*. We can also look at the action of $\mathcal{P}_t$ on observables.

$$\mathcal{P}_t^*\phi(z) = \int \phi(z')\mathcal{P}_t\delta_z(\mathrm{d}z').$$

Then we can define the *forwards operator*, L, associated to $\mathcal{P}_t$ by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_t^*\phi\Big|_{t=0} = L\phi.$$

We begin by looking a Doeblin's theorem. Harris's theorem is a natural successor to Doeblin's theorem. Harris's and Doeblin's theorems are normally stated for a fixed time t_*. In our theorems we work to choose an appropriate t_*.

Hypothesis 3.1 (Doeblin's Condition)**.** *We have a Markov semigroup $\mathcal{P}_{t_*}$ where there exists ν a probability distribution and $\alpha \in (0,1)$ such that for any z in the state space we have*

$$\mathcal{P}_{t_*}\delta_z \geq \alpha\nu$$

Using this we prove

Theorem 3.1 (Doeblin's Theorem)**.** *If we have a semigroup $\mathcal{P}_{t_*}$ satisfying Doeblin's condition then for any two measures μ_1 and μ_2 we have that*

$$\|\mathcal{P}_{t_*}^n\mu_1 - \mathcal{P}_{t_*}^n\mu_2\|_{TV} \leq (1-\alpha)^n\|\mu_1 - \mu_2\|.$$

Proof. This proof is classical and can be found in various versions in [73] and many other places. The key idea is that the minorisation condition tells us that the two measures share a proportion or their distribution after a certain time. First suppose that $\|\mu_1 - \mu_2\|_{TV} = 2\beta$ then there exists $\mu_0, \tilde{\mu}_1, \tilde{\mu}_2$ probability measures such that we can write

$$\mu_i = (1-\beta)\mu_0 + \beta\tilde{\mu}_i$$

and $\|\tilde{\mu}_1 - \tilde{\mu}_2\| = 2$. We find this by setting

$$\mu_0(A) = \frac{1}{1-\beta}\min\{\mu_1(A), \mu_2(A)\}, \quad \tilde{\mu}_i = \frac{1}{\beta}\left(\mu_i - (1-\beta)\mu_0\right).$$

So then since $\mathcal{P}_{t_*}$ is linear we have that

$$\|\mathcal{P}_{t_*}\mu_1 - \mathcal{P}_{t_*}\mu_2\|_{TV} = \beta\|\mathcal{P}_{t_*}\tilde{\mu}_1 - \mathcal{P}_{t_*}\tilde{\mu}_2\|_{TV}.$$

Therefore we may as well assume that $\beta = 1$ initially. Once we have done this, since $\mathcal{P}_{t_*}$ is Markov we can disintegrate over the possible starting conditions to get

$$\|\mathcal{P}_{t_*}\tilde{\mu}_1 - \mathcal{P}_{t_*}\tilde{\mu}_2\|_{TV} = \int\int \|\mathcal{P}_{t_*}\delta_{z_1} - \mathcal{P}_{t_*}\delta_{z_2}\|_{TV}\tilde{\mu}_1(\mathrm{d}z_1)\tilde{\mu}_2(\mathrm{d}z_2).$$

We now use our assumption to rewrite

$$\mathcal{P}_{t_*}\delta_{z_i} = \alpha\nu + (1-\alpha)\gamma_i,$$

where γ_i is a probability measure. This means that,

$$\|\mathcal{P}_{t_*}\delta_{z_1} - \mathcal{P}_{t_*}\delta_{z_1}\|_{TV} = (1-\alpha)\|\gamma_1 - \gamma_2\|_{TV} \leq 2(1-\alpha).$$

Substituting this all back in gives that

$$\|\mathcal{P}_{t_*}\mu_1 - \mathcal{P}_{t_*}\mu_2\|_{TV} \leq (1-\alpha)\|\mu_1 - \mu_2\|_{TV}.$$

We then iterate this to get the final result. □

Harris's theorem extends this to the setting where we cannot prove minorisation uniformly on the whole of the state space. The idea is to use the argument given above on the centre of the state space then exploit the Lyapunov structure to show that any stochastic process will return to the centre infinitely often.

We make two assumptions on the behaviour of $\mathcal{P}_{t_*}$, for some fixed t_*

Hypothesis 3.2. *There exists some function $V : \mathbb{T}^d \times \mathbb{R}^d \to [0,\infty)$ and constants $D \geq 0, \alpha \in (0,1)$ such that*

$$\mathcal{P}^*_{t_*}(V)(x,v) \leq \alpha V(x,v) + D.$$

Remark. *We use the name Lyapunov condition as it is the standard name used for this condition in probability literature. However, we should stress this condition is not closely related the Lyapunov method for proving convergence to equilibrium. We do prove monotonicity of a functional.*

Remark. *In our situation where we have an equation on the law $f(t)$ this is equivalent to the statement*

$$\int_S f(t,z)V(z)\mathrm{d}z \leq \alpha \int_S f(0,z)V(z)\mathrm{d}z + D.$$

We normally verify this by showing that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_S f(t,z)V(z)\mathrm{d}z \leq -\lambda \int_S f(t,z)V(z)\mathrm{d}z + K,$$

for some positive constants K and λ.

The idea behind verifying the Lyapunov structure in our case comes from [94] where they use similar Lyapunov structures for the kinetic Fokker-Planck equation. When we work on the torus the Lyapunov structure is only needed in the v variable and the result is purely about how moments in v are affected by the collision operator.

The next assumption is a minorisation condition as in Doeblin's theorem

Hypothesis 3.3. *There exists a probability measure ν and a constant $\beta \in (0,1)$ such that*

$$\inf_{(x,v)\in\mathcal{C}} \mathcal{P}_{t_*}\delta_x \geq \beta\nu,$$

where

$$\mathcal{C} = \{(x,v) \ : \ V(x,v) \leq R\}$$

for some $R > 2D/(1-\alpha)$.

Remark. *Production of quantitative lower bounds as a way to quantify the* positivity *of a solution has been proved and used in kinetic theory before. For example it is an assumption required for the works of Desvillettes and Villani [46, 47]. Such lower bounds have been proved for the non-linear Boltzmann equation in [101, 28, 29].*

This second assumption is more challenging to verify in our situations. Here we use a strategy based on our observation about how noise is transferred from the v to the x variable as described earlier. The actual calculations are based on the PDE governing the evolution and iteratively using Duhamel's formula.

We define a distance on probability measures for every $a > 0$

$$\rho_a(\mu_1, \mu_2) = \int (1 + aV(x, v))|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Theorem 3.2 (Harris's Theorem)**.** *If hypotheses 3.2 and 3.3 hold then there exists $\bar{\alpha} \in (0,1)$ and $a > 0$ such that*

$$\rho_a(\mathcal{P}_{t_*}\mu_1, \mathcal{P}_{t_*}\mu_2) \leq \bar{\alpha}\rho_a(\mu_1, \mu_2).$$

Explicitly if we choose $\beta_0 \in (0, \beta)$ and $\alpha_0 \in (\alpha + 2D/R, 1)$ then we can set $\gamma = \beta_0/K$ and $\bar{\alpha} = (1 - (\beta - \beta_0)) \wedge (2 + R\gamma\alpha_0)/(2 + R\gamma)$.

Remark. *We have that*

$$\min\{1, a\}\rho_1(\mu_1, \mu_2) \leq \rho_a(\mu_1, \mu_2) \leq \max\{1, a\}\rho_1(\mu_1, \mu_2).$$

We can also iterate Theorem 3.2 to get

$$\rho_a(\mathcal{P}_{nt_*}\mu_1, \mathcal{P}_{nt_*}\mu_2) \leq \bar{\alpha}^n\rho_a(\mu_1, \mu_2).$$

Therefore we have that

$$\rho_1(\mathcal{P}_{nt_*}\mu_1, \mathcal{P}_{nt_*}\mu_2) \leq \bar{\alpha}^n \frac{\max\{1, a\}}{\min\{1, a\}}\rho_1(\mu_1, \mu_2).$$

Remark. *In this paper we always consider functions V where $V(z) \to \infty$ as $|z| \to \infty$. In this case we can replace $\mathcal{C}$ in hypothesis 3.3 with some ball of radius R' which will contain $\mathcal{C}$.*

In order to prove Harris's theorem we follow [73] and formulate the weighted total variation norm as dual Lipschitz.

Sketch proof of Harris's Theorem. We are looking at convergence in a weighted total variation which is the dual of

$$\|\phi\| = \sup_x \frac{|\phi(x)|}{1 + U(x)}.$$

We wish to write this as a Lipschitz norm. We also introduce a parameter which we can tune to help adapt the proof to give quantitative rates in a simple way. If we write

$$d_a(z_1, z_2) = \begin{cases} 0 & z_1 = z_2 \\ 2 + aV(z_1) + aV(z_2) & z_1 \neq z_2 \end{cases},$$

and define

$$\|\phi\|_a = \sup_z \frac{|\phi(z)|}{1 + aV(z)},$$

Then

$$\|\phi\|_{Lip_{d_a}} = \inf_{c \in \mathbb{R}} \|\phi + c\|_a.$$

For details of this see [73].

Lets look when $V(z_1) + V(z_2) \geq R$. We have that if $\|\phi\|_{Lip_{d_a}} \leq 1$ then

$$\begin{aligned}|\mathcal{P}^*_{t_*}\phi(z_1) - \mathcal{P}^*_{t_*}\phi(z_2)| \leq& 2 + a\mathcal{P}^*_{t_*}V(z_1) + a\mathcal{P}^*_{t_*}V(z_2)\\ \leq& 2 + a\alpha\,(V(z_1) + V(z_2)) + aD\\ \leq& \tilde{\alpha}\,(2 + a(V(z_1) + V(z_2))) = \tilde{\alpha}d_a(z_1, z_2).\end{aligned}$$

Where

$$\tilde{\alpha} = \frac{2 + aR\alpha}{2 + aR}.$$

Now we look at the case $V(z_1)+V(z_2) \leq R$ and hence z_1 and z_2 are in $\mathcal{C}$. By similar arguments to Doeblin's theorem we get that

$$|\mathcal{P}^*_{t_*}\phi(z_1) - \mathcal{P}^*_{t_*}\phi(z_2)| \leq 2(1-\beta) + a\alpha(V(z_1) + V(z_2)) + aD.$$

So now we choose a sufficiently small to mean that this is a contraction. □

There are versions of Harris's theorem adapted to weaker Lyapunov conditions which give subgeometric convergence [9]. We use the following theorem which can be found in section 4 of [70].

Theorem 3.3 (Subgeometric Harris's Theorem)**.** *Given the forwards operator, L, of our Markov semigroup $\mathcal{P}$, suppose there exists a continuous function M valued in $[1,\infty)$ with pre compact level sets such that*

$$LM \leq K - \phi(M),$$

for some constant K and some strictly concave function $\phi : \mathbb{R}_+ \to \mathbb{R}$. Assume that for every $C > 0$ we have the minorisation condition like 3.3. i.e. for some T a time and ν a probability distribution and $\alpha \in (0,1)$, then for all z with $M(z) \leq C$

$$\mathcal{P}_t\delta_z \geq \alpha\nu.$$

With these conditions we have that

- *There exists a unique invariant measure μ for the Markov process and it satisfies*

$$\int \phi(M(z))\mathrm{d}\mu \leq K.$$

- *Let H_ϕ be the function defined by*

$$H_\phi = \int_1^u \frac{\mathrm{d}s}{\Phi(s)}$$

then there exists an constant C such that for every z_1, z_2 we have

$$\|\mathcal{P}_t\delta_{z_1} - \mathcal{P}_t\delta_{z_2}\| \leq C\frac{M(z_1) + M(z_2)}{H_\phi^{-1}(t)}.$$

- *There exists another constant C such that*

$$\|\mathcal{P}_t\delta_z - \mu\| \leq \frac{CM(z)}{H_\phi^{-1}(t)} + \frac{C}{(\phi \circ H_\phi^{-1})(t)}$$

We will apply this abstract theorem as well as Harris's theorem to the PDEs we study to show convergence when they only satisfy a weaker confinement condition.

3.3 Linear relaxation equation

We begin with the linear relaxation equation. This is the simplest of our equations. This allows us to present the key ideas of the strategy of proof which can then be built on to show similar results for the more complicated cases. On the torus without confining potential we do not even need to use a Lyapunov structure so we use Doeblin's theorem. Convergence to equilibrium for this equation has been shown in [33] in H^1 to converge faster than any polynomial function of t. It was then shown to converge exponentially fast in both H^1 and L^2 using hypocoercivity techniques [77, 102, 50].

3.3.1 On the flat torus

We consider

$$\partial_t f + v \cdot \nabla_x f = Lf, \tag{3.4}$$

posed for $(x, v) \in \mathbb{T}^d \times \mathbb{R}^d$, where

$$Lf(x,v) := L^+f(x,v) - f(x,v) := \left(\int_{\mathbb{R}^d} f(x,u)\,\mathrm{d}u\right) M(v) - f(x,v). \tag{3.5}$$

As in the introduction either of these equations is a PDE on the law of a Markov jump process which can be written

$$X_t = X_0 + \int_0^t V_s \mathrm{d}s, \tag{3.6}$$

$$V_t = V_0 + \int_0^t \int_{\mathbb{R}^d} (w - V_{s^-})\, P(\mathrm{d}s, \mathrm{d}w). \tag{3.7}$$

P is a Poisson random measure with intensity measure $\lambda_{\mathbb{R}_+} \otimes \gamma_d$. Here $\lambda_{\mathbb{R}_+}$ is Lebesgue measure on $\mathbb{R}_+$ and γ_d is the d dimensional standard Gaussian measure.

Theorem 3.4. *The solutions to equation (3.4) converge exponentially fast to equilibrium in total variation distance. This rate is explicitly calculable. i.e. There exists some $\lambda > 0$ and $C > 0$, which we can compute, such that*

$$\|f(t) - \mu\|_{TV} \leq Ce^{-\lambda t}\|f(0) - \mu\|_{TV},$$

where $f(t)$ is a solution to (3.4) at time t.

As already mentioned here we can show a global Doeblin condition so we do not need a Lyapunov condition. We focus on showing uniform minorisation. We define $(T_t)_{t\geq 0}$ as the transport semigroup associated to the operator $-v\cdot\nabla_x f$; that is, $t \mapsto T_t f_0$ solves the equation $\partial_t f + v\nabla_x f = 0$ with initial condition f_0. When f_0 is a function one can write T_t explicitly as

$$T_t f_0(x,v) = f_0(x - tv, v). \tag{3.8}$$

Lemma 3.1. *Let $f(t)$ be a solution of (3.4) with initial data f_0, ($f(t) = \mathcal{P}_t f_0$) then we have that*

$$e^t f(t) \geq \int_0^t \int_0^s T_{t-s} L^+ T_{s-r} L^+ T_r f_0 \, \mathrm{d}r \, \mathrm{d}s. \tag{3.9}$$

Remark. *This is to say that f_t will be bounded by the density restricted to the set where there are only two jumps in time t. i.e.*

$$\mathbb{P}(Z_t \in A) \geq \mathbb{P}(Z_t \in A \textit{ and there are exactly two jumps before time } t).$$

Proof. Here we look at measures as well as densities in which case we take $\geq$ to mean when integrated against a positive test function. We prove this only in the case of functions but the same proof works with small adaptations when understanding everything by duality. First we note that

$$\partial_t \left(T_{-t} f(t) \right) = T_{-t} \left(L^+ f - f \right).$$

Therefore we have

$$\partial_t \left(e^t T_{-t} f(t) \right) = e^t T_{-t} L^+ f(t).$$

So Duhamel's formula gives

$$e^t T_{-t} f(t) = f(0) + \int_0^t e^s T_{-s} L^+ f(s) \mathrm{d}s,$$

changing variables we have

$$e^t f(t) = T_t f(0) + \int_0^t e^s T_{t-s} L^+ f(s) \mathrm{d}s.$$

Looking at just the first term, and noting that L^+ is positive, we have that

$$e^t f(t) \geq T_t f(0),$$

this also clearly holds replacing t by s. Again as L^+ is a positive operator so we can substitute this into the second term to get

$$e^t f(t) \geq \int_0^t T_{t-s} L^+ T_s f(0) \mathrm{d}s.$$

We can then again substitute this into Duhamel's formula to get

$$e^t f(t) \geq \int_0^t \int_0^s T_{t-s} L^+ T_{s-r} L^+ T_r f(0) \mathrm{d}r \mathrm{d}s.$$

□

We will now check two properties. The first one says that the operator L always allows jumps to any velocity:

Lemma 3.2. *There exist $\alpha_L, r_L > 0$ such that for all non-negative measures ν on the velocity space, $\mathbb{R}^d$, we have*

$$L^+\nu \geq \alpha_L \nu(\mathbb{R}^d) 1_{|v| \leq r_L}. \tag{3.10}$$

Here $\geq$ is understood by duality.

Proof. This is true for any r_L and α_L will depend on r_L just by setting

$$\mathcal{M}(v) \geq \mathcal{M}(r_L) 1_{|x| \leq r_L}.$$

□

The second one says that the transport part can move a particle to a neighbourhood of 0, given that one starts out with the correct velocity:

Lemma 3.3. *For all $R > 0$ there exist $\alpha_T, r_T, t_0 > 0, t_0 > \epsilon > 0$ (possibly depending on R) such that for all non-negative measures μ on $\mathbb{T}^d$ we have*

$$T_t(\mu \otimes 1_{B(0,r_L)})(A \times \mathbb{R}^d) \geq \alpha_T \mu(B(0,R)) |B(0,r_T) \cap A| \ \ \forall t \in (t_0 - \epsilon, t_0 + \epsilon), \quad A \in \mathcal{B}(\mathbb{T}^d). \tag{3.11}$$

Proof. We show this assuming that μ has density h. This is just in order to write the transport semi group more explicitly. Exactly the same proof works in the measure case. We have that

$$T_t \left(h(x) 1_{B(0;r_L)}(v) \right) = h(x - vt) 1_{B(0;r_L)}(v).$$

Integrating this and changing variables gives that

$$\int T_t \left(h(x) 1_{B(0;r_L)}(v) \right) \mathrm{d}v = \frac{1}{t^d} \int h(y) 1_{B(0;r_L)} \left(\frac{x-y}{t} \right) \mathrm{d}y.$$

We have that

$$1_{B(0;r_L)} \left(\frac{x-y}{t} \right) \geq 1_{B(0;r_L/2)}(x/t) 1_{B(0,r_L/2)}(y/t).$$

Therefore for all $t > 2R/r_L$ we have

$$1_{B(0;r_L)} \left(\frac{x-y}{t} \right) \geq 1_{B(0;R)}(x) 1_{B(0;R)}(y).$$

Hence

$$T_t \left(h(x) 1_{B(0;r_L)}(v) \right) \geq \frac{1}{t^d} \int_{|y| \leq R} h(y) \mathrm{d}y 1_{B(0,R)}(x).$$

So if we take t to be in $(2R/r_L, 2R/r_L + \epsilon)$ we have our assumption with $\alpha_T = (2R/r_L + \epsilon)^d$ and $r_T = R$. □

Proof of Theorem 3.4. Take now $f_0 = \delta_{(x_0,v_0)}$. Let us write $Z_r(x,v) = (X_r, V_r)(x,v)$ for the solution to the equation $\dot{X}_r = V_r, \dot{V}_r = 0$ with initial data (x,v). With these two properties we have the following lower bounds, which we will use to obtain a lower bound in (3.9):

$$T_r f_0 = \delta_{Z_r(x_0,v_0)}.$$

Using Lemma 3.2,

$$L^+ T_r f_0 \geq \alpha_L \, \delta_{X_r(x_0,v_0)}(x) \, 1_{|v|\leq r_L} =: \alpha_L h(x) \, 1_{|v|\leq r_L}.$$

Using Lemmas 3.2 and 3.3, whenever $s-r \in (t_0-\epsilon, t_0+\epsilon)$ we have

$$\begin{aligned} L^+ T_{s-r} L^+ T_r f_0 &\geq \alpha_L \left(\int_{\mathbb{R}^d} T_{s-r} L^+ T_r f_0 \, \mathrm{d}u \right) 1_{|v|\leq r_L} \\ &\geq \alpha_L^2 \left(\int_{\mathbb{R}^d} T_{s-r}\big(h(x) \, 1_{|u|\leq r_L}\big) \, \mathrm{d}u \right) 1_{|v|\leq r_L} \\ &\geq \alpha_L^2 \alpha_T \, 1_{|x|\leq r_T} \, 1_{|v|\leq r_L}. \end{aligned}$$

We now need to allow for a final bit of movement along the flow T_{t-s}. Let us assume that ϵ is sufficiently small that $r_L \epsilon \leq r_T/2$ then

$$T_{t-s} 1_{B(0;r_T)}(x) 1_{B(0;r_L)}(v) = 1_{B(0;r_T)}(x-vt) 1_{B(0;r_L)}(v) \geq 1_{B(0;r_T/2)}(x) 1_{B(0;r_L)}(v).$$

This means that for all t,s,r such that $r \leq \epsilon, t-s \leq \epsilon$ and $s-r \in (2R/r_L, 2R/r_L + 2\epsilon)$ we have

$$T_{t-s} L^+ T_{s-r} L^+ T_r f_0 \geq \alpha_L^2 \alpha_T 1_{|x|\leq r_T/2} 1_{|v|\leq r_L}.$$

We now integrate this setting $t = 2R/r_L + 2\epsilon$ then

$$\begin{aligned} \int_0^t \int_0^s T_{t-s} L^+ T_{s-r} L^+ T_r f_0 \mathrm{d}r \mathrm{d}s &\geq \alpha_L^2 \alpha_T \int_{t-\epsilon}^t \int_0^\epsilon 1_{|x|\leq r_T/2} 1_{|v|\leq r_L} \mathrm{d}r \mathrm{d}s \\ &\geq \alpha_L^2 \alpha_T \epsilon^2 1_{|x|\leq \delta_T/2} 1_{|v|\leq r_L}. \end{aligned}$$

Now since we are on the torus we can just make R large enough so the ball of size R covers the whole torus. Therefore this bound is uniform in starting positions. This means we can use Doeblin's theorem rather than Harris's theorem and get that

$$\|f_t - \mu\|_{TV} \leq e^{-\lambda t} \|f_0 - \mu\|_{TV}.$$

□

3.3.2 With a Confining Potential

Consider the equation

$$\partial_t f + v \cdot \nabla_x f - \nabla_x U(x) \cdot \nabla_v f = Lf, \tag{3.12}$$

here L is defined as before and $x, v \in \mathbb{R}^d$. Now we have to use the full power of Harris's theorem to show convergence to equilibrium. This is because the behaviour in x is necessarily local and we cannot expect the trajectories to reach the centre at the same time for any x. This means we also have to find a Lyapunov function for the flow and weight our norm by this function. This means weighting the TV norm with moments in v as well as x in order to see the confinement through the transport operator which mixes x and v we need weights in v as well.

Theorem 3.5. *Suppose that $U(x)$ is a function satisfying*

$$|\nabla_x U(x)| \leq C U(x)^\eta$$

for some $\eta \in (0,1)$ and

$$x \cdot \nabla_x U(x) \geq \gamma_1 |x|^2 + \gamma_2 U(x) - A$$

for positive constants and $\gamma_1 \leq 1$. Then the solution to (3.12) converges exponentially fast to equilibrium in a weighted total variation norm. More specifically there exists $C > 0$ and $\lambda > 0$ which we can calculate explicitly and do not depend on the initial data such that

$$\rho(\mathcal{P}_t \mu_1, \mathcal{P}_t \mu_2) \leq C e^{-\lambda t} \rho(\mu_1, \mu_2),$$

where

$$\rho(\mu_1, \mu_2) = \int \left(1 + U(x) + \frac{1}{2}|v|^2 + \frac{1}{4} x \cdot v + \frac{1}{8}|x|^2\right) |\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Furthermore as U is super quadratic at infinity ρ is equivalent to the distance weighted by the Hamiltonian

$$\tilde{\rho}(\mu_1, \mu_2) = \int (1 + H(x,v)) |\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Remark. *The conditions on U hold for any C^2 function which is super quadratic and grows polynomially at infinity. This includes multiple wells. The standard assumption is that U needs to satisfy a Poincaré inequality in [50] and others. In this paper they also need that*

$$\lim_{|x| \to \infty} \left(|\nabla_x V|^2 - \Delta_x V \right) > 0.$$

Therefore our condition seems to be stronger as it limits how fast the potential can grow at infinity.

To show the minorization condition we use that the jump operator instantaneously produces large velocities.

Lemma 3.4. *For any R_1 there exists δ_M and R_2 such that there will be a range (s_*, s^*) for which if $s \in (s_*, s^*)$ we have*

$$\int T_s(\delta_{x_0} 1_{|v| \leq R_2}) \mathrm{d}v \geq 1_{|x| \leq \delta_M}$$

for any $U(x_0) \leq R_1$.

Proof. Let X_t, V_t be the solutions to the flow T_t. Then

$$\dot{X}_t = V_t, \quad \dot{V}_t = -\nabla_x U(X_t).$$

So by Taylor expanding we have

$$X_t = \tilde{X}_t + \mathrm{error} = x + vt + \mathrm{error}.$$

Furthermore

$$|X_t - \tilde{X}_t| \leq \max_{\tau \leq t} \frac{1}{2} t^2 |\nabla_x U(X_\tau)|.$$

The energy $U(x) + |v|^2/2$ is fixed by the flow of the equation and the energy of an initial point in the set we wish to evolve is bounded by $R_1 + R_2^2/2$. Using our assumption this means that

$$|X_t - \tilde{X}_t| \leq Ct^2 \left(R_1 + \frac{R_2^2}{2}\right)^\eta.$$

If we evolve the set under $\tilde{X}_t$ we get the ball of radius $R_2 t$ around x_0. So if we want to hit the target set $B(0, \delta_M)$ then we need that

$$X_t\left(\delta_{x_0} \times B(0, R_2)\right) \supset B(0, \delta_M).$$

This is if and only if

$$\delta_{x_0} \times B(0, R_2) \supset X_t^{-1} B(0, \delta_M).$$

Furthermore, since $\tilde{X}_t$ is a bijection we have that this is if and only if

$$\tilde{X}_t\left(\delta_{x_0} \times B(0, R_2)\right) \supset \tilde{X}_t X_t^{-1} B(0, \delta_M).$$

In the same was as above

$$|\tilde{X}_t(X_t^{-1}(x,v)) - (x,v)| = |\tilde{X}_t(X_t^{-1}(x,v)) - X_t(X_t^{-1}(x,v))| \leq Ct^2\left(R_1 + \frac{R_2^2}{2}\right)^\eta.$$

Therefore $\tilde{X}_t X_t^{-1} B(0, \delta_M)$ is contained within $B(0, \delta_M + Ct^2(R_1 + R_2^2/2)^\eta)$ so we want to show that

$$\tilde{X}_t\left(\delta_{x_0} \times B(0, R_2)\right) \supset B\left(0, \delta_M + Ct^2\left(R_1 + \frac{R_2^2}{2}\right)^\eta\right).$$

That is all possible displacements of the target ball by amounts less that $Ct^2(R_1 + R_2^2/2)^\eta$ are contained in the ball of radius $R_2 t$ around x_0. This will happen if

$$R_2 t \geq 2R_1 + \delta_M, \;\; Ct^2\left(R_1 + \frac{R_2^2}{2}\right)^\eta \leq R_1,$$

$$\Leftrightarrow t^2 \geq \frac{4R_1^2}{R_2^2}, \;\; t^2 \leq \frac{R_1}{C(R_1 + R_2^2/2)^\eta}.$$

The denominator in the lower bound grows faster than the denominator in the upper bound so for R_2 sufficiently large we have our required bound. □

Proof of Theorem 3.5. Now the strategy is to verify both the hypotheses 3.2 and 3.3 of Harris's theorem then use the quantitative version of Harris's theorem to get convergence with explicit rates. We start by proving the minorisation condition. Proof of hypothesis 3.3. If we take a point (x_0, v_0) with

$$U(x_0) + \frac{1}{2}|v_0|^2 \leq R_1,$$

then we take $f_0 = \delta_{(x_0,v_0)}$ and evolve it we get

$$T_r f_0 = \delta_{(x_1,v_1)}$$

where the energy bound is still satisfied. Then for our required R_2 we have that

$$L^+ T_r f_0 \geq \alpha_{R_2} \delta_{x_1} 1_{B(0,R_2)}.$$

We now apply our lemma to get that if $(s-r)$ is in our given range then we have

$$T_{s-r}L^+T_rf_0 \geq \alpha_{R_2}\delta_{v=G(x)}1_{B(0,\delta_M)}(x).$$

Then applying hypothesis 3.2 again gives

$$L^+T_{s-r}L^+T_rf_0 \geq \alpha_{R_2}^2 1_{B(0,\delta_M)}(x)1_{B(0,R_2)}(v).$$

Now if we make $t-s$ small we can bound

$$\begin{aligned}|T_{t-s}(x,v)-(x,v)| \leq & \max_{\tau \leq t-s}\sqrt{|V_\tau|^2+|\nabla_x U(X_\tau)|^2}(t-s)\\ \leq & C\max_{\tau\leq t-s}\left(\frac{1}{2}|V_\tau|^2+U(X_\tau)\right)(t-s)\\ \leq & C\left(\frac{1}{2}|v|^2+U(x)\right)(t-s).\end{aligned}$$

We can find some K depending on δ_M, R_2 such that the energy is bounded by K on $1_{B(0,R_2)}(v)1_{B(0,\delta_M)}(x)$. Also $B(0,R_2)\times B(0,\delta_M)$ contains some ball $B(0,\delta_M')$ so we make $t-s\leq \delta_M'/2K$ which means that

$$T_{t-s}L^+T_{s-r}L^+T_rf_0 \geq \alpha_{R_2}^2 1_{B(0,\delta_M'/2)},$$

provided that $t-s$ is sufficiently small and $s-r$ lies in the correct range. We then integrate over our permissible range of times to get minorisation.

Now we look at the Lyapunov condition. Proof of hypothesis 4.2. We look at the forwards operator

$$Sf = v\nabla_x f - \nabla_x U(x)\cdot\nabla_v f + L^+f - f.$$

We want a function $M(x,v)$ s.t

$$SM \leq -\lambda M + C$$

for some constants $\lambda > 0, C\geq 0$. We need to make the assumption that

$$x\cdot\nabla U(x) \geq \gamma_1|x|^2+\gamma_2 U(x) - A.$$

with $\gamma_1 \leq 1$. We then try the function

$$M(x,v) = U(x)+\frac{1}{2}|v|^2+ax\cdot v+b|x|^2.$$

We want this to be positive so we impose $a^2 < 2b$. We calculate that

$$\begin{aligned} SM =& \frac{1}{2} - \frac{1}{2}|v|^2 - ax \cdot v + a|v|^2 - ax \cdot \nabla_x U(x) + 2bx \cdot v \\ \leq& C' - \left(\frac{1}{2} - a\right)|v|^2 + (2b-a)x \cdot v - a\gamma_1 |x|^2 - a\gamma_2 U(x) \\ (a = 1/4, b = 1/8) \qquad =& C' - \frac{1}{4}|v|^2 - \frac{\gamma_1}{4}|x|^2 - \frac{\gamma_2}{4}U(x) \\ \leq& C' - \frac{\gamma_1}{4}(|x|^2 + |v|^2) - \frac{\gamma_2}{4}U(x) \\ \leq& C' - \frac{\gamma_1}{4}\left(\frac{1}{2}|v|^2 + \frac{1}{4}x \cdot v + \frac{1}{8}|x|^2\right) - \frac{\gamma_2}{4}U(x) \end{aligned}$$

So $M(x,v)$ works with

$$\lambda = \frac{\min(\gamma_1, \gamma_2)}{4}.$$

□

3.3.3 Subgeometric convergence

When we have the sub quadratic behaviour of the confining potential at infinity we can still use a Harris type theorem to show convergence to equilibrium. This translates into having subgeometric rates of convergence. Now instead of our earlier assumption on the Φ we instead make a weaker assumption

Theorem 3.6. *Suppose that $U(x)$ is a function satisfying*

$$|\nabla_x U(x)| \leq CU(x)^\eta, \quad x \cdot \nabla_x U(x) \geq \gamma_1 \langle x \rangle^\beta + \gamma_2 U(x) - A.$$

Where

$$\langle x \rangle = 1 + |x|^2,$$

and $\beta \in (0,1)$. Then the solution to (3.12) converges to equilibrium in a weighted total variation norm in the following way. We define the function M by

$$M(x,y) = U(x) + \frac{1}{2}|v|^2 + \frac{1}{4}x \cdot v + \frac{1}{8}|x|^2.$$

Then there exists a constant $C > 0$ such that

$$\|\mathcal{P}_t \delta_{z_1} - \mathcal{P}_t \delta_{z_2}\|_{TV} \leq C(M(z_1) + M(z_2))(1+t)^{-1/(1-\beta)},$$

and

$$\|\mathcal{P}_t \delta_z - \mu\|_{TV} \leq CM(z)(1+t)^{-1/(1-\beta)} + C(1+t)^{-\beta/(1-\beta)}.$$

Proof. The proof of the minorisation condition is exactly the same. We can also replicate the calculations for the Lyapunov function as in the proof of Theorem 3.5 to get that in this new situation we have for $a = 1/4, b = 1/8$ that

$$SM \leq C' - \frac{1}{4}|v|^2 - \frac{\gamma_1}{4}\langle x \rangle^\beta - \frac{\gamma_2}{4}U(x).$$

For $x, y \geq 1$
$$(x+y)^\beta \leq x^\beta + y^\beta.$$

So we have

$$\begin{aligned}
SM \leq & C'' - \frac{\gamma_1}{4}(\langle v\rangle + \langle x\rangle^\beta) - \frac{\gamma_2}{4}U(x)\\
\leq & C'' - \frac{\gamma_1}{4}(1+|x|^2+|v|^2)^\beta - \frac{\gamma_2}{4}U(x)^\beta\\
\leq & C'' - \lambda\left(1 + \frac{1}{2}|v|^2 + \frac{1}{4}x\cdot v + \frac{1}{8}|x|^2\right)^\beta - \lambda U(x)^\beta\\
\leq & C'' - \lambda\left(U(x) + \frac{1}{2}|v|^2 + \frac{1}{4}x\cdot v + \frac{1}{8}|x|^2\right)^\beta.
\end{aligned}$$

So we have that
$$SM \leq -\lambda M^\beta + C''.$$

This means we can take $\phi(s) = 1 + s^\beta$ Therefore
$$H_\phi(u) = \int_1^u \frac{1}{1+t^\beta}\mathrm{d}t \sim 1 + u^{1-\beta}$$

for u large. Therefore
$$H_\phi^{-1}(t) \sim 1 + t^{1/(1-\beta)}$$

for t large and
$$\phi \circ H_\phi^{-1} \sim (1+t)^{\beta/(1-\beta)}.$$

□

3.4 The Linear Boltzmann Equation on the Torus

We now look at the Linear Boltzmann equation. This has been studied in the spatially homogeneous case in [17, 31]. Here the interest is partly that this is a more complex and physically relevant operator. Also, it presents less globally uniform behaviour in v which means that we have to use a Lyapunov function even on the torus. Apart from this the strategy is very similar to that from the linear relaxation Boltzmann equation. The Lyapunov condition on the torus and the bound below on the jump operator have to be verified in this situation.

We consider for $x \in \mathbb{T}^d$

$$\partial_t f + v\cdot\nabla_x f = \int_{\mathbb{R}^d}\int_{\mathbb{S}^{d-1}} B\left(\frac{v-v_*}{|v-v_*|}\cdot\sigma, |v-v_*|\right)(f(v')\mathcal{M}(v'_*) - f(v)\mathcal{M}(v_*))\,\mathrm{d}\sigma\mathrm{d}v_*. \tag{3.13}$$

We assume that B splits as

$$B\left(\frac{v-v_*}{|v-v_*|}\cdot\sigma, |v-v_*|\right) = b\left(\frac{v-v_*}{|v-v_*|}\cdot\sigma\right)|v-v_*|^\gamma.$$

We make the cut off assumption that b is integrable in σ. In fact we make the much stronger assumption that b is bounded below by a constant. We also work in the hard spheres/Maxwell molecules regime that is to suppose $\gamma \geq 0$. We have

$$\partial_t f + v \cdot \nabla_x f = L^+ f - \sigma(v) f.$$

We have that

$$\sigma(v) \geq 0,$$

and $\sigma(v)$ behaves like $|v|^\gamma$ for large v. See [31] Lemma 2.1 for example.

We can write this as the equation on the law of the following jump process

$$\begin{aligned} X_t =& X_0 + \int_0^t V_s \mathrm{d}s, \\ V_t =& V_0 + \int_0^t \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} \int_0^\infty \int_0^\infty \left(\left(\frac{w + V_{s^-}}{2} + \frac{|w - V_{s^-}|}{2} \sigma \right) - V_{s^-} \right) 1_{p \leq b\left(\frac{V_{s^-} - w}{|V_{s^-} - w|} \cdot \sigma \right)} \\ & 1_{q \leq |V_{s^-} - w|^\gamma} P(\mathrm{d}s, \mathrm{d}\sigma, \mathrm{d}w, \mathrm{d}q, \mathrm{d}p). \end{aligned}$$

Here P is a Poisson random measure with intensity measure given by $\lambda_{\mathbb{R}_+} \otimes \lambda_{\mathbb{S}^{d-1}} \otimes \gamma_d \otimes \lambda_{\mathbb{R}_+} \otimes \lambda_{\mathbb{R}_+}$. Here λ_S is Lebesgue measure on S and γ_d is d-dimensional standard Gaussian.

Theorem 3.7. *If $f(t)$ is the solution to the linear Boltzmann equation, (3.13), for Maxwell molecules with cut off and b bounded below then there exists $C > 0$ and $\lambda > 0$, which we can compute explicitly, such that*

$$\rho(\mathcal{P}_t \mu_1, \mathcal{P}_t \mu_2) \leq C e^{-\lambda t} \rho(\mu_1, \mu_2),$$

where

$$\rho(\mu_1, \mu_2) = \int (1 + |v|^2) |\mu_1 - \mu_2| (\mathrm{d}x \mathrm{d}v).$$

We want to reduce to a similar situation the linear relaxation equation.

Lemma 3.5. *For f a solution to (3.13) we have that*

$$f(t,x,v) \geq e^{-tC(1+(R+r_L^2/2)^{\gamma/2})} \int_0^t \int_0^s T_{t-s} L^+ T_{s-r} L^+ T_r f_0(x,v) 1_{|v|^2 \leq R + 2r_L^2} \mathrm{d}r \mathrm{d}s.$$

Proof. We look at Duhamel's formula again and let us call

$$\Sigma(s,t,x,v) = \int_s^t \sigma(V_r(x,v)) \mathrm{d}r.$$

We get

$$f(t,x,v) \geq \int_0^t \int_0^s T_{t-s} e^{-\Sigma(s,t,x,v)} L^+ T_{s-r} e^{-\Sigma(r,s,x,v)} L^+ T_r e^{-\Sigma(0,r,x,v)} f_0(x,v) \mathrm{d}r \mathrm{d}s.$$

We restrict to paths where $|v|^2 \leq R + 2r_L^2$ for these

$$\Sigma(s,t,x,v) \leq (t-s) C \left(1 + (R + 2r_L^2)^{\gamma/2} \right).$$

Therefore

$$f(t,x,v) \geq e^{-tC(1+(R+2r_L^2)^{\gamma/2})} \int_0^t \int_0^s T_{t-s} L^+ T_{s-r} L^+ T_r f_0(x,v) 1_{|v|^2 \leq R+2r_L^2} \mathrm{d}r \mathrm{d}s.$$

□

We have that

$$L^+ f = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} b\left(\frac{v-v_*}{|v-v_*|} \cdot \sigma\right) |v - v*|^\gamma f(v') \mathcal{M}(v'_*) \mathrm{d}\sigma \mathrm{d}v_*.$$

Using Carleman representation we rewrite this as

$$L^+ f = \int_{\mathbb{R}^d} \frac{f(v')}{|v-v'|^{d-1}} \int_{E_{v,v'}} B(|u|, \xi) \mathcal{M}(v'_*) \mathrm{d}v'_*.$$

We want to bound this in the manner of the lemma from the first part. We look at hard spheres and no angular dependence this means

$$B(|u|, \xi) = C|u|^\gamma \xi^{d-2}$$

with $\gamma \geq 0$. We also have that

$$\xi = \frac{|v-v'|}{|2v - v' - v'_*|}, \quad |u| = |2v - v' - v'_*|.$$

So we have that

$$L^+ f = \int_{\mathbb{R}^d} \frac{f(v')}{|v-v'|} \int_{E_{v,v'}} |2v - v' - v'_*|^{\gamma - d - 2} \mathcal{M}(v'_*) \mathrm{d}v'_*.$$

We want to prove something similar to Lemma 3.2 but we look at this localised so we want

Lemma 3.6. *L^+ from the linear Boltzmann equation for hard spheres and no angular dependence satisfies for all R_L, r_L, there exists α such that*

$$L^+ g \geq \alpha \int_{B(0;R_L)} g(u) \mathrm{d}u 1_{|v| \leq r_L}.$$

Therefore the semigroup generated by this equation satisfies a minorisation equation for all starting points with $|v| \leq r_L$.

Proof. First we note that on $E_{v,v'}$ we have

$$|2v - v' - v'_*|^{-d-2} \geq C_d \exp\left(-\frac{1}{2}|v - v'_*|^2 - \frac{1}{2}|v - v'|^2\right).$$

Then since $\gamma \geq 0$ we have

$$|2v - v' - v - v'_*|^\gamma = \left(|v - v'|^2 + |v - v'_*|^2\right)^{\gamma/2} \geq |v - v'_*|^\gamma.$$

So this means that

$$\begin{aligned}\int_{E_{v,v'}} |2v - v' - v'_*|^{\gamma-d-2}\mathcal{M}(v'_*)\mathrm{d}v'_* \geq & Ce^{-|v-v'|^2/2}\int_{E_{v,v'}} |v-v'_*|^\gamma \exp\left(-\frac{1}{2}|v-v'_*|^2 - \frac{1}{2}|v'_*|^2\right)\mathrm{d}v'_* \\ \geq & Ce^{-|v-v'|^2/2-|v|^2}\int_{E_{v,v'}} |v-v'_*|^\gamma e^{-|v-v'_*|^2/2}\mathrm{d}v'_* \\ = & C'e^{-|v-v'|^2/2-|v|^2}.\end{aligned}$$

So we have that

$$\begin{aligned}L^+ f \geq & C\int_{\mathbb{R}^d} f(v')|v-v'|^{-1}e^{-|v-v'|^2/2-|v|^2}\mathrm{d}v' \\ \geq & C\int_{\mathbb{R}^d} f(v')e^{-2|v'|^2-3|v|^2} \\ \geq & Ce^{-2R_L^2}\int_{B(0,R)} f(v')\mathrm{d}v' e^{-3|v|^2}.\end{aligned}$$

So we see we have bounded by a Maxwellian as before.

For the minorisation we can argue almost exactly as for the linear relaxation equation. Suppose that we satisfy the condition in Lemma 3.6. Then we choose a starting point $\delta_{(x_0,v_0)}$ where $|v_0|^2 \leq R$ after the first transport this energy is preserved. Then after at each jump we add at most $r_L^2/2$ to the total energy, as we only follow trajectories where the velocity jumps to something smaller than r_L. This means for a path with two jumps we always stay within the the set $|v|^2 \leq R + 2r_L^2$. Then the largest velocity we can reach is $\sqrt{2(R+2r_L^2)}$ so if we set R_L to be this in Lemma 3.6. Therefore as for the linear relaxation equation we have that

$$T_r\delta_{(x_0,v_0)} = \delta_{(x_1,v_0)}.$$

Then since $|v_0|^2 \leq R_L$ we have

$$L^+T_r\delta_{(x_0,v_0)} \geq \alpha 1_{|v|\leq r_L}.$$

Now we have as for the linear relaxation Boltzmann.

$$T_{s-r}L^+T_r\delta_{(x_0,v_0)} \geq \alpha 1_{x\in B(x_1,tr_L)}1_{v\in G(x)}.$$

Now we can see that $G(x) \subset B(0,r_L)$ therefore if $(s-r)$ is sufficiently large so that $B(x_1,tr_L)$ covers the whole torus we have

$$L^+T_{s-r}L^+T_r\delta_{(x_0,v_0)} \geq \alpha^2 1_{|v|\leq r_L}.$$

Then since T_{t-s} doesn't alter the v-variable we have

$$T_{t-s}L^+T_{s-r}L^+T_r\delta_{(x_0,v_0)} \geq \alpha^2 1_{|v|\leq r_L}.$$

Minorisation follows from integrating over the permitted values of s, r as for the linear relaxation equation. □

We now reprove the moment control result from [17] in more detail On the torus we do need a Lyapunov functional in this case. We want to test with $M = v^2$. Let us use the n-representation for the collisions:

$$v' = v - n(u \cdot n), \qquad v'_* = v_* + n(u \cdot n).$$

We assume the collision kernel can be written as

$$B(|v - v_*|, |\xi|) = |v - v_*|^\gamma b(|\xi|),$$

where

$$\xi := \frac{u \cdot n}{|u|}, \qquad u := v - v_*.$$

We also assume that b is normalised, that is,

$$\int_{\mathbb{S}^d} b(|w \cdot n|)\, \mathrm{d}n = 1$$

for all unit vectors $w \in \mathbb{S}^{d-1}$.

Lemma 3.7. *Let L be the linear Boltzmann operator. There are constants $C, K > 0$ such that*

$$\int_{\mathbb{R}^d} L(f)|v|^2 \,\mathrm{d}v \leq -C \int_{\mathbb{R}^d} |v|^2 f \,\mathrm{d}v + K \int_{\mathbb{R}^d} f$$

for all non-negative measures f.

Proof. Using the weak formulation of the operator,

$$\int_{\mathbb{R}^d} L(f)|v|^2 \,\mathrm{d}v = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} f(v)\mathcal{M}(v_*)|v - v_*|^\gamma b(|\xi|)(|v'|^2 - |v|^2)\,\mathrm{d}n\,\mathrm{d}v\,\mathrm{d}v_*.$$

Now we notice that

$$\begin{aligned} |v'|^2 - |v|^2 = |v_*|^2 - |v'_*|^2 &= -(u \cdot n)^2 - 2(v_* \cdot n)(u \cdot n) \\ &= -|u|^2\xi^2 - 2(v_* \cdot n)(v \cdot n) + 2(v_* \cdot n)^2 \\ &= -|v|^2\xi^2 - |v_*|^2\xi^2 + v \cdot v_*\xi^2 - 2(v_* \cdot n)(v \cdot n) + 2(v_* \cdot n)^2. \end{aligned}$$

Note that the first term is negative and quadratic in v, and the rest of the terms are of lower order in v. Hence, calling $\gamma_b := \int_{\mathbb{S}^{d-1}} \xi^2 b(|\xi|)\,\mathrm{d}\xi$ we have

$$\begin{aligned}
\int_{\mathbb{R}^d} L(f)|v|^2\,\mathrm{d}v = &- \gamma_b \int_{\mathbb{R}^d} |v|^2 f(v) \int_{\mathbb{R}^d} \mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v \\
&- \gamma_b \int_{\mathbb{R}^d} f(v) \int_{\mathbb{R}^d} |v_*|^2 \mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v \\
&+ \gamma_b \int_{\mathbb{R}^d} v f(v) \int_{\mathbb{R}^d} v_* \mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v \\
&- 2 \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} (v\cdot n) f(v) \int_{\mathbb{R}^d} (v_*\cdot n)\mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v\,\mathrm{d}n \\
&+ \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} f(v) \int_{\mathbb{R}^d} (v_*\cdot n)^2\mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v\,\mathrm{d}n \\
\leq &- \gamma_b \int_{\mathbb{R}^d} |v|^2 f(v) \int_{\mathbb{R}^d} \mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v \\
&+ (2+\gamma_b) \int_{\mathbb{R}^d} |v| f(v) \int_{\mathbb{R}^d} |v_*|\mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v \\
&+ \int_{\mathbb{R}^d} f(v) \int_{\mathbb{R}^d} |v_*|^2\mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_*\,\mathrm{d}v.
\end{aligned}$$

We can now use the following bound, which holds for all $k \geq 0$ and some constants $0 < A_k \leq C_k$ depending on k:

$$A_k(1+|v|^\gamma) \leq \int_{\mathbb{R}^d} |v_*|^k \mathcal{M}(v_*)|v-v_*|^\gamma\,\mathrm{d}v_* \leq C_k(1+|v|^\gamma), \qquad v \in \mathbb{R}^d.$$

We get

$$\begin{aligned}
\int_{\mathbb{R}^d} L(f)|v|^2\,\mathrm{d}v \leq &- A_0\gamma_b \int_{\mathbb{R}^d} |v|^2(1+|v|^\gamma) f(v)\,\mathrm{d}v + C_1(2+\gamma_b)\int_{\mathbb{R}^d} |v|(1+|v|^\gamma)f(v)\,\mathrm{d}v \\
&+ C_2 \int_{\mathbb{R}^d} f(v)(1+|v|^\gamma)\,\mathrm{d}v \\
\leq &\int_{\mathbb{R}^d} f(v)(C_2 + C_1(1+\gamma_b/2)/\epsilon)\,(1+|v|^\gamma)\,\mathrm{d}v \\
&- (A_0\gamma_b - \epsilon C_1(1+\gamma_b/2)) \int_{\mathbb{R}^d} |v|^2(1+|v|^\gamma) f(v)\mathrm{d}v \\
\leq &\int_{\mathbb{R}^d} f(v)\left(C_2 + C_1(1+\gamma_b/2)/\epsilon + (\epsilon C_1(1+\gamma_b/2) - A_0\gamma_b)|v|^2\right)(1+|v|^\gamma) f(v)\mathrm{d}v \\
&- (A_0\gamma_b - \epsilon C_1(1+\gamma_b/2)) \int_{\mathbb{R}^d} |v|^2 f(v)\mathrm{d}v \\
\leq &\alpha_1 \int_{\mathbb{R}^d} f(v)\mathrm{d}v - \alpha_2 \int_{\mathbb{R}^d} |v|^2 f(v)\mathrm{d}v.
\end{aligned}$$

Here we choose ϵ sufficiently small to make the constant in front of the second moment negative. This also means that

$$(C_2 + C_1(1+\gamma_b/2)/\epsilon + (\epsilon C_1(1+\gamma_b/2) - A_0\gamma_b|v|^2)(1+|v|^\gamma)$$

is bounded above. These things together give the final line. □

Proof of Theorem 3.7. We have minorisation hypothesis 3.3 from lemma 3.2 and the Lyapunov structure 4.2 from lemma 3.7. □

3.5 Kinetic non-local diffusion equation

Here the equation is

$$\partial_t f + v \cdot \nabla_x f = K * f - f + \nabla_v \cdot (vf), \quad x \in \mathbb{T}^d, v \in \mathbb{R}^d. \tag{3.14}$$

or

$$\partial_t f + v \cdot \nabla_x f - \nabla_x U \cdot \nabla_v f = K * f - f + \nabla_v \cdot (vf), \quad x \in \mathbb{R}^d, v \in \mathbb{R}^d. \tag{3.15}$$

Where K is a smooth, radial, compactly supported function which integrates to one. The big difference with this equation is that we only spread the velocities out in a small ball around our current velocity. This means we cannot hope to prove something like Lemma 3.2. However we can instead make a large number of jumps where we only travel for small distances in between them. We make the assumption that

$$K(w) \geq \alpha 1_{B(0,\delta)}.$$

This equation gives the evolution of the law of the following Markov jump process.

$$\begin{aligned} X_t =& X_0 + \int_0^t V_s \mathrm{d}s, \\ V_t =& V_0 - \left(\int_0^t \nabla_x U(X_s) \mathrm{d}s \right) - \int_0^t V_s \mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} w P(\mathrm{d}s, \mathrm{d}w). \end{aligned}$$

Here P is a Poisson random measure with intensity measure $\lambda_{\mathbb{R}_+} \otimes K$.

We take a slightly different transport map to before which takes into account the confinement term in velocity as well. We can see that if

$$\dot{X}_t = V_t, \quad \dot{V}_t = -\nabla_x U(X_t) - V_t, X_0 = x, V_0 = v$$

then we have that

$$\partial_t \left(e^{-(d-1)t} f(t, X_t, V_t) \right) = e^{-(d-1)t} (K * f)(t, X_t, V_t).$$

We write

$$T_t f(x, v) = f(X_t, V_t)$$

and

$$L^+ f = K * f.$$

This means we get a similar result as before that

$$e^{-(d-1)t} f(t, x, v) \geq \int_{t_n}^t \int_{t_{n-1}}^t \cdots \int_0^t T_{t-t_n} L^+ T_{t_n - t_{n-1}} L^+ \ldots L^+ T_{t_1} f_0(x, v) \mathrm{d}t_1 \ldots \mathrm{d}t_n.$$

We also have that $H(x, v) = U(x) + |v|^2/2$ decreases under the flow T_t.

3.5.1 On the torus

Let us begin with looking on the torus so the transport part is simplified.

Theorem 3.8. *The solution to* (3.14) *converges exponentially fast in a weighted TV distance. Specifically there exists $C > 0$ and $\lambda > 0$, which we can compute explicitly, such that*

$$\rho(\mathcal{P}_t\mu_1, \mathcal{P}_t\mu_2) \le Ce^{-\lambda t}\rho(\mu_1, \mu_2),$$

where

$$\rho(\mu_1, \mu_2) = \int (1 + |v|^2)|\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Proof. Our strategy is to show that for any R, there exists n such that

$$\int T_{t-t_n}L^+T_{t_n-t_{n-1}}L^+ \dots T_{t_2-t_1}L^+T_{t_1}\delta_{x_0,v_0}\mathrm{d}x \ge \alpha_L 1_{B(0,\sqrt{d})} \quad \forall \quad v_0 \in B(0, R).$$

If we have this we can just let the transport part run so that we then cover the whole of the torus then we jump n times again to decouple the space and velocity.

Here the transport map reduces velocity

$$T_{t_1}\delta_{x_0,v_0} = \delta_{x_1,v_1}.$$

Then we have

$$L^+\delta_{x_1,v_0} \ge \alpha\delta_{x_1}(x)1_{B(v_1,\delta)}.$$

After transporting for a short time we have

$$T_{t_2-t_1}\alpha\delta_{x_1}(x)1_{B(v_0,\delta)} \ge 1_{x\in G(v)}1_{B(v_1,3\delta/4)}.$$

$$\begin{aligned} L^+1_{x\in G(v)}1_{B(v_0,\delta)} &\ge \alpha 1_{x\in G_2(v)}\left|B\left(w + \frac{v_1 - w}{|v_1 - w|}\frac{3\delta}{4}, \frac{\delta}{4}\right)\right|1_{B(v_1,5\delta/4)}(w) \\ &\ge \alpha C(\delta/4)^d 1_{x\in G_3(v)}1_{B(v_0,5\delta/4)}. \end{aligned}$$

We continue like this to get that

$$\begin{aligned} \int T_{t-t_n}L^+T_{t_n-t_{n-1}}L^+ \dots T_{t_2-t_1}L^+T_{t_1}\delta_{x_0,v_0}\mathrm{d}x &\ge \alpha^n(\delta/4)^{(n-1)d}1_{B(v_1,\delta(1+n/4))} \\ &\ge \alpha_L 1_{B(0,(1-e^{-1})^{-1}e\sqrt{d})}. \end{aligned}$$

Now we have that

$$\begin{aligned} T_1\left(\delta_{x\in G(v)}1_{B(0,(1-e^{-1})^{-1}e\sqrt{d})}\right) &= \int_{G(v)} T_1(\delta_{\tilde{x}}(x)1_{B(0,(1-e^{-1})^{-1}e\sqrt{d})}(v)\mathrm{d}\tilde{x} \\ &= \int_{G(v)} \delta_{\tilde{x}+(1-e^{-1})v}(x)1_{B(0,(1-e^{-1})^{-1}\sqrt{d})}(v)\mathrm{d}\tilde{x} \\ &= \int_{G(v)} 1_{B(\tilde{x},\sqrt{d})}\delta_{v\in H(x,\tilde{x})}\mathrm{d}\tilde{x} \\ &= \int_{G(v)} \delta_{v\in H(x,\tilde{x})}\mathrm{d}\tilde{x} = 1_{v\in K}. \end{aligned}$$

Here $H(x, \tilde{x})$ is the set of velocities which in $B(0, \sqrt{d})$ which will go from $\tilde{x}$ to x. So we end up v lying within some complicated set, K, inside $B(0, \delta)$. We now repeat our n steps and since we began with uniform distribution in x. We have that

$$T_{t_n - t_{n-1}} L^+ T_{t_{n-1} - t_{n-2}} L^+ \dots L^+ T_{t_1} (1_{v \in K}) \geq \alpha_L 1_{v \in B(0, (1-e^{-1})^{-1} e\sqrt{d})}.$$

This gives the minorisation condition.

Now we need a Lyapunov condition since our estimate cannot be made uniform in the starting velocity. We look at the kinetic energy and see that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \int f(t,x,v)|v|^2 \mathrm{d}x\mathrm{d}v = & \int \int K(v-u) f(t,x,u)|v|^2 \mathrm{d}v\mathrm{d}u\mathrm{d}x - \int f(t,x,v)|v|^2 \mathrm{d}x\mathrm{d}v \\
& + \int \nabla_v \cdot (v f(t,x,v))|v|^2 \mathrm{d}x\mathrm{d}v \\
\leq & 2 \int \int K(v-u)(|u|^2 + \delta^2) f(t,x,u) \mathrm{d}u\mathrm{d}v\mathrm{d}x - 3 \int f(t,x,v)|v|^2 \\
= & 2\delta^2 - \int f(t,x,v)|v|^2 \mathrm{d}v\mathrm{d}x.
\end{aligned}$$

Therefore we have both the minorisation 3.3 and Lyapunov 4.2 hypotheses satisfied. □

3.5.2 With a confining potential

Theorem 3.9. *Suppose that $U(x)$ is a function satisfying*

$$|\nabla_x U(x)| \leq C U(x)^\eta$$

for some $\eta \in (0,1)$ and

$$x \cdot U(x) \geq \gamma_1 |x|^2 + \gamma_2 \nabla_x U(x) - A$$

for positive constants and $\gamma_1 \leq 1$. Then the solution to (3.15) converges exponentially fast to equilibrium in a weighted total variation norm. More specifically there exists $C > 0$ and $\lambda > 0$ which we can calculate explicitly such that

$$\rho(\mathcal{P}_t \mu_1, \mathcal{P}_t \mu_2) \leq C e^{-\lambda t} \rho(\mu_1, \mu_2),$$

where

$$\rho(\mu_1, \mu_2) = \int \left(1 + U(x) + \frac{1}{2}|v|^2 + \frac{1}{2} x \cdot v + \frac{1}{4}|x|^2\right) |\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

Furthermore if U is super quadratic at infinity (which maybe implied by earlier assumptions) then ρ is equivalent to the distance weighted by the Hamiltonian

$$\tilde{\rho}(\mu_1, \mu_2) = \int (1 + H(x,v)) |\mu_1 - \mu_2|(\mathrm{d}x\mathrm{d}v).$$

We emulate the strategy from earlier. We want to restrict the amount we move when jumping to build up. Suppose we start with (x_0, v_0) such that $H(x_0, v_0) \leq R_1$ then we will add at most R_2 to this by increasing the velocity. So we stay in the set with $H(x,v) \leq R_1 + R_2$. We also have that same assumption as earlier that

$$|\nabla_x U| \le CU(x)^\eta.$$

First we don't need to use the result with η. Using these facts we have

$$|X_t - X_0| \le tC(R_1 + R_2), \quad |V_t - V_0| \le tC(R_1 + R_2).$$

So now we see that

$$T_{t_1}\delta_{(x_0,v_0)} = \delta_{(x_1,v_1)},$$

where (x_1, v_1) are also inside our good set. Then

$$L^+ T_{t_1}\delta_{(x_0,v_0)} \ge \alpha \delta_{x_1}(x) 1_{B(v_1,\delta)}.$$

Then we want to transport this set. If we make $t_2 - t_1$ suitable small then the velocity variable can have moved at most $\delta/2$ so we get to

$$T_{t_2-t_1} L^+ T_{t_1}\delta_{(x_0,v_0)} \ge \alpha \delta_{x \in G(v)} 1_{B(v_1,\delta/2)}.$$

We proceed as on the torus except this time we must transport for only small amounts of time and pay the price of shrinking the set. So we get

$$\begin{aligned}\int T_{t-t_n} L^+ T_{t_n - t_{n-1}} L^+ \dots T_{t_2-t_1} L^+ T_{t_1}\delta_{x_0,v_0} \mathrm{d}x &\ge \alpha^n (\delta/8)^{(n-1)d} 1_{B(v_1,\delta(1+n/2))/2} \\ &\ge \alpha_L 1_{B(0,R_3)}.\end{aligned}$$

Here how large we make n depends on R_3.

Now we want to use the earlier strategy

Lemma 3.8. *For any R there exists δ_M and R_4 and a range (s_*, s^*) such that for all $s \in (s_*, s^*)$ we have that*

$$\int T_s \left(\delta_{x_0}(x) 1_{|v|<R_2}(v)\right) \mathrm{d}v \ge 1_{|x| \le \delta_M}.$$

Proof. Again we define

$$\tilde{X}_t = x + vt.$$

We have that

$$|X_t - \tilde{X}_t| \le \frac{1}{2} t^2 \max_{\tau \le t} |\nabla_x U(X_\tau) + V_\tau|.$$

So as before we can bound

$$|X_t - \tilde{X}_t| \le Ct^2 \left(R + \frac{1}{2}R_4^2\right)^\eta.$$

So we now are in almost the same situation as the earlier lemma. However we need to worry about the fact that the Hamiltonian might be increasing if we go backwards in time so when we are estimating

$$|\tilde{X}_t(X_t^{-1}(x,v)) - X_t(X_t^{-1}(x,v))| \le Ct^2 e^t \left(R + \frac{1}{2}R_4^2\right)^\eta.$$

However this extra factor of e^t is negligible for very small t so the same proof works. □

For the Lyapunov function we suppose that K is an even function and note that

$$\int\int K(v-u)f(u)|v|^2 \mathrm{d}u\mathrm{d}v = \int\int f(u)K(w)|u+w|^2\mathrm{d}u\mathrm{d}w \leq \int f(u)(|u|^2+\delta^2)\mathrm{d}u.$$

Also,

$$\int\int K(v-u)f(u)v\mathrm{d}u\mathrm{d}v = \int\int f(u)K(w)(w+u)\mathrm{d}w\mathrm{d}u = \int f(u)u\mathrm{d}u.$$

So we look for a Lyapunov function of the form

$$M(x,v) = U(x) + \frac{1}{2}|v|^2 + ax\cdot v + b|x|^2.$$

We can calculate that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\int Mf\mathrm{d}x\mathrm{d}v &\leq \int f(a|v|^2 + 2bx\cdot v - ax\cdot\nabla_x U(x) + \delta^2 - |v|^2 - ax\cdot v)\\
\text{let } a = 1/2, b = 1/4 \quad &\leq \int f\left(\delta^2 + A/2 - \frac{1}{2}|v|^2 - \frac{\gamma_1}{2}|x|^2 - \frac{\gamma_2}{2}U(x)\right)\\
&\leq \int f\left(C - \frac{2}{3}\left(\frac{1}{2}|v|^2 + \frac{1}{2}x\cdot v + \frac{1}{4}|x|^2\right) - \frac{\gamma_2}{2}U(x)\right)\\
&\leq -\min\left\{\frac{2}{3},\frac{\gamma_2}{2}\right\}\int fM + C.
\end{aligned}$$

3.5.3 Subgeometric convergence

As with the other equations we can also show subgeometric convergence with weaker conditions on the confining potential

Theorem 3.10. *Suppose that $U(x)$ is a function satisfying*

$$|\nabla_x U(x)| \leq U(x)^\eta, \quad x\cdot\nabla_x U(x) \geq \gamma_1\langle x\rangle^\beta + \gamma_2 U(x) - A.$$

Where

$$\langle x\rangle = 1 + |x|^2,$$

and $\beta\in(0,1)$. Then the solution to the non-local diffusion equation converges to equilibrium in a weighted total variation norm in the following way. We define the function M by

$$M(x,y) = U(x) + \frac{1}{2}|v|^2 + \frac{1}{2}x\cdot v + \frac{1}{4}|x|^2.$$

Then there exits a constant $C>0$ such that

$$\|\mathcal{P}_t\delta_{z_1} - \mathcal{P}_t\delta_{z_2}\|_{TV} \leq C(M(z_1)+M(z_2))(1+t)^{-1/(1-\beta)},$$

and

$$\|\mathcal{P}_t\delta_z - \mu\|_{TV} \leq CM(z)(1+t)^{-1/(1-\beta)} + C(1+t)^{-\beta/(1-\beta)}.$$

Proof. As before we have already got the minorisation condition, we only need to prove a new Lyapunov condition. We take the same Lyapunov function as for the geometric case. The result follows in exactly the same way as for the linear BGK equation. □

Chapter 4

Hypocoercivity for the kinetic Fokker-Planck equation with a confining potential via Hairer and Mattingly's Wasserstein-1 Harris's theorem and Malliavin calculus

4.1 Hypocoercivity and hypoellipticity

In this chapter we return to one of the first equations which was studied in the context of hypocoercivity, the kinetic Fokker-Planck or Langevin equation

$$\partial_t f + v \cdot \nabla_x f - \nabla_x U \cdot \nabla_v f = \Delta_v f + \nabla_v \cdot (vf).$$

Here $\mu = M \exp(-|v|^2/2 - U(x))$ for some normalising constant M. Hypocoercivity for the kinetic Fokker-Planck equation has been shown by many authors. It was shown in $L^2(\mu^{-1})$ in [80]. This paper then inspired the mémoire of Villani [116] where he proves a general theorem in the first section which he then applies to the kinetic Fokker-Planck. The L^2 and H^1 results are also given as special cases of the theorems proven in [50] and [102] respectively.

The kinetic Fokker-Planck equation is an equation in the sum of squares form given in [116] with $B = v \cdot \nabla_x - \nabla_x U \cdot \nabla_v$ and $A = -\nabla_v$. Then

$$\partial_t f + Bf + A^* A f = 0.$$

This equation is also hypoelliptic. The hypocoercivity and hypoellipticity of some degenerate diffusions can be proved using similar techniques and the name hypocoercivity was inspired by this similarity. The main examples of this is the paper [80] where they prove hypocoercivity and hypoellipticity simultaneously using pseudo differential techniques and the new proof of hypoellipticity for the kinetic Fokker-Planck equation given in [79]. The link is expressed clearly in [79]. These proofs of both hypocoercivity and hypoellipticity for kinetic Fokker-Planck equation use crucially the fact that

$$[B, A] = -\nabla_x.$$

More generally both hypocoercivity and hypoellipticity rely on the diffusion being spread to the other direction seen by taking successive iterated commutators between the vector fields [81].

Some degenerate diffusions equations are also the Kolmogorov backwards equations for the law of the SDE

$$\mathrm{d}Z_t = \sum_i \tilde{A}_i \mathrm{d}W_t^i + \tilde{B}\mathrm{d}t.$$

Where the tilde vector fields are closely related to the the ones appearing in the PDE. In [116] (Part 1, Prop 5) Villani shows that all SDEs which converge to an equilibrium state have backwards equations which can be written in the form

$$\partial_t f + \sum_i A_i^* A_i f + Bf = 0.$$

This is the form for which it is possible to state his hypocoercivity theorem. Here the vector fields are different to those in the Itō SDE form of the equation. Hypoellipticity has been understood on the level of SDEs via Malliavin calculus see for example [92, 103]. The machinery of Malliavin calculus allow one to see how the effect of the Brownian motions is transferred along different directions given by the iterated commutators of the driving vector fields.

Kinetic Fokker-Planck equations were shown to converge to equilibrium in [94] using techniques from [96]. These works use probabilistic techniques, relying on Harris's Theorem which gives exponential convergence to equilibrium based on a Lyapunov condition and a minorization condition. The minorization condition is typically of the form that for all R there exists some probability measure ν and constant α such that for all z in $B(0, R)$ we have

$$f_t^z \geq \alpha\nu.$$

Here f_t^z is the solution to the PDE at time t, with initial condition δ_z.

These proofs do not give explicit constants and this lack of quantifiability arises when showing the minorisation condition. They first show that f_t^z has a density using hypoellipticity theory. Then they show via control theory that for some compact C then there is some $y \in C$ such that for any δ we have $t_1(\delta)$ with

$$\mathcal{P}_{t_1}(x, B_\delta(y)) > 0 \quad \forall x \in C.$$

They then use these to prove a minorisation condition. Its not clear how to make this argument quantitative as it would require us to be able to estimate $p_t(x, y)$ from below at a specific point and uses compactness arguments. As the proof of hypoellipticity can be made using Malliavin calculus it makes sense to ask whether the minorisation condition can be shown directly and quantitatively using Malliavin calculus. This would then allow one to prove hypocoercivity for the SDE quantitatively on the level of the SDE itself rather than via the PDE. Convergence to equilibrium in Wasserstein for the kinetic Fokker-Planck equation is shown very nicely in [52] by a direct coupling approach. In [52] they use a Lyapunov structure to show that the solution concentrates in the centre of the state space. Within this centre they show contraction in Wasserstein by using a mixture of reflection and synchronisation couplings. In this setting the reflection coupling should push the x coordinates of the processes towards each other and the synchronisation coupling should push the v-coordinates towards each other The final result of this paper is very similar to the one given here. However, our techniques for looking at the behaviour in the centre of the space are very different. We use a much less trajectorial viewpoint. This means we are unlikely to get as sharper constants as with a coupling approach. It does allow us to see how we are exploiting the hypoelliptic structure of the equations more clearly.

We could not show something as strong as the minorisation condition quantitatively. This is because we use Malliavin calculus to approximate our solutions by Gaussians for which spreading out in all directions is clear but we then get an error from this process which is not bounded in L^∞ as we would need to show minorisation. However this error is sufficiently well behaved that we can bound below the probability that any two solutions to the SDE started within a compact will be within a distance δ from each other at some time T, i.e.

$$\inf_{|x|,|y|\leq C} \sup_{\Gamma \in \mathcal{C}(\mathcal{P}_T^*\delta_x, \mathcal{P}_T^*\delta_y)} \Gamma\{(x',y') : d(x',y') < \delta\} \geq a.$$

Where

$$\mathcal{C}(\mathcal{P}_T^*\delta_x, \mathcal{P}_T^*\delta_y)$$

is the set of couplings of the solutions at time T. This is one of the assumptions of the Wasserstein-1 version of Harris's theorem proved by Hairer and Mattingly in [72] to show spectral gaps in Wasserstein for the stochastically forced Navier-Stokes equation.

Therefore the goal is to show exponentially fast convergence to equilibrium in a weighted Wasserstein-1 distance for the kinetic-Fokker Planck or Langevin equation

$$\mathrm{d}X_t = V_t \mathrm{d}t, \tag{4.1}$$

$$\mathrm{d}V_t = -(V_t + \nabla_x U(X_t))\mathrm{d}t + \sqrt{2}\mathrm{d}W_t. \tag{4.2}$$

4.2 Harris's theorem in Wasserstein

We are going to use the version of Harris's theorem in a Wasserstein-1 distance proved by Hairer and Mattingly in [72] for use in giving explicit rates of convergence to equilibrium for the 2D Navier-Stokes equation. We first introduce the distance for some function L

$$\rho_r(x,y) = \inf_\gamma \int_0^1 L^r(\gamma(t))\|\dot{\gamma}(t)\|\mathrm{d}t,$$

where r is an exponent and the infimum runs over all paths γ between x and y. Let us write $\rho_1 = \rho$.

The assumptions of this theorem are

Assumption 1. *There exists a continuous function $L \geq 1$ which has the following properties:*

1. *There exist strictly increasing functions L_*, L^* such that*

$$L_*(|z|) \leq L(z) \leq L^*(|z|),$$

with $\lim_{a\to\infty} L_*(a) = \infty$.

2. There exist constants C and $\kappa \geq 1$ such that for all a

$$aL^*(a) \leq CL_*^{\kappa}(a).$$

3. Finally, there exist constants $C_* > 0, 0 < r_0 < 1$ and a function $\xi : [0,1] \to [0,1]$ which is non increasing with $\xi(1) < 1$ such that for every h with $|h| = 1$ we have

$$L^r(\Phi_t(z))(1 + \|\nabla_z \Phi_t(z)h\|) \leq C_* L^{r\xi(t)}(z),$$

for every z and every $r \in [r_0, 2\kappa]$ and every $t \in [0,1]$. Here Φ_t is the flow map which takes an initial position z to the random variable which is the solution to the SDE at time t.

Assumption 2. *There exists a $C_1 > 0$ and $p \in [0,1)$ so that for every $\alpha \in (0,1)$ there exists positive $T(\alpha), C(\alpha)$ with*

$$\|\nabla_z \mathcal{P}_t \phi(z)\| \leq L(z)^p \left(C(\alpha)\sqrt{(\mathcal{P}_t|\phi|^2)(z)} + \alpha \sqrt{(\mathcal{P}_t \|\nabla_z \phi\|^2)(z)} \right),$$

for every $z \in \mathbb{R}^d$, $\phi \in C_b^1$ and every $t > T(\alpha)$.

Assumption 3. *For any $C > 0, r \in (0,1)$ and $\delta > 0$, there exists a T_0 so that for any $T \geq T_0$ there exists and $a > 0$ so that*

$$\inf_{|z_1|,|z_2| \leq C} \sup_{\pi \in \Pi(\mathcal{P}_T^* \delta_{z_1}, \mathcal{P}_T^* \delta_{z_2})} \pi\{(z_1', z_2') : \rho_r(z_1', z_2') < \delta\} \geq a.$$

Here $\Pi(\mu,\nu)$ is the set of couplings of μ and ν. In our situation we actually only use the coupling where they are independent. This depends on L through the distance ρ but not very strongly. We can rewrite this as

$$\inf_{|z_1|,|z_2| \leq C} \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} 1_{\rho_r(z_1', z_2') < \delta} \mathcal{P}_T^*(\mathrm{d}z_1') \mathcal{P}_T^*(\mathrm{d}z_2') \geq a.$$

Then the theorem is

Theorem 4.1 (Hairer & Mattingly 2008)**.** *If the semigroup $\mathcal{P}_t$ satisfies the assumptions above then for all μ, ν there exists C and λ which we can calculate from the constants in the assumptions so that*

$$\mathcal{W}_\rho(\mathcal{P}_t^* \mu, \mathcal{P}_t^* \nu) \leq Ce^{-\lambda t} \mathcal{W}_\rho(\mu, \nu),$$

for any μ, ν. Here $\mathcal{W}_\rho$ is the Wasserstein-1 distance corresponding to the distance ρ. i.e.

$$\mathcal{W}_\rho(\mu,\nu) = \inf_{\pi \in \Pi} \int \rho(z_1, z_2) \pi(\mathrm{d}z_1, \mathrm{d}z_2).$$

Π is the set of couplings of μ and ν probability measures on $\mathbb{R}^{2d}$ which have marginals μ and ν on the first and last d dimensions.

Our goal is to verify each of these assumptions with explicit constants. I will briefly describe the strategy.

- The first assumption is a Lyapunov structure. We verify this using more tools from [72] and known Lyapunov functions for the kinetic Fokker-Planck equation from [94].

- The second assumption is a gradient bound. This is an additional condition needed for the Wasserstein proof to work and is not present in Harris's theorem in any form. We verify this using tools similar to those of Bakry-Emery calculus. Some work on Hypoelliptic diffusions via Bakry-Emery stuff has been done in [12, 98] and papers referenced therein. We need the Hessian of the confining potential to be bounded for this to work but it seems plausible to relax this assumption.

- The third assumption is a kind of uniform boundedness condition. We verify this using Malliavin calculus by showing that for any positive the solution spreads out in all directions. This part should work for any equation satisfying the Hörmander bracket condition provided that it also satisfies the very strong assumptions that all the vector fields appearing in the commutator conditions are constant.

Theorem 4.2. *Suppose that $\mathcal{P}_t$ is a semigroup corresponding to the solution to the kinetic Fokker-Plank with the confining potential U being a smooth function satisfying*

$$Hess(U)(x) \leq M, \quad x \cdot \nabla_x U(x) \geq c_1 U(x) + c_2 x^2 - c_3$$

for some strictly positive constants M, c_1, c_2, c_3. Then we can choose constants a_* and k depending on these other constants to define the function

$$L(x,v) = \exp\left(a_*\left(v^2 + 2U(x) + 2kx^2 + kxv\right)\right).$$

We define ρ corresponding to L with

$$\rho(z_1, z_2) = \inf_{\gamma \in \Gamma} \int_0^1 L(\gamma(t)) \|\dot{\gamma}(t)\| \mathrm{d}t.$$

Here Γ is the set of all C^1 paths between z_1 and z_2. Then if $\mathcal{W}_\rho$ is the Wasserstein-1 distance associated to ρ we have constants $C > 0$ and $\lambda > 0$ which we can calculate explicitly such that

$$\mathcal{W}_\rho(\mathcal{P}_t\mu, \mathcal{P}_t\nu) \leq Ce^{-\lambda t}\mathcal{W}_\rho(\mu, \nu).$$

Remark. *The conditions on U are equivalent to requiring it to behave roughly like a quadratic at infinity. This allows it to have 'bad' behaviour on a compact set. For example multiple wells or being flat in large areas. In particular this would allow for the double well potential which behaves quadratically at infinity in 1D.*

Remark. *$\mathcal{W}_\rho(\mu, \nu)$ bounds the Wasserstein 1, distance associated to the euclidean metric. We can see that there exists some M such that*

$$|z_1 - z_2| \leq \rho(z_1, z_2) \leq |z_1 - z_2| \exp\left(M\left(|z_1|^2 + |z_2|^2\right)\right).$$

We structure the paper as follows. We split the proof of Theorem 4.2 into three parts relating to the three assumptions. We then deal with each of these parts separately. We rely on the theorem of Hairer and Mattingly but in order to make it clear how the proofs work we include a proposition showing how each assumptions will allow us to show contraction for a different part of the space. These propositions follow closely Hairer and Mattingly's proof of theorem 4.1 and are not original. They are intended for expository purposes and to make this chapter more self contained.

Proof of 4.2. We prove Theorem 4.2 by showing that we can verify all the assumptions of Theorem 4.1 and then applying this result. Assumption 1 is verified in Lemma 4.2. Assumption 2 is verified in Lemma 4.3. Assumption 3 is verified in Lemma 4.5.

We also give the proof of 4.1 in our context. We note that for any distance d we have

$$\mathcal{W}_{1,d}(\mathcal{P}_t\mu, \mathcal{P}_t\nu) \leq \inf_{\pi\in\Pi(\mu,\nu)} \int \mathcal{W}_{1,d}(\mathcal{P}_t\delta_{z_1}, \mathcal{P}_t\delta_{z_2})\pi(\mathrm{d}z_1, \mathrm{d}z_2).$$

Therefore if we can show for each z_1, z_2 that

$$\mathcal{W}_{1,d}(\mathcal{P}_t\delta_{z_1}, \mathcal{P}_t\delta_{z_2}) \leq \alpha d(z_1, z_2)$$

then we have

$$\mathcal{W}_{1,d}(\mathcal{P}_t\mu, \mathcal{P}_t\nu) \leq \alpha\mathcal{W}(\mu, \nu).$$

We do not work directly with the distance ρ and instead look at the equivalent distance

$$d(z_1, z_2) = \left(\frac{\rho_r(z_1, z_2)}{\delta} \wedge 1\right) + \beta\rho(z_1, z_2).$$

For any $r < 1$ and δ, β to be chosen later.

In Proposition 4.1, we show that there exists some K such that for all $r \in [r_0, 1)$ and for all $\beta \in (0, 1)$ we have that $\mathcal{P}_t$ gives a contraction between measures δ_{z_1} and δ_{z_2} in $\mathcal{W}_{1,d}$ uniformly over the set $\rho(z_1, z_2) > K$ and uniformly over all t sufficiently large. In Proposition 4.2, we then show that there exists an $r \in [r_0, 1)$ and a $\delta > 0$ such that $\mathcal{P}_t$ is a contraction in $\mathcal{W}_{1,d}$ uniformly over the set $\rho_r(z_1, z_2) < \delta$, $\beta \in (0, 1)$ and t sufficiently large. Finally in Proposition 4.3, we show that for this given r, δ and K we can choose β such that, for every t sufficiently large, $\mathcal{P}_t$ gives a contraction in $\mathcal{W}_{1,d}$ uniformly over the set $\rho(z_1, z_2) \leq K$ and $\rho_r(z_1, z_2) > \delta$.

□

4.3 Proofs

4.3.1 Assumption 1

We would like to show that these assumptions hold with explicit constants for the kinetic Fokker-Planck equation. We begin with assumption 1 where our treatment closely mirrors that of Hairer and Mattingly in [72]. Here the Lyapunov function we find is essentially the exponential of the Lyapunov function used by Mattingly, Stuart and Higham in [94]. We write $J_{0,t} = \nabla_z\Phi_{0,t}(z)$

Let us define

$$Q(x, v) = |v|^2 + 2U(x) + \frac{1}{2}|x|^2 + x \cdot v, \quad P_k(x, v) = 2(|x|^2 + |v|^2 + kU(x)).$$

We will choose k later.

Lemma 4.1. *Let U be a smooth function satisfying that for all x $x \cdot \nabla_x U(x) \geq c_1U(x) + c_2x^2 - c_3$, for strictly positive constants c_1, c_2, c_3 and $Hess(U) \leq M$ for some $M > 0$. Define $L_a(x, v) = \exp(aQ(x, v))$. Then we show there exists $a_* > 0$ such that, for $0 < a \leq a_*$ and uniformly over*

$t \in [0,1]$, *there is a constant* $\beta > 0$ *such that*

$$\mathbb{E}(L_a(\Phi_t(x,v))\|J_{0,t}\|) \leq L_{ae^{-\beta t/4}}(x,v).$$

Proof. Note first that we may as well choose $c_1 \leq 1$. We have that

$$\mathrm{d}(aQ(Z_s)) = \left(-a|V_s|^2 - aX_s \cdot \nabla_x U(X_s) + 2a\right)\mathrm{d}s + a(X_s + 2V_s)\mathrm{d}W_s = -aH_s\mathrm{d}s + a(X_s + 2V_s)\mathrm{d}W_s.$$

Where

$$H_s = |V_s|^2 + X_s \cdot \nabla_x U(X_s) - 2.$$

Therefore with $k = c_1$ we have that as functions of z

$$H_s(z) \leq \beta P_k(z) + c_3,$$

for some β which depends on c_1, c_2. We also have that $Q(z) \leq P(z)/c_1$. Now we define

$$Y_s = e^{\gamma(s-t)}aQ(Z_s) + \gamma\int_0^s e^{\gamma(r-t)}ac_1P(Z_s), \quad M_s = \int_0^s e^{\gamma(r-t)}a(2V_r + X_r)\mathrm{d}W_r.$$

Differentiating this gives us that,

$$dY_s = e^{\gamma(s-t)}(aH_s + a\gamma(Q(Z_s) + c_1P(Z_s)))\mathrm{d}s + \mathrm{d}M_s.$$

Hence for $s < t$ we have

$$\begin{aligned} Y_s &\leq M_s + Y_0 + a\int_0^s e^{\gamma(r-t)}(H_r + \gamma(Q(Z_r) + c_1P(Z_r)))\mathrm{d}r \\ &\leq M_s + Y_0 + a\int_0^s e^{\gamma(r-t)}((2\gamma - \beta)P(Z_r) + 2 + c_3)\mathrm{d}r. \end{aligned}$$

Therefore we have that

$$Y_s \leq M_s + Y_0 + C + a\int_0^s e^{\gamma(r-t)}(2\gamma - \beta)P(Z_r)\mathrm{d}r.$$

We now note that we have

$$Y_0 = ae^{-\gamma t}Q(Z_0),$$

and that

$$Y_t \geq aQ(Z_t) + ac_1\gamma e^{-\gamma t}\int_0^t P(Z_s)\mathrm{d}z.$$

We also have that

$$\langle M\rangle_s \leq 16a^2\int_0^s e^{\gamma(r-t)}P(Z_r)\mathrm{d}r,$$

therefore for every $s < t$ we have

$$M_s - (\beta - 2\gamma)c_1a\int_0^s e^{\gamma(r-t)}P(Z_r) \leq M_s - \frac{c_1(\beta - 2\gamma)}{16a}\langle M\rangle_s.$$

The exponential martingale inequality gives that

$$\mathbb{P}\left(\sup_{s\leq t}\left(M_s - \frac{c_1(\beta-2\gamma)}{16a}\langle M\rangle_s\right) > K\right) \leq \exp\left(-\frac{Kc_1(\beta-2\gamma)}{8a}\right).$$

Now we choose $\gamma = \beta/4$ this gives

$$Y_s - Y_0 - C \leq M_s - \frac{\beta}{2}ac_1\int_0^s e^{\gamma(r-t)}P(Z_r)\mathrm{d}r \leq M_s - \frac{c_1\beta}{32a}\langle M\rangle_s.$$

Combining this with our earlier assumptions we have

$$aQ(Z_t) + ac_1\frac{\beta}{4}e^{-\beta t/4}\int_0^t P(Z_s)\mathrm{d}s - ae^{-\beta t/4}Q(Z_0) - C \leq M_s - \frac{\beta c_1}{32a}.$$

Therefore,

$$\mathbb{P}\left(\exp\left(aQ(Z_t) + ac_1\frac{\beta}{4}e^{-\beta t/4}\int_0^t P(Z_s)\mathrm{d}s - ae^{-\beta t/4}Q(Z_0) - C\right) > x\right) \leq x^{-c_1\beta/16a}.$$

We can make a smaller than $a^* = \beta c_1/32$ we have the exponent is bigger than 2 so we integrate to get

$$\mathbb{E}\left(\exp\left(aQ(Z_t) + ac_1\frac{\beta}{4}e^{-\beta t/4}\int_0^t P(Z_s)\mathrm{d}s - ae^{-\beta t/3}Q(Z_0) - C\right)\right) \leq \frac{c_1\beta}{c_1\beta - 16a}.$$

Therefore,

$$\mathbb{E}\left(\exp\left(aQ(Z_t) + ac_1\frac{\beta}{4}e^{-\beta t/4}\int_0^t P(Z_s)\mathrm{d}s\right)\right) \leq C(a)\exp\left(aQ(Z_0)\right).$$

Now we have that

$$\mathrm{d}J_{0,t} = \begin{pmatrix} 0 & I \\ -\mathrm{Hess}(U)(X_t) & -1 \end{pmatrix} J_{0,t}\mathrm{d}t.$$

It therefore follows that

$$\mathrm{d}\|J_{0,t}h\| = \frac{(J_{0,t}h)^T}{\|J_{0,t}h\|}\begin{pmatrix} 0 & I \\ -\mathrm{Hess}(U)(X_t) & -1 \end{pmatrix} J_{0,t}h\mathrm{d}t \leq (1+M)\|J_{0,t}h\|\mathrm{d}t.$$

This means that for every $t \in [0,1]$ we have

$$\|J_{0,t}h\| \leq e^{1+M}.$$

Then we have that for $t \in [0,1]$, h a unit vector, $\eta > 0$

$$\|J_{0,t}h\| \leq e^{1+M}\exp\left(\left(\eta\int_0^t (|X_s|^2 + |V_s|^2 + kU(X_s))\mathrm{d}s\right)\right).$$

Therefore for any $a < a_*$ and η small enough in terms of a we have

$$\|J_{0,t}h\| \leq e^{1+M}\exp\left(ac_1\frac{\beta}{4}e^{-\beta t/4}\int_0^t P(Z_s)\mathrm{d}s\right).$$

This combined with our earlier result gives the lemma. □

Lemma 4.2. *Provided that U is a smooth function satisfying*

$$x\cdot\nabla_x U(x)\geq c_1U(x)+c_2x^2-c_3,\quad HessU(x)\leq M$$

for some positive constants we can choose a_*, k such that

$$L(x,v)=\exp\left(a_*\left(v^2+2U(x)+2kx^2+kxv\right)\right)$$

is a function satisfying assumption 1.

Proof. We can add a constant in the definition of U so we may as well take $U\geq 0$. Since $Hess(U)\leq M$ we have

$$\frac{3}{4}(|x|^2+|v|^2)\leq Q(x,v)\leq(2+M)(|x|^2+|v|^2).$$

We also have that

$$|z|e^{a(2+M)|z|^2}\leq\frac{1}{a}e^{(3+M)|z|^2}\leq\frac{1}{a}\left(e^{3a|z|^2/4}\right)^{4(3+M)/3}\leq\frac{1}{a}e^{4a(3+M)Q(z)/3}.$$

Therefore if $8a(3+M)/3\leq a^*$ Then by lemma 4.1 we have that

$$\mathbb{E}\left(\left(|\Phi_t(z)|e^{a(2+M)|\Phi_t(z)|^2}\right)^2\right)\leq\frac{1}{a}e^{4a(3+M)e^{-\beta t/4}Q(z)/3}$$

Therefore if we set

$$a_*=3a^*/8(3+M)$$

then we can set

$$L(z)=e^{a_*Q(z)},L_*(z)=e^{3a_*|z|^2/4},L^*(z)=e^{(2+M)a_*|z|^2}.$$

Then our calculation shows that

$$L_*\leq L\leq L^*,$$

and furthermore that

$$|z|L^*(|z|)\leq L_*(z)^\kappa,$$

with $\kappa=3(3+M)/3$. Then lemma 4.1 shows that

$$\mathbb{E}(L^r(\Phi_t(z)))\leq L^{re^{-\beta t/4}}(z),$$

for all $r\leq 2\kappa$. □

Now we briefly describe how the proof of Hairer and Mattingly uses this lemma to show convergence for $\rho(z_1,z_2)\geq 4C_1$ with C_1 given below.

Proposition 4.1. *If we define ρ as above then for every $\alpha\geq 1/2, T_1>0$ there exists constants C_1, C such that for all $t\geq T_1$*

$$\mathbb{E}(\rho(\Phi_t(z_1),\Phi_t(z_2)))\leq C\rho(z_1,z_2),$$

$$\mathbb{E}(\rho(\Phi_t(z_1), \Phi_t(z_2))) \leq C_1 + \alpha\rho(z_1, z_2).$$

Furthermore, there exists some radius R_2 such that if $|z_1|$ or $|z_2| \geq R_2$ then,

$$\mathbb{E}(\rho(\Phi_t(z_1), \Phi_t(z_2))) \leq \alpha\rho(z_1, z_2).$$

Proof. Fix z_1, z_2, $t > T_1$ then there exists some curve joining z_1, z_2 such that

$$\int_0^1 L^r(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s \leq \rho_r(z_1, z_2) + \epsilon.$$

So then we can evolve every point along this curve by Φ_t to make a curve joining $\Phi_t(z_1), \Phi_t(z_2)$. Using lemma 4.2 this gives

$$\mathbb{E}(\rho(\Phi_t(z_1), \Phi_t(z_2)) \leq \mathbb{E}\left(\int_0^1 L(\Phi_t(\gamma(s)))|J_{0,t}\dot{\gamma}(s)|\mathrm{d}s\right) \leq C\int_0^1 L(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s \leq C(\rho(z_1, z_2) + \epsilon).$$

ϵ was arbitrary. In fact we could have written

$$\mathbb{E}(\rho(\Phi_t(z_1), \Phi_t(z_2))) \leq C\int_0^1 L^{e^{-\beta t/4}}(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s.$$

Then since L grows at infinity there is some R so that $CL^{e^{-\beta t/4}}(z) \leq \alpha L(z)$ for $|z| \geq R$. Therefore

$$\mathbb{E}(\rho(\Phi_t(z_1), \Phi_t(z_2))) \leq \alpha\rho(z_1, z_2) + \int_0^1 L(\gamma(s))|\gamma\dot{(s)}|1_{\gamma(s)\in B(0,R)}\mathrm{d}s$$

Now we recall that there exists constants m and M so that

$$Ce^{m|z|^2} \leq L(z) \leq e^{M|z|^2}.$$

If we replace the segment of γ in $B(0, R)$ by a straight line segment this means we can never need to pick up more than

$$Re^{MR^2} + \epsilon$$

in our integral while travelling through $B(0, R)$ so we have that

$$\mathbb{E}(\rho(\Phi_t(z_1), \Phi_t(z_2))) \leq \alpha\rho(z_1, z_2) + CRe^{MR^2}.$$

So we know we are contractive if $\rho(z_1, z_2) \geq 4C_1$ say, and also we can see from this proof that we will be contractive whenever almost optimal paths between z_1, z_2 do not pass through the $B(0, R)$. We can calculate that the distance, ρ, from z to $B(0, R)$ is bounded below by

$$C\int_R^{|z|} e^{mr^2}\mathrm{d}r.$$

Therefore we have R_2 such that if $|z| > R_2$ then this will be greater than $4C_1$. This means that if γ is a path from z_1, z_2 going through $B(0, R)$ with $|z_1|$ or $|z_2|$ greater than R_2 then

$$\int_\gamma L(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s \geq 4C_1.$$

This means that if $|z_1| \geq R_2$ or $|z_2| \geq R_2$ then either close to optimal paths do not go through $B(0,R)$ or $\rho(z_1,z_2) \geq 4C_1$. Therefore

$$\mathbb{E}(\rho(\Phi_t(z_1),\Phi_t(z_2))) \leq \alpha\rho(z_1,z_2),$$

for $|z_1| \geq R_2$ or $|z_2| \geq R_2$. □

4.3.2 Assumption 2

Assumption 2 looks very similar to the gradient bounds found in Malliavin's proof of Hörmander's theorem see for example [103]. It seems to be more of a technical challenge than anything else to make the estimates here explicit. However, it is simpler to use more standard hypocoercive techniques based on point wise Bakry-Emery style estimates on the semigroup $\mathcal{P}_t$. Let us write

$$\Gamma(f,g) = 2\nabla_x f \cdot \nabla_x g - \nabla_x f \cdot \nabla_v g - \nabla_v f \cdot \nabla_x f + 2\nabla_v f \cdot \nabla_v g.$$

Now write

$$L = \Delta + v \cdot \nabla_x - v \cdot \nabla_v - \nabla_x U \cdot \nabla_v$$

this is the forwards operator for the solution to the SDE. We set $\Gamma_2(f) = L\Gamma(f,f) - 2\Gamma(f, Lf)$.

Lemma 4.3. *For $\mathcal{P}_t$ the semigroup associated to the SDE when U'' is bounded we have that for an explicit constant C_M*

$$|\nabla_x \mathcal{P}_t f|^2 + |\nabla_v \mathcal{P}_t f|^2 \leq C_M \mathcal{P}_t(f^2) + 3e^{-t/3}\mathcal{P}_t\left(|\nabla_x f|^2 + |\nabla_v f|^2\right).$$

Proof.

$$\begin{aligned}
\Gamma_2(f) =& 4|\nabla_x\nabla_v f|^2 - 4\nabla_x\nabla_v f : \nabla_v\nabla_v f + 4|\nabla_v\nabla_v f|^2 + 4\nabla_x f \mathrm{Hess}(U)\nabla_v f - 2\nabla_v f\mathrm{Hess}(U)\nabla_v f \\
&+ 2|\nabla_x f|^2 - 2\nabla_x f \cdot \nabla_v f + 4|\nabla_v f|^2 - 4\nabla_x f \cdot \nabla_v f \\
\geq& 4\nabla_x \mathrm{Hess}(U)\nabla_v f - 2\nabla_v \mathrm{Hess}(U)\nabla_v f + 2|\nabla_x f|^2 - 6\nabla_x f \cdot \nabla_v f + 4|\nabla_v f|^2 \\
\geq& (2 - 3\epsilon_1 - 2M\epsilon_2)|\nabla_x f|^2 + \left(4 - \frac{3}{\epsilon_1} - \frac{2M}{\epsilon_2} - 2M\right)|\nabla_v f|^2
\end{aligned}$$

We set $\epsilon_1 = 1/6$ and $\epsilon_2 = 1/4M$ to get

$$\Gamma_2(f) \geq |\nabla_x f|^2 - (14 + 6M^2 + 2M)|\nabla_v f|^2.$$

Let $\tilde{\Gamma}(f) = \Gamma(f) + (15 + 6M^2 + 2M)f^2$, and write $C_M = 15 + 6M^2 + 2M$. Then we get

$$L\tilde{\Gamma}(f) - 2\tilde{\Gamma}(f, Lf) \geq |\nabla_x f|^2 + |\nabla_v f|^2 \geq \frac{1}{3}\Gamma(f) = \frac{1}{3}\left(\tilde{\Gamma}(f) - C_M f^2\right).$$

Therefore, let

$$\psi(s) = \mathcal{P}_s\tilde{\Gamma}(\mathcal{P}_{t-s}(f)).$$

Then

$$\dot{\psi}(s) \geq \frac{1}{3}\left(\mathcal{P}_s\tilde{\Gamma}(\mathcal{P}_{t-s}f) - C_M\mathcal{P}_s(\mathcal{P}_{t-s})^2\right)$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}s}(e^{-s/3}\psi(s)) \geq -\frac{C_M}{3}e^{-s/3}\mathcal{P}_s(\mathcal{P}_{t-s}f)^2 \geq -\frac{C_M}{3}e^{-s/3}\mathcal{P}_t(f^2).$$

So

$$e^{-s/3}\psi(s) - \psi(0) \geq -C_M\left(1 - e^{-s/3}\right)\mathcal{P}_t(f^2)$$

which means that

$$e^{-t/3}\mathcal{P}_t(\Gamma(f)) - \Gamma(\mathcal{P}_tf) - C_M(\mathcal{P}_tf)^2 \geq -C_M\mathcal{P}_t(f^2).$$

Rearranging this gives

$$\Gamma(\mathcal{P}_tf) + C_M(\mathcal{P}_tf)^2 \leq C_M\mathcal{P}_t(f^2) + e^{-t/3}\mathcal{P}_t(\Gamma(f)).$$

We also have that

$$|\nabla_x f|^2 + |\nabla_v f|^2 \leq \Gamma(f) \leq 3\left(|\nabla_x f|^2 + |\nabla_v f|^2\right).$$

So we have that

$$|\nabla_x\mathcal{P}_tf|^2 + |\nabla_v\mathcal{P}_tf|^2 \leq C_M\mathcal{P}_t(f^2) + 3e^{-t/3}\mathcal{P}_t\left(|\nabla_x f|^2 + \nabla_v f|^2\right).$$

□

Now we look at how this is used to show convergence in the main theorem. We define a new metric

$$d(z_1,z_2) = \left(\frac{\rho_r(z_1,z_2)}{\delta}\wedge 1\right) + \beta\rho(z_1,z_2).$$

We see that for $\rho(z_1,z_2) > 4C_1$ proposition 4.1 still gives a contraction in this metric for every β.

Proposition 4.2. *If $\rho_r(z_1,z_2) < \delta$ then we have that for t sufficiently large*

$$\mathcal{W}_{1,d}(\mathcal{P}_t\delta_{z_1},\mathcal{P}_t\delta_{z_2}) \leq \gamma d(z_1,z_2)$$

for some explicit $\gamma < 1$.

Proof. In this section we want to use the dual Lipschitz formulation of the Wasserstein 1 distance. We have that

$$\mathcal{W}_{1,d}(\mathcal{P}_t\delta_{z_1},\mathcal{P}_t\delta_{z_2}) = \sup_\phi(\mathcal{P}_t\phi(z_1) - \mathcal{P}_t\phi(z_2)).$$

Here the infimum is taken over all Lipschitz ϕ with $|\phi|_{Lip} \leq 1$. In fact by density and adding and subtracting we can take a supreme over $phi \in C^1$ with $\phi(0) = 0$. If ϕ is such a function then

$$|\phi(z)| \leq (1+\beta)|z|L^*(z), |\nabla\phi(z)| \leq (1/\delta + \beta)L^*(z).$$

Therefore by lemma 4.3 and lemma 4.1 we have that

$$|\nabla\mathcal{P}_t\phi(z)| \leq L^{\kappa e^{-\beta t/4}}(z)(C + 3e^{-t/3}(1/\delta + \beta))$$

Therefore for t sufficiently large so that $\kappa e^{-\beta/4} \leq r$ and $3e^{-t/3} \leq 1/4$ we have that

$$|\nabla\mathcal{P}_t\phi(z)| \leq (\delta(C+2) + 1/4)\frac{1}{\delta}L^r(z).$$

Now take $\delta \le 1/2(C+2)$ so we have

$$|\nabla \mathcal{P}_t\phi(z)| \le \frac{3}{4}\frac{1}{\delta}L^r(z).$$

So we have that

$$\mathcal{P}_t\phi(z_1) - \mathcal{P}_t\phi(z_2) \le \int_0^1 \nabla \mathcal{P}_t\phi(\gamma(s)) \cdot \dot{\gamma}(s)\mathrm{d}s \le \frac{3}{4}\frac{1}{\delta}\int_0^1 L^r(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s.$$

For any path γ joining z_1 and z_2. Therefore we have

$$\mathcal{P}_t\phi(z_1) - \mathcal{P}_t\phi(z_2) \le \frac{3}{4}\frac{1}{\delta}\rho_r(z_1, z_2).$$

Since $\rho_r(z_1, z_2) \le \delta$ this means

$$\mathcal{W}_{1,d}(\mathcal{P}_t\delta_{z_1}, \mathcal{P}_t\delta_{z_2}) \le \frac{3}{4}d(z_1, z_2).$$

□

4.3.3 Assumption 3

Before starting we need some material from Malliavin calculus

Malliavin Calculus

The material in this section is all standard and follows [104, 71]. Malliavin calculus is a way of ‘differentiating’ a random variable whose randomness comes from some Brownian motion with respect to this Brownian motion. Since it is the driving Brownian motion which causes the diffusive behaviour of the solutions to SDEs, the Malliavin derivative allows us to measure the strength and direction of this diffusion. We will denote the Malliavin derivative of a function by $\mathcal{D}F$, this derivative is in fact a function and if F is a functional of $W_s, 0 \le s \le t$ then the Malliavin derivative is a function on $[0,t]$ we denote the evaluation of this function at a particular time s by $\mathcal{D}_sF$. We quickly introduce some of the definitions in Malliavin calculus. First we need to know what kind of functions can be differentiated. Let

$$\Omega = C_0 = \{f \mid f \in C([0,T]^n; \mathbb{R}^d), f(0) = 0\},$$

be Wiener space, and P the Wiener measure. Let H be the Hilbert space $H = L^2([0,T])$. Then we define a simple type of Weiner functional

$$W : H \to \mathbb{R}, \quad W(h) = \int_0^T h(t)\mathrm{d}W_t$$

by Ito integration. We have that $\mathcal{D}W(h) = h$. For each $h \in H, W(h)$ is a random variable. Let $\mathscr{G}$ be the sigma-algebra generated by $\{W(h) : h \in H\}$. We want to look a Weiner functionals which are in the Hilbert space G,

$$G = L^2(\Omega, \mathscr{G}, P).$$

The Malliavin derivative operator is $\mathcal{D} : G \to H$ is a closable, unbounded operator much like the weak derivative operator on L^2. Since, we are dealing mainly with SDEs we wish to know how to find the Malliavin derivative of the solution to an SDE. If we work purely formally we can derive an SDE for the Malliavin derivative to an SDE, writing in integral form we have

$$Z_t = Z_0 + \sum_{k=1}^{n} \int_0^t A_k(Z_s)\mathrm{d}W_{k,s} + \int_0^t B(Z_s)\mathrm{d}s$$

then we can formally take derivatives

$$\mathscr{D}_r^k Z_t = A_k(Z_r) + \sum_{j=1}^{n} \int_r^t \nabla A_j(Z_s) \cdot \mathscr{D}_r^k(Z_s)\mathrm{d}W_{j,s} + \int_r^t \nabla B(Z_s) \cdot \mathscr{D}_r^k(Z_s)\mathrm{d}s.$$

Here the k in the exponent corresponds to the Malliavin derivative with respect to the k^{th} Brownian motion. The Malliavin derivative can be constructed rigorously and in the case that A_k are smooth and uniformly Lipschitz it can be shown that $\mathscr{D}_r^k$ will satisfy this SDE, see [104, 71].

We now wish to look at our solution in a different form. If we write the map

$$\Phi_{s,t}^{\omega}(Z_s) = Z_t,$$

the solution map. Then we can differentiate with respect to the initial conditions to get

$$\partial\Phi_{s,t} = J_{s,t}.$$

Then we would like to write an SDE for $J_{s,t}$. Let us write

$$J_{s,t}Z_s = Z_s + \int_s^t \nabla A_k(Z_r) \cdot J_{s,r}Z_s\mathrm{d}W_{k,r} + \int_s^t \nabla B(Z_r) \cdot J_{s,r}Z_s\mathrm{d}r.$$

Comparing this with the SDE for $\mathscr{D}_s Z_t$ shows that, formally anyway,

$$\mathscr{D}_s Z_t = J_{s,t}A(Z_s).$$

Furthermore we can write an SDE for $J_{s,t}$ on its own in both Ito and Stratanovich form.

$$\begin{aligned} J_{s,t} &= I + \sum_{k=1}^{n} \int_s^t \nabla A_k(Z_r) \cdot J_{s,r}\mathrm{d}W_{k,r} + \int_s^t \nabla B(Z_r) \cdot J_{s,r}\mathrm{d}r, \\ &= I + \sum_{k=1}^{n} \int_s^t \nabla A_k(Z_r) \cdot J_{s,r} \circ \mathrm{d}W_{k,r} + \int_s^t \nabla A_0(Z_r) \cdot J_{s,r}\mathrm{d}r. \end{aligned}$$

We also notice that as

$$\Phi_{s,t} = \Phi_{r,t} \circ \Phi_{s,r}$$

the chain rule gives us that

$$J_{s,t} = J_{r,t}J_{s,r}.$$

We can also show that $J_{s,t}$ is invertible by writing a suitable SDE for $J_{s,t}$ and showing that the solution will not blow up. This lack of blow up comes from global controls on the size of ∇A and

∇B which we would like to impose. This SDE is

$$J_{s,t}^{-1} = I - \sum_{k=1}^{n} \int_s^t J_{s,r}^{-1} \nabla A_k(Z_r) \circ \mathrm{d}W_{k,r} - \int_s^t J_{s,r}^{-1} \nabla B(Z_r) \mathrm{d}r.$$

Putting these two facts together gives that

$$J_{s,t} = J_{0,t} J_{0,s}^{-1} \Rightarrow \mathscr{D}_s Z_t = J_{0,t} J_{0,s}^{-1} A(Z_s).$$

This is useful because $J_{0,s}^{-1} A(Z_s)$ is a measurable function of $Z_r, r \leq s$ so we could write an SDE purely on this quantity. This will be useful later, we do this in Stratanovich form where V is any smooth bounded vector field,

$$\begin{aligned}
\circ \mathrm{d} \left(J_{0,t}^{-1} V(Z_t) \right) &= \left(\circ \mathrm{d} J_{0,t}^{-1} \right) V(Z_t) + J_{0,t}^{-1} \left(\mathrm{d} V(Z_t) \right) \\
&= -\sum_{k=1}^{n} \nabla A_k(Z_t) J_{0,t}^{-1} V(Z_t) \circ \mathrm{d}W_t^{(k)} - \nabla A_0(Z_t) J_{0,t}^{-1} V(Z_t) \mathrm{d}t \\
&\quad + J_{0,t}^{-1} \nabla V(Z_t) \left[\sum_{k=1}^{n} A_k(Z_t) \circ \mathrm{d}W_t^{(k)} + A_0(Z_t) \mathrm{d}t \right] \\
&= \sum_{k=1}^{n} J_{0,t}^{-1} [A_k, V](Z_t) \circ \mathrm{d}W_t^{(k)} + J_{0,t}^{-1} [A_0, V](Z_t) \mathrm{d}t.
\end{aligned}$$

Converting this to Ito form gives

$$\mathrm{d} \left(J_{0,t}^{-1} V(Z_t) \right) = \sum_{k=1}^{n} J_{0,t}^{-1} [A_k, V](Z_t) \mathrm{d}W_t^{(k)} + J_{0,t}^{-1} \left(\frac{1}{2} \sum_{k=1}^{n} [A_k, [A_k, V]](Z_t) + [A_0, V](Z_t) \right) \mathrm{d}t.$$

We also need another important theorem from Malliavin calculus

Theorem 4.3 (Clark-Ocone Representation Formula)**.** *If F is Malliavin differentiable and $\mathbb{E}(F^2) < \infty, \mathbb{E}((\mathcal{D}_s F)^2) < \infty$ and W is a Brownian motion with natural filtration $\mathscr{F}_t$ then,*

$$F = \mathbb{E}(F) + \int_0^t \mathbb{E}(\mathcal{D}_s F | \mathscr{F}_s) \mathrm{d}W_s.$$

This could be considered a version of the fundamental theorem of calculus in this context. A proof of this can be found in [104].

Back to Assumption 3

Now we return to assumption 3. We are now in the setting of looking the the kinetic Fokker-Planck SDE

$$\mathrm{d}X_t = V_t \mathrm{d}t, \qquad \mathrm{d}V_t = -V_t \mathrm{d}t - \nabla_x U(X_t) \mathrm{d}t + \mathrm{d}W_t.$$

For this SDE we have that $n = 1$ and $A_1 = (0, 1)$ and $B = (v, -v - \nabla_x U(x))$. We define C_1 by

$$C_1 := [A_1, B](z) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

The key idea of this sections is that we can use Malliavin calculus to show that for very small t the solution behaves approximately like

$$\mathbb{E}(Z_t) + A_1 W_t + C_1 \int_0^t s \mathrm{d}W_s.$$

Then because $(W_t, \int_0^t s\mathrm{d}W_s)$ is a $2d$ dimensional non-degenerate Gaussian and because A_1 and C_1 are linearly independent this shows that the solution spreads out in every direction. In particular if we take two independent realisations Z_t^1 and Z_t^2 with different starting points the solutions will spread in the direction $\mathbb{E}(Z_t^1) - \mathbb{E}(Z_t^2)$ which allows us to show there is some positive probability of them becoming close.

Lemma 4.4. *Let U be smooth and satisfy $Hess(U) \leq M$ and fix δ and R. There exists $T = T(\delta, R)$ such that for fixed $0 < t < T$ there exists an $\alpha = \alpha(t, \delta, R)$ with the property that for any two independent solutions to the SDE, Z_t^1, Z_t^2 with initial points having $z_1, z_2 \in B(0, R)$, then*

$$\mathbb{P}(|Z_t^1 - Z_t^2| < \delta) \geq \alpha.$$

We have that

$$\alpha(t, \delta, R) = 1 - C\delta^2 \frac{1}{t_2} \exp\left(-\frac{k}{t^3} m^2\right) + 8 \exp\left(-\frac{\delta^2}{16Ct^5}\right).$$

Here k and m are explicit numerical constants. This value of $\alpha(t, \delta, R)$ is only positive for t sufficiently small and T is the value for which $\alpha(T, \delta, R) = 0$.

Proof. The key idea of this proof is to use the fact that the solution spreads out in every direction due to hypoelliptic effects. We represent the solution by a deterministic part, a Gaussian part and a small error. We begin by approximating the Malliavin derivative of the solution using the SDEs

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} J_{s,t} A_1 &= J_{s,t} C_1, \\
\frac{\mathrm{d}}{\mathrm{d}s} J_{s,t} C_1 &= J_{s,t} C_1 - U''(X_s) J_{s,t} A_1
\end{aligned}$$

We can then Taylor expand and use the Clarke-Ocone formula to get

$$Z_t = \mathbb{E}(Z_t) + \int_0^t \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} - (t-s) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + E_{s,t} \right) \mathrm{d}W_s.$$

$$E_{s,t} = -\mathbb{E}\left(\int_s^t (J_{r,t} C_1 - U''(X_r) J_{r,t} A_1)(t-r) \mathrm{d}r | \mathscr{F}_s \right)$$

At this point we have to assume that U'' is bounded in order to get bounds on $E_{s,t}$. Using the first part with the Lyapunov structure we know that $J_{0,t}$ can be bounded in terms of Lyapunov function we have

$$\|J_{s,t}\| \leq e^{1+M} \exp\left(\eta \int_s^t (X_r^2 + V_r^2 + U(X_r)) \mathrm{d}r\right) \leq e^{1+M} \exp\left(\eta \int_0^t (X_r^2 + V_r^2 + U(X_r)) \mathrm{d}r\right) \leq CL(Z_0).$$

Taking the supremum over possible starting points in $B(0, R)$ we have $E_{s,t} \leq C(t-s)^2$ for some constant C. Let us write

$$\mathcal{E}_t = \int_0^t E_{s,t} \mathrm{d}W_s.$$

We would like to get bounds on the expectation of $\exp c|\mathcal{E}_t|$. Since

$$\mathcal{E}_r = \int_0^r E_{s,t} \mathrm{d}W_s$$

is a Martingale for $r \leq t$ then by the exponential martingale inequality

$$\mathbb{E}\left(\exp\left(\xi \cdot \mathcal{E}_t\right)\right) \leq \exp\left(\int_0^t C|\xi|^2(t-s)^4 \mathrm{d}s\right) \leq \exp\left(C|\xi|^2 t^5\right).$$

Alternatively, we can bound $J_{s,t}$ in a way that doesn't depend on the initial data but does use that $Hess(U) \leq M$. We can use the equation to see that

$$|J_{s,t}A_1 + J_{s,t}C_1|^2 \leq 4e^{(2+M)t}.$$

Then the rest follows exactly as before but we replace C with $Ce^{(2+M)t}$. Since we are looking at the asymptotics for small t this makes no difference.

So we have decomposed Z_t into a deterministic part $\mathbb{E}(Z_t)$ a Gaussian part which we call G_t and an error which has exponential moments.

$$\begin{aligned}\mathbb{P}\left(Z_t^1 - Z_t^2 \notin B(0,\delta)\right) \leq & \mathbb{P}\left(\mathbb{E}(Z_t^1) - \mathbb{E}(Z_t^2) + G_t^1 - G_t^2 \notin B(z,\delta/2)\right) \\ & + \mathbb{P}\left(\mathcal{E}_t^1 \notin B(0,\delta/4)\right) + \mathbb{P}\left(\mathcal{E}^2 \notin B(0,\delta/4)\right).\end{aligned}$$

So we have by Markov's inequality

$$\mathbb{P}\left(\mathcal{E}_t \notin B(0,\delta/2)\right) \leq 4\exp\left(C\eta^2 t^5 - \eta\delta/2\right).$$

Optimising over η gives

$$\mathbb{P}\left(\mathcal{E}_t \notin B(0,\delta/2)\right) \leq 4\exp\left(-\frac{\delta^2}{16Ct^5}\right).$$

We can write down the density for $G_t^1 - G_t^2$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}(|Z_t^1 - Z_t^2|^2) \leq (2+M)\mathbb{E}(|Z_t^1 - Z_t^2|^2) + 4d.$$

This implies that

$$\mathbb{E}(|Z_t^1 - Z_t^2|^2) \leq e^{(2+M)t}(\mathbb{E}(|Z_0^1 - Z_0^2|^2) + 4d) \leq e^{(2+M)t}(R^2 + 4d).$$

We can therefore find the smallest that the density of $G_t^1 - G_t^2$ can be on a ball of size $\delta/2$ at the point $-\mathbb{E}(Z_t^1 - Z_t^2)$ when G^1 and G^2 are independent. Using this we make the two processes independent. The covariance matrix for G_t^1, G_t^2 has eigenvalues $(t/3 + o(t^3), t - t^2 + t/3 + o(t^3))$. Lets call $\sigma_m(t)$ the smallest eigenvalue and $\sigma_M(t)$ the largest eigenvalue. We also have that $z \leq L(z)/a_*$ so using Lemma 4.2 we have that $\mathbb{E}(Z_t^1 - Z_t^2) \leq 2C_* \max_{|z|\leq R}(L(z))/a_* =: m$ So we can bound the probability by

$$1 - \delta^2(2\pi\sigma_M(t))^{d/2}\exp\left(-\frac{m^2}{\sigma_m(t)}\right).$$

We then have that can approximate for $t \leq 1$,

$$\mathbb{P}\left(\mathbb{E}(Z_t^1 - Z_t^2) + G_t^1 - G_t^2 \notin B(0, \delta/2)\right) \leq 1 - C\delta^2 \frac{1}{t^2} \exp\left(-\frac{k}{t^3} m^2\right).$$

Here k and m are constants we can calculate explicitly. In total we have that

$$\mathbb{P}(Z_t^1 - Z_t^2 \notin B(0, \delta)) \leq 1 - C\delta^2 \frac{1}{t_2} \exp\left(-\frac{k}{t^3} M^2\right) + 8 \exp\left(-\frac{\delta^2}{16Ct^5}\right)$$

So as $t \to 0$ we can see that for a fixed sufficiently small t we have

$$\mathbb{P}(Z_t \in B(z, \delta)) \geq \alpha.$$

Where we can calculate α explicitly in terms of t, δ, R and the other constants appearing in the equation. □

Lemma 4.5. *Suppose we fix δ, t and R. Then there exists α such that for any two independent solutions to the SDEs Z_t^1, Z_t^2 with initial points having $z_1 - z_2 \in B(0, R)$ then*

$$\mathbb{P}(|Z_t^1 - Z_t^2| < \delta) \geq \alpha.$$

Furthermore if they start with initial points both in $B(0, R)$ then

$$\mathbb{P}(\rho_r(Z_t^1, Z_t^2) < \delta) \geq \alpha'.$$

Proof. We want to extend the previous Lemma to larger times by showing that if two solutions start with $z^1 - z^2 \in B(0, R)$ then they stay there with some positive probability. To do this we repeat the calculation but replacing δ by R then since the two processes are independent the probability that their difference stay inside $B(0, R)$ is given by the first lemma. So we have for some t_*

$$\mathbb{P}(Z_{t_*}^1 - Z_{t_*}^2 \in B(0, R) \mid Z_0^1 - Z_0^2 \in B(0, R)) \geq b$$

Therefore if $t = nt_* + s$ with $s \leq t_*$ then

$$\mathbb{P}(Z_t^1 - Z_t^2 \in B(0, \delta) \mid Z_t^1 - Z_t^2 \in B(0, R)) \geq ab^n.$$

Here a, b, t_* are explicitly calculable constants depending on M, F. However, we in fact need to look at ρ_r instead of the normal distance. In order to do this we need to look at

$$\mathbb{P}(Z_t^1 - Z_t^2 \in B(0, \delta), Z_t^1, Z_t^2 \in B(0, R')),$$

for some R'. We have that

$$\begin{aligned}\mathbb{P}(Z_t^1 - Z_t^2 \in B(0, \delta), Z_t^1, Z_t^2 \in B(0, R')) =& \mathbb{P}(Z_t^1, Z_t^2 \in B(0, R')) \\ &- \mathbb{P}(Z_t^1 - Z_t^2 \notin B(0, \delta), Z_t^1, Z_t^1 \in B(0, R'))\end{aligned}$$

So we bound

$$\begin{aligned}\mathbb{P}(Z_t^1 - Z_t^2 \notin B(0,\delta), Z_t^1, Z_t^2 \in B(0,R')) \leq & \mathbb{P}(Z_t^1, Z_t^2 \in B(0,R)) - \\ & \mathbb{P}(\mathbb{E}(Z_t^1 - Z_t^2) + G_t^1 - G_t^2 \in B(0,\delta/2), Z_t^1, Z_t^2 \in B(0,R')) \\ & + \mathbb{P}(\|E_t^1\| \leq \delta/4) + \mathbb{P}(\|E_t^2\| \leq \delta/4)\end{aligned}$$

Furthermore we have

$$\mathbb{P}(\mathbb{E}(Z_t^1 - Z_t^2) + G_t^1 - G_t^2 \in B(0,\delta/2), Z_t^1, Z_t^2 \in B(0,R')) \geq C\delta^2 R' \frac{1}{t^2} \exp\left(-K/t^3\right)$$

for explicitly computable constants C and K. So in the same way we have for all t, R, R' there is $a(t, R, R', \delta) > 0$ such that

$$\mathbb{P}(Z_t^1 - Z_t^2 \in B(0,\delta), Z_t^1, Z_t^2 \in B(0,R') \mid Z_0^1, Z_0^2 \in B(0,R)) \geq a(t, R, R', \delta).$$

Then we can find an R'' such that on any optimal path between two points in $B(0,R')$ we have $L(\gamma(t)) \leq R''$ so this implies that for $x, y \in B(0,R')$ we have

$$\rho_r(x,y) = \inf_\gamma \int_0^t L(\gamma(t))\dot{\gamma}(t)\mathrm{d}t \leq \inf_\gamma R'' \int_0^t \dot{\gamma}(t)\mathrm{d}t = R''|x-y|.$$

We mean that the two distances are equivalent on compact sets. So if $|x-y| \leq \delta/R''$ we have that $\rho_r(x,y) \leq \delta$ therefore

$$\mathbb{P}(\rho_r(Z_t^1, Z_t^2) \leq \delta \mid Z_0^1, Z_0^2 \in B(0,R)) \geq a(t, R, R', \delta/R'').$$

□

Now for this section we look again at how this shows contraction in the theorem of Hairer and Mattingly. We have that

Proposition 4.3. *If $\rho(z_1, z_2) \leq 4C_1$ and $\rho_r(z_1, z_2) > \delta$ then there exists γ such that*

$$\mathcal{W}_{1,d}(\mathcal{P}_t\delta_{z_1}, \mathcal{P}_t\delta_{z_2}) \leq \gamma d(z_1, z_2).$$

Proof. Suppose that we have that $\rho(z_1, z_2)_r \geq \delta$ and $\rho(z_1, z_2) \leq 4C_1$ then we have that z_1, z_2 are contained in some ball. There is some R such that for $|z| \geq R$ we have

$$L^*(z)^r \leq \frac{\delta}{8C_1} L_*(z).$$

Then as we discussed there is some R' such that

$$\int_R^{R'} L_*(r)\mathrm{d}r \geq 8C_1.$$

Therefore if $|z_1|, |z_2| \geq R'$ and $\rho(z_1, z_2) \leq 4C_1$ then if γ is a path such that

$$\int_0^1 L(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s \leq \rho(z_1, z_2) + \epsilon$$

then γ must not pass through $B(0,R)$. and for such a path

$$\rho_r(z_1,z_2) \leq \int_0^1 L^r(\gamma(s))|\dot{\gamma(s)}|\mathrm{d}s \leq \frac{\delta}{8C_1}\int_0^1 L(\gamma(s))|\dot{\gamma}(s)|\mathrm{d}s \leq \frac{\delta}{8C_1}(4C_1+\epsilon).$$

Since ϵ is arbitrary this shows that $\rho(z_1,z_2) \leq \delta$. Therefore if $\rho(z_1,z_2) \leq 4C_1$ and $\rho_r(z_1,z_2) \geq \delta$ we have that $z_1,z_2 \in B(0,R')$. Then for this R' we can apply lemma 4.5 to get that there is some a such that if we make Z^1,Z^2 independent then we have

$$\mathbb{P}(\rho_r(Z_t^1,Z_t^2) \leq \delta/2 \ | \ Z_0^1,Z_0^2 \in B(0,R')) \geq a.$$

Using this we have for the independent coupling

$$\mathbb{E}(d(Z_t^1,Z_t^2)) \leq \frac{1}{2}\mathbb{P}(\rho(Z_t^1,Z_t^2) \leq \delta/2) + (1-\mathbb{P}(\rho(Z_t^1,Z_t^2) \leq \delta/2)) + \beta\mathbb{E}(\rho(Z_t^1,Z_t^2))$$

$$\leq (1-a/2) + \beta(\mathbb{E}(\rho(0,Z_t^1)) + \mathbb{E}(\rho(0,Z_t^1))).$$

Now we can see that

$$\mathbb{E}(\rho(0,Z_t^1)) \leq \mathbb{E}(|Z_t^1|L^*(Z_t^1)) \leq CL^\kappa(z_1) \leq C_*$$

since $z_1 \in B(0,R')$. So if we take $\beta \leq a/8C_*$ then we have that

$$\mathbb{E}(d(Z_t^1,Z_t^2)) \leq 1-a/4 \leq (1-a/4)d(z_1,z_2).$$

$\square$

Chapter 5

Hypocoercivity in Φ-entropy for the linear relaxation Boltzmann equation

5.1 Introduction

In this chapter we constructively prove convergence to equilibrium for the linear relaxation Boltzmann equation on the torus in relative entropy. We also look at other entropy functionals, the p-entropies. The equation is

$$\partial_t f + v \cdot \nabla_x f = \lambda \tilde{\Pi}(f) - \lambda f. \tag{5.1}$$

Where $f = f(t,x,v) : \mathbb{R}_+ \times \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}$ and λ is a positive constant. We always consider f to be a probability density so it is positive and of mass one, this is well known to be preserved by the equation. The operator $\tilde{\Pi}$ is defined by

$$\tilde{\Pi}(f) = \left(\int_{\mathbb{R}^d} f(t,x,u)\mathrm{d}u \right) (2\pi)^{-d/2} \exp\left(-\frac{1}{2}|v|^2 \right) =: \left(\int_{\mathbb{R}^d} f(t,x,u)\mathrm{d}u \right) \mathcal{M}(v).$$

The equilibrium state of this equation is $\mu(x,v) = \mathcal{M}(v) \times 1$. We give two separate notations here to emphasize when we consider it as a function of v alone or a function of x and v. We will always work in terms of $h = f/\mu$ which satisfies,

$$\partial_t h + v \cdot \nabla_x h = \lambda \Pi h - \lambda h, \tag{5.2}$$

here we define Π by

$$\Pi h = \int_{\mathbb{R}^d} h(t,x,u)\mathcal{M}(u)\mathrm{d}u.$$

So the function Πh does not depend on v.

We want to study the convergence to equilibrium for solutions to this equations in relative entropy, H, and Fisher information, I, of f to μ. Studying the relative entropy has been an important way of showing convergence to equilibrium for kinetic equations since Boltzmann's H-theorem [23]. Fisher information was introduced into kinetic theory by McKean to study convergence to equilibrium for a caricature of the Boltzmann equation [95]. These quantities are defined in terms of $h = f/\mu$, and are

$$H(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} h\log(h)\mathrm{d}\mu,$$
$$I(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla h|^2}{h}\mathrm{d}\mu.$$

Villani and Desvillettes demonstrated convergence to equilibrium in weighted H^1 for spatially inhomogeneous kinetic equations including the Boltzmann equation in [46, 47], their techniques were also applied to the linear Boltzmann equation in [33] where they show convergence faster than any power of t. After this the theory of hypocoercivity was developed and the equation is shown to converge to equilibrium in weighted L^2 [77] by Hérau in order to demonstrate the applicability of the tools used in [80]. Convergence in weighted H^1 is also demonstrated in section 5.1 of [102] by Neumann and Mouhot as a consequence of a more general theorem. The techniques used in both these papers exploit commutator relations between the transport and collision part of the equation using the tools of hypocoercivity also see [54, 80, 116, 76, 97, 50]. The paper [2], shows convergence in Sobolev spaces with improved rates, and studies the convergence in relative entropy for models with discrete velocities. The convergence demonstrated in all these papers is of the form

$$H(f(t)|\mu) \leq Ce^{-\gamma t}H(f(0)|\mu),$$

where C and γ are explicit constants. If $C = 1$ the equation would be coercive in this norm. When $C > 1$, we use the terminology introduced in [116] and say that it is hypocoercive. We can see that our equation is hypocoercive not coercive as if it were coercive for all initial data that would be equivalent to the inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}H(f(t)|\mu) \leq -\gamma H(f(t)|\mu).$$

If we call the left hand side of this inequality the functional $-D(f(t)|\mu)$ then having this inequality for all initial data in some set $\mathcal{A}$ is equivalent to

$$D(f|\mu) \geq \gamma H(f|\mu) \quad \forall f \in \mathcal{A}.$$

We can check that this last inequality does not hold for the functionals we consider when f is in local equilibrium (i.e. of the form $\rho(x)\mathcal{M}(v)$). More precisely we can check that $D(\rho\mathcal{M}|\mu) = 0$.

Entropic hypocoercivity was introduced by Villani in [116]. More recently entropic hypocoercivity and hypocoercivity in different Φ entropies have been studied for diffusion operators [12, 98, 11] using Bakry-Emery type methods, in [19] for non-linear diffusions and in[7, 3] for linear or close to linear operators to find optimal rates. Working in relative entropy allows us to show convergence to equilibrium for a different class of initial data than if we were to use the results in Hilbert spaces. Another important advantage of working in entropy and Fisher information is that these distances behave well as the dimension of the space increases. The proofs also rely on logarithmic Sobolev inequalities where the constants do not depend on dimension. In part 1 section 6 of [116], Villani studies entropic hypocoercivity for derivative operators in a '$A^*A + B$' form. As in the Hilbert space theory this is done by constructing a 'twisted norm' which he then shows will converge to equilibrium. Here the role of the 'twisted norm' is taken by a distorted Fisher information like term

$$\int \frac{\nabla h \cdot S \nabla h}{h} \mathrm{d}\mu,$$

where S is a non-diagonal matrix. Crucially, as in many previous works we need to introduce a term with mixed derivatives. This term allows us to use the transport part of the equation to generate dissipation in the directions not dissipated by the collision operator.

The main purpose of this work is to demonstrate that entropic hypocoercivity can be proved for an equation which is not in '$A^*A + B$' form. The key difference between the proofs given here and those of previous hypocoercivity results arises because we do not have a diffusion operator. Therefore we cannot use the chain rule or understand the dissipation in terms of commutators or compositions of first order derivatives as is done in the first section of [116]. We find that these terms produce more extra terms which do not have an analogy in the Hilbert space case in the proof of Theorem 1.1 in [102]. Therefore, we need to add an extra entropy term to the functional. This term can be bounded above by $H(f)$ so we can still state our results in terms of the entropy and fisher information.

Theorem 5.1. *If f is a solution to (5.1) with initial data f_0 such that*

$$\int_{\mathbb{R}^d \times \mathbb{T}^d} \frac{|\nabla_{x,v} h_0|^2}{h_0} \mathrm{d}\mu < \infty, \qquad f_0 \in W^{1,1}(\mu),$$

then there exist constants $\Lambda > 0$ and $\alpha > 0, \beta > 0$, which we can calculate explicitly depending on λ but not on the dimension such that

$$I(h_t) + \beta H(\Pi h_t) \leq \exp(-\Lambda t) \left(\alpha I(h_0) + 2\beta H(\Pi h_0)\right).$$

This implies that for some γ,

$$H(h_t) \leq \exp(-\Lambda t) \left(\gamma I(h_0)\right).$$

We then look at the convergence to equilibrium in p-entropy, that is for $p \in (1, 2]$ we consider entropies of the form

$$H^{(p)}(h) = \int_{\mathbb{R}^d \times \mathbb{T}^d} \frac{h^p - h}{p(p-1)} \mathrm{d}\mu,$$

where h is as in the first section, and here the analogy of Fisher information is

$$I^{(p)}(h) = \int_{\mathbb{R}^d \times \mathbb{T}^d} h^{p-2} |\nabla_{x,v} h|^2 \mathrm{d}\mu.$$

These quantities interpolate between the Hilbert space case $p = 2$, and the Boltzmann entropy case, $p \sim 1$. They are used in [8, 20] to study Fokker-Planck equations and convergence to equilibrium. Here we have inequalities due to Beckner in [13] which play the same role as the logarithmic Sobolev inequality does in showing hypocoercivity in Boltzmann entropy. They are of the form

$$\int_{\mathbb{R}^d \times \mathbb{T}^d} \frac{h^p - h}{p(p-1)} \mathrm{d}\mu \leq C \int_{\mathbb{R}^d \times \mathbb{T}^d} h^{p-2} |\nabla_{x,v} h|^2 \mathrm{d}\mu.$$

These can be shown by interpolating between Poincaré and logarithmic Sobolev inequality [5]. Using this we can prove a similar theorem in the p-entropy case.

Theorem 5.2. *If f is a solution to (5.1) with initial data f_0 such that*

$$\int_{\mathbb{R}^d\times\mathbb{T}^d} h_0^{p-2}|\nabla h_0|^2 \mathrm{d}\mu < \infty, \qquad f_0 \in W^{1,1}(\mu),$$

then there exist constants $\Lambda > 0$ and $\alpha > 0, \beta > 0$, which we can calculate explicitly depending on λ, p, d such that

$$I^{(p)}(h_t) + \beta H^{(p)}(\Pi h_t) \leq \exp(-\Lambda t)\left(\alpha I^{(p)}(h_0) + \beta H^{(p)}(\Pi h_0)\right).$$

This implies that for some γ,

$$H^{(p)}(h_t) \leq \exp(-\Lambda t)\left(\gamma I_\mu^{(p)}(h_0)\right).$$

Remark. *For the case $p = 2$ we recover the result of section 5.1 in [102].*

Lastly, we briefly look at the kinetic Fokker-Planck equation and its convergence to equilibrium in p-entropy. It is already known that this is hypocoercive in H^1 and relative entropy see for example [80, 116]. We show here that, as in the linear relaxation Boltzmann case, we can extend this result to the p-entropies. The proof in p-entropies is very similar to that of these other results. The equation in terms of h is

$$\partial_t h + v\cdot\nabla_x h = (\nabla_v + v)\cdot\nabla_v h. \tag{5.3}$$

Theorem 5.3. *If f is a solution to (5.3), with finite initial Fisher information, then there exists an explicit constant k such that*

$$I^{(p)}(h_t) + \frac{27}{2}H^{(p)}(h_t) \leq e^{-kt}\left(3I^{(p)}(h_0) + \frac{27}{2}H^{(p)}(h_0)\right).$$

This implies for some C we have,

$$H^{(p)}(h_t) \leq e^{-kt}(CI^{(p)}(h_t)).$$

Here k does depend on p.

Remark. *We now briefly consider the case where $x \in \mathbb{R}^d$ and the transport operator also involves a confining potential term. For the kinetic Fokker-Planck equation Villani shows convergence in H^1 and Boltzmann entropy in the first section of [116]. In [98] Monmarché proves a general theorem which shows that hypocoercivity holds for the kinetic Fokker-Planck equation with confining potential in a class of Φ entropies which include the p-entropies. The proof here is different in strategy to the one given here or in [116] but very similar calculations to the ones used here in the proof of Theorem 3 can show hypocoercivity for the kinetic Fokker-Planck in the confining potential case. The situation is different for the linear relaxation Boltzmann equation. It is shown to be hypocoercive in L^2 in [77, 50]. To show hypocoercivity for the linear relaxation Boltzmann equation with a confining potential in Φ-entropies would involve a very different strategy to our proofs in this equation. However, in the near to quadratic case it is possible to exploit additional cancellations happening in the operator to show convergence as is shown in [99].*

5.2 Boltzmann entropy

Throughout the main parts of this paper we work with an h which is bounded above and below by constants and has bounded derivatives of all orders. In this set of possible h, all the integration by parts and differentiating through the integral are justified. In the appendix we show that these properties are propagated by the equation and that we can extend the result to a wider set using a density argument.

We now outline our strategy for the proof. Our goal is to get constructive rates of convergence to equilibrium by closing a Grönwall estimate on a functional that we construct. This functional is composed from the components of Fisher information and an entropy term. We introduce the components of Fisher information.

$$\begin{aligned} I^X &:= I^X(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu, \\ I^V &:= I^V(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu, \\ I^M &:= I^M(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu. \end{aligned}$$

We note here that I^M does not have a sign. We also introduce a projected entropy which we use in our functional,

$$H_\Pi(h) = \int_{\mathbb{T}^d} \Pi h \log(\Pi h)\mathrm{d}x.$$

We have several more terms which only appear in the intermediate steps of the proof,

$$\begin{aligned} I^{\Pi X/X} &:= \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x(\Pi h/h)|^2}{(\Pi h/h)} h\mathrm{d}\mu, \\ I^{\Pi X} &:= \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x(\Pi h)|^2}{(\Pi h)}\mathrm{d}\mu, \\ I^{\Pi V/V} &:= \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v(\Pi h/h)|^2}{(\Pi h/h)} h\mathrm{d}\mu = \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v h|^2}{h}\frac{\Pi h}{h}\mathrm{d}\mu. \end{aligned}$$

We prove later in this section that $I^X - I^{\Pi X} \geq 0$.

By differentiating along the flow of the equation we show that

$$\frac{\mathrm{d}}{\mathrm{d}t} I^X \leq -\lambda I^{\Pi X/X} - \lambda\left(I^X - I^{\Pi X}\right), \tag{5.4}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} I^M \leq -I^X - \lambda I^M + \frac{\lambda}{\epsilon} I^{\Pi X/X} + \lambda\epsilon I^{\Pi V/V}, \tag{5.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} I^V \leq -2I^M - \lambda I^V - \lambda I^{\Pi V/V}. \tag{5.6}$$

We begin by constructing a functional of the form

$$J = A_1 I^X + A_2 I^M + A_3 I^V,$$

with $A_1A_3 - A_2^2/4 \geq 0$. This inequality means that J is equivalent to the Fisher information I.

We now give a strategy for choosing the A_i. Whenever $A_1A_3 - A_2^2/4 \geq 0$ we can choose ϵ so that the sum of terms in the derivative of J which involve $I^{\Pi X/X}, I^{\Pi V/V}$ will be negative. We need that A_2 is non-zero since inequality 5.5 provides the negative I^X which we want in the derivative. The most natural next step would be to use the Cauchy-Schwarz inequality to control I^M by I^X, and I^V. However, we can check that the quantity of I^M is too large for this to be possible. We need to utilise inequality 5.4. We do this by showing that

$$-I^M \leq \frac{\eta}{2} I^V + \frac{1}{2\eta}\left(I^X - I^{\Pi X}\right) - \frac{\mathrm{d}}{\mathrm{d}t} H_\Pi. \tag{5.7}$$

This is the key new element in our proof.

By adding a quantity of H_Π to the functional and using inequality 5.7, we can now control I^M by I^V and $I^X - I^{\Pi X}$. Since the inequality 5.4 doesn't produce bad terms we are free to add as much I^X to the functional as we need. Therefore, by adding a large amount of H_Π and I^X to our functional we can cancel out the positive $I^X - I^{\Pi X}$. Therefore we can make η small. This means the sum of the positive I^V from controlling I^M and the negative I^V from inequality 5.6 will sum to a negative amount of I^V. We recall that we also have some negative I^X for inequality 5.5. So we have,

$$\frac{\mathrm{d}}{\mathrm{d}t}(J + A_4 H_\Pi) \leq -C(I^X + I^V).$$

We then use the equivalence between J and I and the logarithmic Sobolev inequality to get

$$\frac{\mathrm{d}}{\mathrm{d}t}(J + A_4 H_\Pi) \leq -C(J + A_4 H_\Pi).$$

So we can close a Gronwall estimate and then use the equivalence between J and I again to translate this to an inequality on I.

Before beginning the main proof it is helpful to separate out some lemmas. The first result relates the quantities involving only Πh to quantities coming from the full Fisher information. For this we define the local average speed $U(x)$, of a solution to (5.1) by

$$U(x) := \int_{\mathbb{R}^d} v h(v,x) \mathcal{M}(v) \mathrm{d}v = \int_{\mathbb{R}^d} v f(v,x) \mathrm{d}v.$$

Lemma 5.1. *For any h we have that*

$$I^{\Pi X}(h) = \int_{\mathbb{R}^d \times \mathbb{T}^d} \frac{|\nabla_x \Pi h|^2}{\Pi h} \mathrm{d}\mu \leq \int_{\mathbb{R}^d \times \mathbb{T}^d} \frac{|\nabla_x h|^2}{h} \mathrm{d}\mu.$$

This implies that for all h there exists a constant C such that

$$H_\Pi(h) = \int_{\mathbb{T}^d} \Pi h \log(\Pi h) \mathrm{d}x \leq C \int_{\mathbb{T}^d \times \mathbb{R}^d} \frac{|\nabla_x h|^2}{h} \mathrm{d}\mu.$$

Finally, if h is a solution to (5.2) then

$$\frac{\mathrm{d}}{\mathrm{d}t} H_\Pi(h(t)) = -\int_{\mathbb{T}^d} \log(\Pi h) \nabla_x \cdot U(x) \mathrm{d}x.$$

Proof. We can see that the first inequality will follow if

$$\frac{|\nabla_x \Pi h|^2}{\Pi h} \leq \Pi \left(\frac{|\nabla_x h|^2}{h} \right).$$

Since Π is integrating against a probability measure we would like to use Jensen's inequality. Instead of looking at h we consider $H = (\nabla_x h, h)$ and the function $\phi(\mathbf{x}, y) = |\mathbf{x}|^2/y$ which is convex so we have from Jensen's inequality that,

$$\phi(\Pi H) \leq \Pi(\phi(H)),$$

which implies our desired result since Π commutes with ∇_x. (Here Π acts component wise on vectors).

The second inequality follows since the log-sobolev inequality on the torus with uniform measure says that

$$\int_{\mathbb{T}^d} \Pi h \log(\Pi h) \mathrm{d}x \leq C \int_{\mathbb{T}^d} \frac{|\nabla_x \Pi h|^2}{\Pi h} \mathrm{d}x.$$

We then use the first inequality to get the final result.

For the last part,

$$\begin{aligned} \partial_t \Pi h = & - \int_{\mathbb{R}^d} v \cdot \nabla_x h \mathcal{M}(v) \mathrm{d}v + \lambda \Pi(\Pi h) - \lambda \Pi h \\ = & - \nabla_x \cdot U(x). \end{aligned}$$

Since,

$$\int_{\mathbb{T}^d} \partial_t \Pi h \mathrm{d}x = \partial_t \int_{\mathbb{R}^d \times \mathbb{T}^d} h \mathrm{d}\mu = 0,$$

we have that

$$\partial_t H_\Pi(h) = \int_{\mathbb{T}^d} (\partial_t \Pi h) \log(\Pi h) \mathrm{d}x = - \int_{\mathbb{T}^d} \log(\Pi h) \nabla_x \cdot U(x) \mathrm{d}x.$$

□

We now prove inequalities 5.4, 5.5, 5.6. First, for simplicity, we introduce some notation. Let $T = -v \cdot \nabla_x$ and $L = \lambda(\Pi - I)$ then let $(\mathrm{d}/\mathrm{d}t)_O$ represent the derivative along the semi-group generated by any given operator O.

Lemma 5.2. *We have the following inequalities or inequalities on the derivatives of the components* *of Fisher information,*

$$
\begin{aligned}
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^X =&0,\\
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^V =&2I^M,\\
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^M =&I^X,\\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^V =&-\lambda\left(I^V + I^{\Pi V/V}\right),\\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X \leq&-\lambda I^{\Pi X/X} - \lambda\left(I^X - I^{\Pi X}\right),\\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^M \leq&\lambda\epsilon I^{\Pi V/V} + \frac{\lambda}{\epsilon} I^{\Pi X/X} - \lambda I^M.
\end{aligned}
$$

Here ϵ is any strictly positive number.

Remark. *The first three equalities simply follow in a very similar way to [102]. When we are differentiating under the flow of L we begin to see terms appearing which correspond to the relative Fisher information of h to Πh or vice versa. These terms are similar in spirit to terms like $\|\nabla_x(h-\Pi h)\|$ appearing in [102].*

Proof. The first three equalities are straightforward calculations. They are also given in a different form in [116].

$$
\begin{aligned}
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu =&2\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x v\cdot\nabla_x h\cdot\nabla_x h}{h}\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\frac{v\cdot\nabla_x h}{h}\mathrm{d}\mu\\
=&\int_{\mathbb{R}^d\times\mathbb{T}^d} v\cdot\nabla_x\left(\frac{|\nabla_x h|^2}{h}\right)\mathrm{d}\mu\\
=&0.
\end{aligned}
$$

$$
\begin{aligned}
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu =&2\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_v v\cdot\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v h|^2}{h}\frac{v\cdot\nabla_x h}{h}\mathrm{d}\mu\\
=&2\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu + \int_{\mathbb{R}^d\times\mathbb{T}^d} v\cdot\nabla_x\left(\frac{|\nabla_v h|^2}{h}\right)\mathrm{d}\mu\\
=&2\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu.
\end{aligned}
$$

$$
\begin{aligned}
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu =& \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_v v\cdot\nabla_x h\cdot\nabla_x h}{h}\mathrm{d}\mu + \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_v h\cdot\nabla_x v\cdot\nabla_x h}{h}\mathrm{d}\mu\\
&- \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\frac{v\cdot\nabla_x h}{h}\mathrm{d}\mu\\
=& \int_{\mathbb{R}^d\times\mathbb{T}^d} v\cdot\nabla_x\left(\frac{\nabla_x h\cdot\nabla_v h}{h}\right)\mathrm{d}\mu + \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu\\
=& \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu.
\end{aligned}
$$

For the last three terms we have that,

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu =& 2\lambda \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_v(\Pi - I)h\cdot\nabla_v h}{h}\mathrm{d}\mu - \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_v h|^2}{h}\frac{(\Pi-I)h}{h}\mathrm{d}\mu \\ =& -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_v h|^2}{h}\frac{(I+\Pi)h}{h}\mathrm{d}\mu.\end{aligned}$$

where this last result follows from the fact that $\nabla_v \Pi h = 0$.

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu =& 2\lambda \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x(\Pi - I)h\cdot\nabla_x h}{h}\mathrm{d}\mu - \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x h|^2}{h}\frac{(\Pi-I)h}{h}\mathrm{d}\mu \\ =& -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x h|^2}{h}\mathrm{d}\mu + 2\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x\Pi h\cdot\nabla_x h}{h}\mathrm{d}\mu \\ & -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x h|^2}{h}\frac{\Pi h}{h}\mathrm{d}\mu \\ =& \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x \Pi h|^2}{\Pi h}\mathrm{d}\mu - \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x h|^2}{h}\mathrm{d}\mu \\ & -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\left|\frac{\nabla_x\Pi h}{\Pi h} - \frac{\nabla_x h}{h}\right|^2\Pi h\mathrm{d}\mu \\ =& \lambda\left(\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x \Pi h|^2}{\Pi h}\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x h|^2}{h}\mathrm{d}\mu\right) \\ & -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x(\Pi h/h)|^2}{(\Pi h/h)}h\mathrm{d}\mu \\ \leq & -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x(\Pi h/h)|^2}{(\Pi h/h)}h\mathrm{d}\mu.\end{aligned}$$

Here the last inequality follows from the first inequality in Lemma 5.1.

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu =& \lambda \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x(\Pi - I)h\cdot\nabla_v h}{h}\mathrm{d}\mu + \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x h\cdot\nabla_v(\Pi-I)h}{h}\mathrm{d}\mu \\ & -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x h\cdot\nabla_v h}{h}\frac{(\Pi-I)h}{h}\mathrm{d}\mu \\ =& -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x h\cdot\nabla_v h}{h}\frac{(I+\Pi)h}{h}\mathrm{d}\mu + \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x\Pi h\cdot\nabla_v h}{h}\mathrm{d}\mu \\ =& \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\nabla_x\left(\frac{\Pi h}{h}\right)\cdot\nabla_v h\mathrm{d}\mu - \lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu \\ \leq & \lambda\epsilon\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_v h|^2}{h}\frac{\Pi h}{h}\mathrm{d}\mu + \lambda\frac{1}{\epsilon}\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|\nabla_x(\Pi h/h)|^2}{(\Pi h/h)}h\mathrm{d}\mu \\ & -\lambda\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu.\end{aligned}$$

We can check that all the terms appearing on the right hand side can be bounded in terms of

$$\int_{\mathbb{R}^d\times\mathbb{T}^d}\left(1+\frac{\Pi h}{h}\right)\frac{|\nabla h|^2}{h}\mathrm{d}\mu,$$

which justifies switching integration and differentiation. □

It might appear at this point that if we were to split $I^M \leq I^X/2\eta + \eta I^V/2$ then we would be able to close a Gronwall type estimate but we cannot close an estimate doing this. We can see that unlike in [102] we have not bounded $(\mathrm{d}/\mathrm{d}t)_L I^M$ by terms only involving v-derivatives and the distance between h and Πh with x-derivatives, this produces some extra mixed term in the derivative which has to be dealt with.

Lemma 5.3. *For any positive η we have*

$$-I^M \leq \frac{\eta}{2} I^V + \frac{1}{2\eta}(I^X - I^{\Pi X}) - \frac{\mathrm{d}}{\mathrm{d}t} H_\Pi.$$

Proof. First we notice that

$$\begin{aligned}
\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{h}{\Pi h}\frac{|\nabla_x(\Pi h/h)|^2}{(\Pi h/h)} h\mathrm{d}\mu &= \int_{\mathbb{R}^d\times\mathbb{T}^d} |\nabla_x \log(\Pi h/h)|^2 h\mathrm{d}\mu \\
&= \int_{\mathbb{R}^d\times\mathbb{T}^d} |\nabla_x \log(\Pi h) - \nabla_x \log h|^2 h\mathrm{d}\mu \\
&= \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x \Pi h|^2}{(\Pi h)^2} h\mathrm{d}\mu \\
&\quad - 2\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x \Pi h\cdot\nabla_x h}{\Pi h}\mathrm{d}\mu \\
&\quad + \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu \\
&= \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x h|^2}{h}\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x \Pi h|^2}{\Pi h}\mathrm{d}\mu,
\end{aligned}$$

where here we push the integration in v onto either h or $\nabla_x h$ and use the fact that Π commutes with ∇_x. Now we can see that

$$\begin{aligned}
-\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu =& -\int_{\mathbb{R}^d\times\mathbb{T}^d} \nabla_x \log h\cdot\nabla_v h\mathrm{d}\mu + \int_{\mathbb{R}^d\times\mathbb{T}^d} \nabla_x \log(\Pi h)\cdot\nabla_v h\mathrm{d}\mu \\
& - \int_{\mathbb{R}^d\times\mathbb{T}^d} \nabla_x \log(\Pi h)\cdot\nabla_v h\mathrm{d}\mu \\
=& \int_{\mathbb{R}^d\times\mathbb{T}^d} \nabla_x \log(\Pi h/h)\cdot\nabla_v h\mathrm{d}\mu - \int_{\mathbb{T}^d} \nabla_x \log(\Pi h)\cdot\int_{\mathbb{R}^d}\nabla_v h\mathcal{M}(v)\mathrm{d}v\mathrm{d}x \\
=& \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\nabla_x(\Pi h/h)\cdot\nabla_v h}{(\Pi h/h)}\mathrm{d}\mu + \int_{\mathbb{T}^d} \log(\Pi h)\nabla_x\cdot\left(\int_{\mathbb{R}^d} hv\mathcal{M}(v)\mathrm{d}v\right)\mathrm{d}x \\
\leq& \frac{\eta}{2}\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_v h|^2}{h}\mathrm{d}\mu + \frac{1}{2\eta}\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{h}{\Pi h}\frac{|\nabla_x(\Pi h/h)|^2}{(\Pi h/h)}h\mathrm{d}\mu \\
& - \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{\mathbb{T}^d} \Pi h\log(\Pi h)\mathrm{d}x\right) \\
=& \frac{\eta}{2} I^V + \frac{1}{2\eta}\left(I^X - I^{\Pi X}\right) - \frac{\mathrm{d}}{\mathrm{d}t}H_\Pi.
\end{aligned}$$

□

We are now able to prove the main theorem

Proof of Theorem 5.1. We look at

$$J(h) = A_1 I^X + A_2 I^M + A_3 I^V.$$

Using Lemma 5.2 we can see that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}J_\mu(f) =& -\lambda\left(A_1 - \frac{A_2}{\epsilon}\right) I^{\Pi X/X} \\
&- A_1\lambda(I^X - I^{\Pi X}) \\
&- \lambda\left(A_3 - \epsilon A_2\right) I^{\Pi V/V} \\
&- A_2 I^X - \lambda A_3 I^V - (\lambda A_2 + 2A_3) I^M.
\end{aligned}$$

We therefore have for any $\epsilon > 0, \eta > 0$ and $\alpha \in (0,1)$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(J + (\lambda A_2 + 2A_3)H_\Pi\right) \leq& -\left(\lambda A_1 - \frac{\lambda A_2}{\epsilon}\right) I^{\Pi X/X} \\
&- \left(\lambda A_1 - \frac{1}{2\eta}(\lambda A_2 + 2A_3)\right)(I^X - I^{\Pi X}) \\
&- \lambda(A_3 - \epsilon A_2) I^{\Pi V/V} \\
&- \left(\lambda A_3 - \frac{\eta}{2}(\lambda A_2 + 2A_3)\right) I^V \\
&- \alpha A_2 I^X \\
&- (1-\alpha) A_2 I^X \\
\leq& -\left(\lambda A_1 - \frac{\lambda A_2}{\epsilon}\right) I^{\Pi X/X} \\
&- \left(\lambda A_1 - \frac{1}{2\eta}(\lambda A_2 + 2A_3)\right)(I^X - I^{\Pi X}) \\
&- \lambda(A_3 - \epsilon A_2) I^{\Pi V/V} \\
&- \left(\lambda A_3 - \frac{\eta}{2}(\lambda A_2 + 2A_3)\right) I^V \\
&- \alpha A_2 I^X \\
&- (1-\alpha) A_2 C H_\Pi.
\end{aligned}$$

We set,

$$\begin{aligned}
A_1 =& \frac{1}{\eta}\left(\lambda + \frac{2}{\lambda}\right) A_3, \\
A_2 =& \lambda A_3, \\
\epsilon =& \frac{1}{\lambda}, \\
\alpha =& \frac{1}{2}.
\end{aligned}$$

Since if we send η to 0 then A_1 will become much larger than A_2 and A_3 we can choose η sufficiently small so that

$$\frac{1}{2}A_3 I(h) \leq J(h) \leq 2A_1 I(h).$$

Then the inequalities we have to satisfy become

$$\begin{aligned}\frac{1}{\eta}\left(\lambda^2+2\right)A_3-\lambda^3 A_3\geq&0,\\ \frac{1}{2\eta}\left(\lambda+\frac{2}{\lambda}\right)A_3\geq&0,\\ \left(\lambda-\frac{\eta}{2}\left(\lambda^2+2\right)\right)A_3\geq&\frac{\lambda}{2}A_3.\end{aligned}$$

So we can choose η sufficiently small so that all these inequalities are satisfied and we get that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(J(h)+(\lambda^2+2)A_3H_\Pi\right)\leq-\frac{\lambda A_3}{2}I(h)-\frac{\lambda A_3}{2}CH_\Pi.$$

Which becomes,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(J(h)+(\lambda^2+2)A_3H_\Pi\right)\leq-\frac{\lambda^2\eta}{4(\lambda^2+2)}J(h)-\frac{C\lambda}{2(\lambda^2+2)}\left((\lambda^2+2)A_3H_\Pi\right).$$

Since, η is small (or if not we can make η even smaller), and $H_\Pi\geq 0$, the first term will dominate so we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(J(h)+(\lambda^2+2)A_3H_\Pi\right)\leq-\frac{\lambda^2\eta}{4(\lambda^2+2)}\left(J(h)+(\lambda^2+2)A_3H_\Pi\right).$$

So if we now convert to I we have

$$\begin{aligned}I(h)+2(\lambda^2+2)H_\Pi\leq&\frac{2}{A_3}\left(J(h)+(\lambda^2+2)A_3H_\Pi\right)\\ \leq&\frac{2}{A_3}\exp\left(-\frac{\lambda^2\eta t}{4(\lambda^2+2)}\right)\left(J(h_0)+(\lambda^2+2)A_3H_\Pi(h_0)\right)\\ \leq&\exp\left(-\frac{\lambda^2\eta t}{4(\lambda^2+2)}\right)\left(\frac{4}{\eta\lambda}(\lambda^2+2)I(h_0)+2(\lambda^2+2)H_\Pi(h_0)\right)\end{aligned}$$

Since $H(h)\leq I(h)$, and $H_\Pi(h)\leq H(h)$ we can write this as in the theorem. □

5.3 p-entropies

In this section we will prove Theorem 5.2. Here for compactness of notation we suppress the p in the notation for the entropy functional. The proof is similar to the proof of Theorem 5.1 except that it is useful to understand the dissipation of I^X in a different way. Also, a number of extra terms appear which can be shown to have a good sign and are therefore easy to deal with. We can justify switching the order of integration and differentiation as in section 2. We recall that $p\in(1,2]$. We again make some notation

$$I^X=\int_{\mathbb{R}^d\times\mathbb{T}^d}h^{2-p}|\nabla_x h|^2\mathrm{d}\mu,\ \ I^V=\int_{\mathbb{R}^d\times\mathbb{T}^d}h^{2-p}|\nabla_v h|^2\mathrm{d}\mu,\ \ I^M=\int_{\mathbb{R}^d\times\mathbb{T}^d}h^{2-p}\nabla_x h\cdot\nabla_v h\mathrm{d}\mu,$$

and

$$I^{\Pi X}\int_{\mathbb{R}^d\times\mathbb{T}^d}\Pi h^{2-p}|\nabla_x\Pi h|^2\mathrm{d}\mu.$$

Our first lemma is Lemma 5.1 in the p-entropy setting.

Lemma 5.4. *For any h we have that*

$$I_\Pi(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} (\Pi h)^{p-2}|\nabla_x \Pi h|^2 \mathrm{d}\mu \leq \int_{\mathbb{R}^d\times\mathbb{T}^d} h^{p-2}|\nabla_x h|^2 \mathrm{d}\mu.$$

This implies that for all h there exists a constant C such that

$$H_\Pi(h) = \int_{\mathbb{T}^d} \frac{(\Pi h)^p - \Pi h}{p(p-1)} \mathrm{d}x \leq B \int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}|\nabla_x h|^2 \mathrm{d}\mu.$$

Finally, if h is a solution to (5.2) then

$$\frac{\mathrm{d}}{\mathrm{d}t} H_\Pi(h(t)) = -\int_{\mathbb{T}^d} \frac{(\Pi h)^{p-1}}{p-1} \nabla_x \cdot U(x) \mathrm{d}x.$$

Proof. The proof is similar to that of Lemma 5.1 we first note that $|x|^2 y^{p-2}$ is a convex function of x and y for $x \in \mathbb{R}^d, y \in (0,\infty)$. As the sum of convex functions is a convex function, we can particularize to the case $d = 1$ where the Hessian matrix is

$$\begin{pmatrix} 2y^{p-2} & 2(p-2)xy^{p-3} \\ 2(p-2)xy^{p-3} & x^2(p-2)(p-3)y^{p-4} \end{pmatrix}.$$

The trace is $2y^{p-2}+x^2(p-2)(p-3)y^{p-4} \geq 0$, and the determinant is $-2(p-2)(p-1)y^{(2p-6)}x^2 \geq 0$. So the Hessian is positive definite. Therefore by the same argument as before we have that

$$I_\Pi(h) \leq I^X(h).$$

Now by the Beckner inequalities for the torus, [5], we have that

$$H_\Pi(h) \leq C I_\Pi(h),$$

from which the second inequality follows. Lastly we know that

$$\partial_t \Pi h = -\nabla_x \cdot U(x),$$

so

$$\frac{\mathrm{d}}{\mathrm{d}t} H_\Pi(h) = -\int_{\mathbb{T}^d} \frac{(\Pi h)^{p-1}}{p-1} \nabla_x \cdot U(x) \mathrm{d}x.$$

□

We can also prove a Lemma which is very similar to Lemma 5.3

Lemma 5.5. *For any positive η we have*

$$-I^M \leq \frac{\eta}{2} I^V + \frac{1}{2\eta}(I^X - I^{\Pi X}) - \frac{\mathrm{d}}{\mathrm{d}t} H_\Pi.$$

Proof.

$$\begin{aligned}
-\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x h\cdot\nabla_v h \mathrm{d}\mu =& -\int_{\mathbb{T}^d\times\mathbb{R}^d}\nabla_x\left(\frac{h^{p-1}}{p-1}\right)\cdot\nabla_v h\mathrm{d}\mu\\
=& -\int_{\mathbb{T}^d\times\mathbb{R}^d}\nabla_x\left(\frac{h^{p-1}-(\Pi h)^{p-1}}{p-1}\right)\cdot\nabla_v h\mathrm{d}\mu\\
&-\int_{\mathbb{T}^d\times\mathbb{R}^d}\nabla_x\left(\frac{(\Pi h)^{p-1}}{p-1}\right)\cdot\nabla_v h\mathrm{d}\mu\\
\le&\frac{\eta}{2}I^V+\frac{1}{2\eta}\int\frac{\left|\nabla_x\left(\frac{(\Pi h)^{p-1}-h^{p-1}}{p-1}\right)\right|^2}{h^{p-2}}\mathrm{d}\mu\\
&+\int_{\mathbb{T}^d\times\mathbb{R}^d}\frac{(\Pi h)^{p-1}}{p-1}\nabla_x\cdot U(x)\mathrm{d}\mu\\
\le&\frac{\eta}{2}I^V+\frac{1}{2\eta}(I^X-I^{\Pi X})-\frac{\mathrm{d}}{\mathrm{d}t}H_\Pi.
\end{aligned}$$

Here for the last inequality we expand out and use that $\Pi\nabla_x h=\nabla_x\Pi h$ and $\Pi(h^{2-p})\le(\Pi h)^{2-p}$ since x^{2-p} is a concave function.

$$\begin{aligned}
&\int\left|\nabla_x\left(\frac{(\Pi h)^{p-1}-h^{p-1}}{p-1}\right)\right|^2 h^{2-p}=\\
&\int_{\mathbb{R}^d\times\mathbb{T}^d}\left(|\nabla_x\Pi h|^2\Pi h^{2p-4}h^{2-p}-2\nabla_x\Pi h\cdot\nabla_x h(\Pi h)^{p-2}+|\nabla_x h|^2h^{p-2}\right)\mathrm{d}\mu\\
\le&\int_{\mathbb{R}^d\times\mathbb{T}^d}\left(|\nabla_x\Pi h|^2(\Pi h)^{p-2}-2|\nabla_x\Pi h|^2(\Pi h)^{p-2}+|\nabla_x h|^2h^{p-2}\right)\mathrm{d}\mu\\
\le& I^X-I^{\Pi X}.
\end{aligned}$$

□

Before we continue the proof as before we need to be able to deal with another term which appears for the p-entropies but not elsewhere.

Lemma 5.6. *The function $F_p(r)$ defined for $r\in[0,\infty)$ by*

$$F_p(r)=p-1+(2-p)r-r^{2-p},$$

is positive when r is positive. Note that F_p is zero whenever p is 1 or 2.

Proof. $F_p(1)=0, F_p'(1)=0, F_p''(r)=-(2-p)(1-p)r^{-p}\ge 0$. Therefore, by Taylor's theorem we have $F_p(r)=-(2-p)(1-p)s^{-p}(r-1)^2$ for some s between r and 1. So $F_p(r)\ge 0$. □

We now calculate the derivatives of different components of $I(f)$. In order that we can write things compactly we introduce the following extra notation

$$I^{X,F}=\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}|\nabla_x h|^2F_p(\Pi h/h)\mathrm{d}\mu,\ \ I^{V,F}=\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}|\nabla_v h|^2F_p(\Pi h/h)\mathrm{d}\mu,$$

$$I^{V\Pi}=\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}|\nabla_v h|^2(\Pi h/h)^{2-p}\mathrm{d}\mu.$$

Lemma 5.7.

$$
\begin{aligned}
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^X =&0,\\
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^V =&2I^M,\\
-\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^M =&I^X,\\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^V =&-\lambda\left(I^{V,F}+I^{V,\Pi}+I^V\right),\\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X \leq&-\lambda D-\lambda I^{X,F},\\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^M \leq&\frac{\lambda\epsilon_1}{2}I^{V,\Pi}+\frac{\lambda}{\epsilon_1}D+\frac{\lambda}{2\epsilon_2}I^{X,F}+\frac{\lambda\epsilon_2}{2}I^{V,F}-\lambda I^M.
\end{aligned}
$$

Here ϵ_1,ϵ_2 are any strictly positive real numbers and,

$$
D=\int_{\mathbb{T}^d\times\mathbb{R}^d}\frac{\left|\nabla_x\left(\frac{(\Pi h)^{p-1}-h^{p-1}}{p-1}\right)\right|^2}{(\Pi h)^{p-2}}\mathrm{d}\mu.
$$

Proof. The calculation for the first three terms is identical so we do not repeat it. We first expand out D to make it easier to recognise in the calculations later.

$$
\begin{aligned}
D=&\int_{\mathbb{T}^d\times\mathbb{R}^d}\left|(\Pi h)^{p-2}\nabla_x\Pi h-h^{p-2}\nabla_x h\right|^2(\Pi h)^{2-p}\mathrm{d}\mu\\
=&\int_{\mathbb{T}^d\times\mathbb{R}^d}\left((\Pi h)^{p-2}|\nabla_x\Pi h|^2-2h^{p-2}\nabla_x\Pi h\cdot\nabla_x h+(\Pi h)^{2-p}h^{2p-4}|\nabla_x h|^2\right)\mathrm{d}\mu
\end{aligned}
$$

Now we can proceed to calculate the derivatives.

$$
\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X =&-p\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}|\nabla_x h|^2\mathrm{d}\mu+2\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}\nabla_x h\cdot\nabla_x\Pi h\mathrm{d}\mu\\
&-(2-p)\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}\frac{\Pi h}{h}|\nabla_x h|^2\mathrm{d}\mu\\
=&-\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}(\Pi h)^{2-p}|\nabla_x h|^2h^{2p-4}\mathrm{d}\mu+2\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}\nabla_x h\cdot\nabla_x\Pi h\mathrm{d}\mu\\
&-\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}(\Pi h)^{p-2}|\nabla_x\Pi h|^2\mathrm{d}\mu\\
&-p\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}|\nabla_x h|^2\mathrm{d}\mu-(2-p)\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}\frac{\Pi h}{h}|\nabla_x h|^2\mathrm{d}\mu\\
&+\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}(\Pi h)^{2-p}|\nabla_x h|^2h^{2p-4}\mathrm{d}\mu+\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}(\Pi h)^{p-2}|\nabla_x\Pi h|^2\mathrm{d}\mu\\
\leq&-\lambda D+(1-p)\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}h^{p-2}|\nabla_x h|^2\mathrm{d}\mu-(2-p)\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\frac{\Pi h}{h}h^{p-2}|\nabla_x h|^2\mathrm{d}\mu\\
&+\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\left(\frac{\Pi h}{h}\right)^{2-p}h^{p-2}|\nabla_x h|^2\mathrm{d}\mu-\lambda\left(I^X-I^{\Pi X}\right)\\
=&-\lambda D-\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}F\left(\frac{\Pi h}{h}\right)h^{p-2}|\nabla_x h|^2\mathrm{d}\mu-\lambda\left(I^X-I^{\Pi X}\right).
\end{aligned}
$$

Using very similar calculations for the v-derivatives we have,

$$
\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^V =& -p\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}|\nabla_v h|^2\mathrm{d}\mu - (2-p)\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\frac{\Pi h}{h}|\nabla_v h|^2\mathrm{d}\mu\\
=& -\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\left(\frac{\Pi h}{h}\right)^{2-p} h^{p-2}|\nabla_v h|^2\mathrm{d}\mu - \lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} F\left(\frac{\Pi h}{h}\right)h^{p-2}|\nabla_v h|^2\mathrm{d}\mu\\
& -\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}|\nabla_v h|^2\mathrm{d}\mu.
\end{aligned}
$$

$$
\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^{X,V} =& -p\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x h\cdot\nabla_v h\mathrm{d}\mu - (2-p)\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\frac{\Pi h}{h}h^{p-2}\nabla_x h\cdot\nabla_v h\mathrm{d}\mu\\
&+\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x\Pi h\cdot\nabla_v h\mathrm{d}\mu\\
=& -p\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x h\cdot\nabla_v h\mathrm{d}\mu - (2-p)\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\frac{\Pi h}{h}h^{p-2}\nabla_x h\cdot\nabla_v h\mathrm{d}\mu\\
&+\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\left(\frac{\Pi h}{h}\right)^{2-p}\nabla_x\left(\frac{(\Pi h)^{p-1}}{p-1}\right)\cdot\nabla_v h\mathrm{d}\mu\\
=&\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d}\left(\frac{\Pi h}{h}\right)^{2-p}\left(\nabla_x\left(\frac{(\Pi h)^{p-1}}{p-1}-\frac{h^{p-1}}{p-1}\right)\right)\cdot\nabla_v h\mathrm{d}\mu\\
&+\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x h\cdot\nabla_v hF\left(\frac{\Pi h}{h}\right)\mathrm{d}\mu\\
&-\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x h\cdot\nabla_v h\mathrm{d}\mu\\
\leq&\frac{\lambda}{2\epsilon_1}D + \frac{\lambda\epsilon_1}{2}\int_{\mathbb{T}^d\times\mathbb{R}^d}\left(\frac{\Pi h}{h}\right)^{2-p} h^{p-2}|\nabla_v h|^2\mathrm{d}\mu\\
&+\frac{\lambda}{2\epsilon_2}\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}|\nabla_x h|^2 F\left(\frac{\Pi h}{h}\right)\mathrm{d}\mu + \frac{\lambda\epsilon_2}{2}\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}|\nabla_v h|^2 F\left(\frac{\Pi h}{h}\right)\mathrm{d}\mu\\
&-\lambda\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x\cdot\nabla_v h\mathrm{d}\mu.
\end{aligned}
$$

□

Proof of Theorem 5.2. As before, first we consider an entropy of the form

$$J(h) = A_1 I^X + A_2 I^M + A_3 I^V.$$

So by Lemma 5.7 we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}J \leq& -\lambda\left(A_1 - \frac{A_2}{2\epsilon_1}\right)D - \lambda A_1\left(I^X - I^{\Pi X}\right)\\
&-\lambda\left(A_1 - \frac{A_2}{2\epsilon_2}\right)I^{X,F} - A_2 I_X\\
&-\lambda\left(A_3 - \frac{A_2\epsilon_1}{2}\right)I^{V,\Pi} - \lambda\left(A_3 - \frac{A_2\epsilon_2}{2}\right)I^{V,F}\\
&-\lambda A_3 I^V - (\lambda A_2 + 2A_3)\,I^M.
\end{aligned}
$$

We can now use Lemma 5.5 to see

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}J =& -\lambda\left(A_1 - \frac{A_2}{2\epsilon_1}+\right)D - \left(\lambda A_1 - \frac{\lambda A_2 + 2A_3}{2\eta}\right)(I^X - I^{\Pi X})\\ &-\lambda\left(A_1 - \frac{A_2}{2\epsilon_2}\right)I^{X,F} - A_2 I_X\\ &-\lambda\left(A_3 - \frac{A_2\epsilon_1}{2}\right)I^{V,\Pi} - \lambda\left(A_3 - \frac{A_2\epsilon_2}{2}\right)I^{V,F}\\ &-\left(\lambda A_3 - \frac{\eta}{2}(\lambda A_2 + 2A_3)\right)I^V - (\lambda A_2 + 2A_3)\frac{\mathrm{d}}{\mathrm{d}t}H_\Pi.\end{aligned}$$

We set $A_2 = \lambda A_3, \epsilon = 2/\lambda, A_1 = (\lambda + 2/\lambda)/\eta$ to get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}J \leq& -\lambda\left(\frac{1}{\eta}(\lambda + 2/\lambda) - \lambda^2/4\right)A_3 D\\ &-\frac{1}{2\eta}(\lambda^2+2)A_3(I^X - I^{\Pi X})\\ &-\lambda\left(\frac{1}{\eta}(\lambda + 2/\lambda) - \lambda^2/4\right)A_3 I^{X,F}\\ &-\lambda A_3 I^X\\ &-\left(\lambda - \frac{\eta}{2}(\lambda^2+2)\right)A_3 I^V\\ &-(\lambda^2+2)A_3\frac{\mathrm{d}}{\mathrm{d}t}H_\Pi.\end{aligned}$$

Making η small enough we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(J + (\lambda^2+2)A_3 H_\Pi) \leq -\frac{\lambda}{2}A_3(I^X + I^V) - \frac{\lambda}{2}A_3 C H_\Pi.$$

As before we have for η sufficiently small

$$\frac{1}{2}A_3 I \leq J \leq 2A_1 I.$$

Hence, for η possibly even smaller

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(J + (\lambda^2+2)A_3 H_\Pi) &\leq -\frac{\lambda^2\eta}{2(\lambda^2+2)}J - \frac{\lambda C}{\lambda^2+2}\left((\lambda^2+2)A_3 H_\Pi\right)\\ &\leq -\frac{\lambda^2\eta}{2(\lambda^2+2)}\left(J + (\lambda^2+2)A_3 H_\Pi\right).\end{aligned}$$

Phrasing this in terms of I we get

$$I(t) + 2(\lambda^2+2)H_\Pi(t) \leq \exp\left(-\frac{\lambda^2\eta}{2(\lambda^2+2)}t\right)\left(\frac{4(\lambda^2+2)}{\lambda\eta}I + 2(\lambda^2+2)H_\Pi\right).$$

□

5.4 The kinetic Fokker-Planck Equation

Here we consider the case where $L = (\nabla_v - v)\cdot\nabla_v$, and still in the case where I, H represent the p-entropies. The proof is similar to the proof in section one of [116]. We use a different formalism to make the connection with the earlier sections. Again new terms appear which vanish in the case where $p = 1, 2$.

The strategy of the proof is much simpler than in sections 2 and 3. This is because when we differentiate the entropy we get $-I^V$. This means our proof does not need to involve H_Π as we can control I^M terms in the derivative of J by splitting it as a very large amount of I^V and a very small amount of I^X. Then we can cancel out the I^V terms by adding a large amount of H to our functional.

First we calculate the dissipation of the various parts as before.

Lemma 5.8.

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X = -2I^{V,X} - (2-p)(p-1)I^{2,X,V},$$

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^M \leq I^{V,V} + I^{V,X} + \frac{(2-p)(1-p)}{2}\left(I^{2,V} + I^{2,V,X}\right) + \frac{1}{2}I^X + \frac{1}{2}I^V,$$

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^V = -2I^{V,V} - (2-p)(p-1)I^{2,V} + 2I^V.$$

Where,

$$I^{V,X} = \int_{\mathbb{T}^d\times\mathbb{R}^d}\left|\nabla_v\nabla_x\left(\frac{h^{p-1}}{p-1}\right)\right|^2 h^{2-p}\mathrm{d}\mu,\ \ I^{2,X,V} = \int_{\mathbb{T}^d\times\mathbb{R}^d}|\nabla_v h|^2|\nabla_x h|^2 h^{p-4}\mathrm{d}\mu,$$

$$I^{V,V} = \int_{\mathbb{T}^d\times\mathbb{R}^d}\left|\nabla_v\nabla_v\left(\frac{h^{p-1}}{p-1}\right)\right|^2 h^{2-p}\mathrm{d}\mu,\ \ I^{2,V} = \int_{\mathbb{T}^d\times\mathbb{R}^d}|\nabla_v h|^4 h^{p-4}\mathrm{d}\mu.$$

Proof. First we calculate that

$$\begin{aligned}\left|\nabla_c\nabla_{c'}\left(\frac{h^{p-1}}{p-1}\right)\right|^2 h^{2-p} &= \left|\nabla_c\left(h^{p-2}\nabla_{c'}h\right)\right|^2 h^{2-p}\\ &= \left|h^{p-2}\nabla_c\nabla_{c'}h + (p-2)h^{p-3}\nabla_c h\nabla_{c'}h\right|^2 h^{2-p}\\ &= h^{p-2}|\nabla_c\nabla_{c'}h|^2 + 2(p-2)h^{p-3}\nabla_c h\cdot\nabla_c\nabla_{c'}h\nabla_{c'}h\\ &\quad + (p-2)^2h^{p-4}|\nabla_c h|^2|\nabla_{c'}h|^2.\end{aligned}$$

where c, c' are equal to either x or v. A similar argument shows

$$\begin{aligned}h^{2-p}\nabla_v\nabla_v\left(\frac{h^{p-1}}{p-1}\right) : \nabla_v\nabla_x\left(\frac{h^{p-1}}{p-1}\right) =& h^{p-2}\nabla_v\nabla_v : \nabla_v\nabla_x h\\ &+ (p-2)h^{p-3}\nabla_v h\cdot\nabla_x\nabla_v h\nabla_v h\\ &+ (p-2)h^{p-3}\nabla_x h\cdot\nabla_v\nabla_v h\nabla_v h\\ &+ (p-2)^2h^{p-4}|\nabla_v h|^2\nabla_v h\cdot\nabla_x h.\end{aligned}$$

Now we can calculate we have

$$
\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X =&(p-2)\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-3}(\nabla_v+v)\cdot\nabla_v h|\nabla_x h|^2\mathrm{d}\mu \\
&+2\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}\nabla_x((\nabla_v+v)\cdot\nabla_v)\cdot\nabla_x h\mathrm{d}\mu \\
=&-(p-2)\int_{\mathbb{T}^d\times\mathbb{R}^d}\nabla_v\left(h^{p-3}|\nabla_x h|^2\right)\cdot\nabla_v h\mathrm{d}\mu \\
&-2\int_{\mathbb{T}^d\times\mathbb{R}^d}\nabla_v\left(h^{p-2}\nabla_x h\right):\nabla_x\nabla_v h\mathrm{d}\mu \\
=&-(p-2)(p-3)\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-4}|\nabla_x h|^2|\nabla_v h|^2\mathrm{d}\mu \\
&-2(p-2)\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-3}\nabla_x h\cdot\nabla_x\nabla_v h\nabla_v h\mathrm{d}\mu \\
&-2(p-2)\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-3}\nabla_x\cdot\nabla_x\nabla_v h\nabla_v h\mathrm{d}\mu-2\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-2}|\nabla_x\nabla_v h|^2\mathrm{d}\mu \\
=&-2\int_{\mathbb{T}^d\times\mathbb{R}^d}\left|\nabla_v\nabla_x\left(\frac{h^{p-1}}{p-1}\right)\right|^2 h^{2-p}\mathrm{d}\mu-(2-p)(1-p)\int_{\mathbb{T}^d\times\mathbb{R}^d} h^{p-4}|\nabla_v h|^2|\nabla_x h|^2\mathrm{d}\mu.
\end{aligned}
$$

The calculation for I^V is similar except an additional term appears as the commutator of $\partial_{v_i}+v_i$ and ∂_{v_i} is non-zero meaning we gain a $+2I^V$ when doing integration by parts on the second term in the first line. Again for I^M the calculations are similar, this time we only gain $+I^M$ due to the non-zero commutator. □

Remark. *Here again we can see additional terms appearing for $p\in(1,2)$ which are not present at the limit cases similarly to the linear relaxation case.*

We are now able to prove Theorem 5.3.

Proof. Since,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\frac{h^p-h}{p(p-1)}\mathrm{d}\mu=-I^V,$$

we have a lot more flexibility in our proof as we can add H rather that H_Π to the entropy functional and thereby make the coefficient of I^V in the derivative of J as large as we want. So we calculate as before

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}J_\mu\le&-2A_1I^{V,X}-(2-p)(p-1)A_1I^{2,X,V} \\
&-A_2I^X+A_2I^{V,V}+A_2I^{V,X}+\frac{(2-p)(p-1)}{2}A_2(I^{2,V}+I^{2,X,V})+\frac{A_2}{2}I^X+\frac{A_2}{2}I^V \\
&\frac{A_3}{\epsilon}I^X+A_3\epsilon I^V-2A_3I^{V,V}-(2-p)(p-1)A_3I^{2,V}+2A_3I^V \\
=&\left(-2A_1+A_2\right)I^{V,X} \\
&+(2-p)(p-1)\left(-A_1+\frac{A_2}{2}\right)I^{2,X,V}+\left(A_2-2A_3\right)I^{V,V} \\
&+(2-p)(p-1)\left(\frac{A_2}{2}-A_3\right)I^{2,V}+\left(-\frac{A_2}{2}+\frac{A_3}{\epsilon}\right)I^X \\
&+\left(\frac{A_2}{2}+A_3\epsilon+2A_3\right)I^V.
\end{aligned}
$$

Set $\epsilon = 4A_3$ and $A_2 = 1$ then we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} J_\mu \leq &(-2A_1 + 1)\, I^{V,X} + (2-p)(p-1)\left(-A_1 + \frac{1}{2}\right) I^{2,X,V} \\ &+ (1 - 2A_3)\, I^{V,V} + (2-p)(p-1)\left(\frac{1}{2} - A_3\right) I^{2,V} \\ &- \frac{1}{4} I^X + \left(\frac{1}{2} + 4A_3^2 + 2A_3\right) I^V.\end{aligned}$$

So now set $A_1 = A_3 = 1$ to get

$$\frac{\mathrm{d}}{\mathrm{d}t} J_\mu \leq -\frac{1}{4} I^X + \frac{13}{2} I^V.$$

So we have from Beckner's inequalities that

$$H \leq CI$$

for some constant C and we have

$$\frac{1}{2} I \leq J \leq \frac{3}{2} I.$$

Therefore,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(J + \frac{27}{4}H\right) &\leq -\frac{1}{4} I \\ &\leq -\frac{1}{8} I - \frac{1}{8C} H \\ &\leq -\frac{1}{12} J - \frac{1}{8C} H \\ &\leq -\min\{\frac{1}{12}, \frac{4}{216C}\}\left(J + \frac{27}{4}H\right) = -k\left(J + \frac{27}{4}H\right).\end{aligned}$$

Recasting this in terms of I we get

$$I(t) + \frac{27}{2} H(t) \leq e^{-kt}\left(3I(0) + \frac{27}{2} H(0)\right).$$

□

5.5 Quadratic and close to quadratic confinement

This section reviews the work [99] (Monmarché 2017) which proves results based on the calculations in the beginning of this section. We also present it here for completeness, to make some comparisons with Fokker-Planck case and as it motivates the next section which uses similar ideas for the true linear Boltzmann equation. In [7, 3] they study hypocoercivity in ϕ-entropy for kinetic Fokker-Planck equations with linear forces. i.e. equations of the form

$$\partial_t f + v \cdot \nabla_x f - \beta x \cdot \nabla_v = \nabla_v \cdot (\nabla_v f + v f).$$

The goal of these papers is to get sharp rates for the convergence to equilibrium. To do this they take a strategy similar to that of Villani in [116], however they exploit nicely cancellations appearing between the different terms which allow them to close a Grönwall estimate purely on Fisher information. We show these calculations for the above equation in the case of relative entropy rather than general ϕ-entropies. We now write T to be the operator involving confining as well as transport, $v \cdot \nabla_x - \beta x \cdot \nabla_v$.

$$\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^X =& 2\beta I^M, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^V =& -2I^M, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^M =& -I^X + \beta I^V, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X =& -2I^{XV}, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^V =& -2I^{VV} - 2I^V \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^M =& 2\sqrt{I^{XV}I^{VV}} - I^M.
\end{aligned}$$

As before we construct a functional of the form

$$J = AI^X + 2A_2 I^M + A_3 I^V.$$

We choose $A_2^2 < A_1 A_3$ so this means that as before the higher order terms will always give a negative contribution. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} J \leq -2A_2 I^X + (2\beta A_1 - 2A_3 - 2A_2) I^M + (2\beta A_2 - 2A_3) I^V.$$

Therefore we can set $A_1 = 1/\beta + 2, A_2 = 1, A_3 = 2\beta$ to have

$$\frac{\mathrm{d}}{\mathrm{d}t} J \leq -2I^X - 2\beta I^V.$$

We can then close a Grönwall estimate as before. We can choose the coefficients in J better in order to get sharp rates but the key idea is that we can choose A_1 to cancel the rest of the mixed term appearing. We can also extend this to the case where the confining is force is close to being linear, in the sense that

$$U(x) = \frac{\beta}{2}|x|^2 + \tilde{U}(x)$$

where $\mathrm{Hess}\tilde{U} \leq C_\beta$. It was shown in [99] that the same strategy works to get convergence for the linear relaxation Boltzmann case. Here suppose we have the equation

$$\partial_t f + v \cdot \nabla_x f - \beta x \cdot \nabla_v f = \lambda(\tilde{\Pi} f - f).$$

Then we can work in the same way to get

$$\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^X =& 2\beta I^M, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^V =& -2I^M, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T I^M =& -I^X + \beta I^V, \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^X =& -\lambda I^{\Pi X/X} - \lambda(I^X - I^{\Pi X}), \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^V =& -\lambda I^{\Pi V/V} - \lambda I^V \\
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^M =& \lambda\sqrt{I^{\Pi X/X} I^{\Pi V/V}} - \lambda I^M.
\end{aligned}$$

This gives us that

$$\frac{\mathrm{d}}{\mathrm{d}t} J \leq -2A_2 I^X + (2\beta A_1 - 2A_3 - 2\lambda A_2) I^M + (2\beta A_2 - \lambda A_3) I^V.$$

So we can set $A_1 = 2/\lambda + \lambda/\beta, A_2 = 1, A_3 = 2\beta$ to get

$$\frac{\mathrm{d}}{\mathrm{d}t} J \leq -2I^X - 2\beta I^V.$$

So again we can close a Grönwall estimate as before. This can also be extended to close to quadratic confinement when the perturbation away from the quadratic term must be bounded by some constant that depends on both β and λ. The precise statement is given in [99].

It might be interesting at this point to look at how the effect of linear forces looks on the level of SDEs. In the kinetic Fokker-Planck case our equation is the Kolmogorov backwards equation of the SDE

$$\begin{aligned}
\mathrm{d}X_t =& V_t \mathrm{d}t \\
\mathrm{d}V_t =& -\beta X_t \mathrm{d}t - V_t \mathrm{d}t + \sqrt{2}\mathrm{d}W_t.
\end{aligned}$$

Suppose that we generate two coupled solutions to this SDE with the same driving Brownian motion then we have

$$\begin{aligned}
\mathrm{d}(X_t^1 - X_t^2) =& (V_t^1 - V_t^2)\mathrm{d}t, \\
\mathrm{d}(V_t^1 - V_t^2) =& -\beta(X_t^1 - X_t^2)\mathrm{d}t - (V_t^1 - V_t^2)\mathrm{d}t.
\end{aligned}$$

So the difference between these two solutions is deterministic given the initial data and this ODE has an explicit solution from which we can see that

$$(X_t^1 - X_t^2)^2 + (V_t^1 - V_t^2)^2 \leq A_\beta e^{-c_\beta t}\left((X_0^1 - X_0^2)^2 + (V_0^1 - V_0^2)^2\right).$$

Where c_β is the spectral gap of the matrix

$$\begin{pmatrix} 0 & 1 \\ -\beta & -1 \end{pmatrix}.$$

This gives an easy proof of a spectral gap in Wasserstein which captures the rates given in [7, 98] which are there shown to be optimal in relative entropy. This idea can also be extended to the case of close to linear forces as is done in [22]. This is a very different situation to the torus where constructing a coupling to show convergence in Wasserstein is more complicated as is shown in this thesis and in [48]. This is also the case when the confining potential is not a perturbation of quadratic [52]. The couplings used in both these papers really need to use the mixing generated by the Brownian motion to show that the solutions will get close together.

For the linear relaxation Boltzmann equation the stochastic interpretation is less useful. We can write SDEs in integrated form by defining P to be a Poisson point process on $[0,\infty)\times\mathbb{R}^d$ with intensity measure given by the tensor product of Lebesgue in time and Gaussian in velocity space. Then we have

$$\begin{aligned} X_t =& X_0 + \int_0^t V_s \mathrm{d}s, \\ V_t =& V_0 - \int_0^t \beta X_s \mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} (w - V_{s^-})\, P(\mathrm{d}s, \mathrm{d}w). \end{aligned}$$

Now if we couple two processes to have the same law we do not get something deterministic

$$\begin{aligned} X_t^1 - X_t^2 =& X_0^1 - X_0^2 + \int_0^t (V_s^1 - V_s^2)\mathrm{d}s \\ V_t^1 - V_t^2 =& V_0^1 - V_0^1 - \int_0^t \beta(X_s^1 - X_s^2)\mathrm{d}s - \int_0^t \int_{\mathbb{R}^d} (V_{s^-}^1 - V_{s^-}^2) P(\mathrm{d}s, \mathrm{d}w). \end{aligned}$$

So the dynamics of this are that the difference between the solutions follows a deterministic path keeping $(V_t^1 - V_t^2)^2 + \beta(X_t^1 - X_t^2)^2$ constant then at random times the velocity part jumps to zero. It seems very likely that this coupling should give exponential convergence to equilibrium but it is not immediate as it was for the kinetic Fokker-Planck equation.

5.6 The linear Boltzmann equation

We now look at the linear Boltzmann equation for Maxwell molecules with cut off. This equation is written as

$$\partial_t f + v \cdot \nabla_x f - \nabla_x U \cdot \nabla_v f = \int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} B\left(\frac{v - v_*}{|v - v_*|} \cdot \sigma\right) f(v') \mathcal{M}(v_*') \mathrm{d}v_* \mathrm{d}\sigma - f. \tag{5.8}$$

Where

$$v' = \frac{v + v_*}{2} + \frac{|v - v_*|}{2}\sigma = v - (v - v_*) \cdot \tilde{\sigma}\tilde{\sigma},$$

$$v_*' = \frac{v + v_*}{2} = \frac{|v - v_*|}{2}\sigma.$$

Where

$$\tilde{\sigma} = \frac{k-\sigma}{\sqrt{2(1-k\cdot\sigma)}}, \quad k = \frac{v-v_*}{|v-v_*|}.$$

We note at this point that

$$\left|\frac{\partial\tilde{\sigma}}{\partial\sigma}\right| = -4(1-k\cdot\sigma)^3.$$

We also have that

$$\int B(k\cdot\sigma)\mathrm{d}\sigma = 1.$$

Now we look at $h = f/\mathcal{M}$ we get that

$$\partial_t h + v\cdot\nabla_x h = \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)h(v')\mathcal{M}(v_*)\mathrm{d}v_*\mathrm{d}\sigma - h.$$

Our goal is to see if we can prove similar theorems as for the linear relaxation Boltzmann equation for this more complicated operator. Unfortunately, I don't know how to replicate this proof on the torus using techniques similar to the first chapter. However, we can show a result for a close to quadratic confining potential as was discussed in the last section. In order to do this we need work out how to control the dissipation of Fisher information type terms along the flow of the collision operator. We do this using techniques from [115]. In order to do this we need to change the way we write the twisted Fisher information terms. We define

$$I^{a,b} = \int \frac{|a\nabla_x h + b\nabla_v h|^2}{h}\mathrm{d}\mu.$$

We will write our twisted Fisher information as

$$J = I^{a,b} + I^{0,c},$$

and choose a, b, c.

Lemma 5.9. *We can calculate that*

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L I^{a,b} \leq -\int_{\mathbb{R}^d\times\mathbb{R}^d}\left(\frac{b^2}{2}\frac{|\nabla_v h|^2}{h} + ab\frac{\nabla_x h\cdot\nabla_v h}{h}\right)\mathrm{d}\mu.$$

Proof. We define

$$Q_+(h) = \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)h(v')\mathcal{M}(v_*)\mathrm{d}v_*\mathrm{d}\sigma,$$

and we want to bounds

$$\int_{\mathbb{R}^d\times\mathbb{T}^d}\frac{|(a\nabla_x + b\nabla_v)Q_+(h)|^2}{Q_+(h)}\mathrm{d}\mu.$$

Now we follow very closely [115] but adapting it to our case where we work in terms of h not f and use derivatives in x as well as v. We define as in [115]

$$M_{\sigma k}x = (\sigma^\perp\cdot x)k^\perp, P_{\sigma k}x = (\sigma\cdot x)k + M_{\sigma k}x.$$

We use the following lemmas.

Lemma 5.10 (Lemma 1 from [115])**.**

$$\nabla_v(B(k\cdot\sigma)) = \frac{1}{|v-v_*|}B'(k\cdot\sigma)\Pi_{k^\perp}\sigma.$$

Lemma 5.11 (Lemma 2 in [115])**.**

$$\int_{\mathbb{S}^{d-1}} \mathrm{d}\sigma B'(k\cdot\sigma)F(\sigma)\Pi_{k^\perp}\sigma = \int_{\mathbb{S}^{d-1}} \mathrm{d}\sigma B(k\cdot\sigma)M_{\sigma k}\nabla_\sigma F(\sigma).$$

Lemma 5.12 (Lemma 4 in [115])**.**

$$\|P_{\sigma k}x\| \le \|x\|.$$

With these we can compute that

$$\begin{aligned}
(a\nabla_x + b\nabla_v)Q_+(h) &= \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} \bigg(B(k\cdot\sigma)(a\nabla_x h(v') + b\frac{1}{2}(I+\sigma k^T)(\nabla_v h)(v')) \\
&\quad + \frac{b}{|v-v_*|}B'(k\cdot\sigma)\Pi_{k^\perp}\sigma h(v')\bigg)\mathcal{M}(v_*) \\
&= \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)\bigg((a\nabla_x h(v') + b\frac{1}{2}(I+\sigma k^T)(\nabla_v h)(v')) \\
&\quad + \frac{b}{|v-v_*|}M_{\sigma k}\nabla_\sigma(h(v'))\bigg)\mathcal{M}(v_*) \\
&= \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)\left(a\nabla_x h(v') + b\frac{1}{2}(I+\sigma k^T + M_{\sigma k})(\nabla_v h)(v')\right)\mathcal{M}(v_*) \\
&= \int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)\left(a\nabla_x h(v') + b\frac{1}{2}(I+P_{\sigma k})(\nabla_v h)(v')\right)\mathcal{M}(v_*).
\end{aligned}$$

Squaring this and using Jensen's inequality gives and that $|P_{\sigma k}x|^2 \le |x|^2$

$$\begin{aligned}
\frac{|(a\nabla_x + b\nabla_v)Q_+ h|^2(v)}{(Q_+(h))^2} \le &\frac{1}{Q_+(h)}\int_{\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)\mathcal{M}(v_*)\bigg(a^2\frac{|\nabla_x h(v')|^2}{h(v')} + \frac{b^2}{2}\frac{|(\nabla_v h)(v')|^2}{h(v')} \\
&+ ab\frac{\nabla_x h(v')\cdot(\nabla_v h)(v')}{h(v')} + \frac{\left(a\nabla_x h(v') + \frac{b}{2}\nabla_v h(v')\right)\cdot P_{\sigma k}\nabla_v h(v')}{h(v')}\bigg).
\end{aligned}$$

Integrating and switching the primes to not primes we get that

$$\begin{aligned}
\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|(a\nabla_x + b\nabla_v)Q_+(h)|^2}{Q_+(h)}\mathrm{d}\mu \le &\int_{\mathbb{R}^d\times\mathbb{T}^d\times\mathbb{R}^d\times\mathbb{S}^{d-1}} B(k\cdot\sigma)\mathcal{M}(v_*)\bigg(a^2\frac{|\nabla_x h(v)|^2}{h(v)} + \frac{b^2}{2}\frac{|\nabla_v h(v)|^2}{h(v)} \\
&ab\frac{\nabla_x h(v)\cdot\nabla_v h(v)}{h(v)} + \frac{(a\nabla_x h(v) + \frac{b}{2}\nabla_v h(v))\cdot P_{k\sigma}\nabla_v h(v)}{h(v)}\bigg).
\end{aligned}$$

Integrating in σ gets rid of the last term as $P_{k\sigma}$ is an odd function of σ and B is even. So finally we get that

$$\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|(a\nabla_x + b\nabla_v)Q_+(h)|^2}{Q_+(h)}\mathrm{d}\mu \le \int_{\mathbb{R}^d\times\mathbb{T}^d}\left(a^2\frac{|\nabla_x h|^2}{h} + \frac{b^2}{2}\frac{|\nabla_v h|^2}{h} + ab\frac{\nabla_x h\cdot\nabla_v h}{h}\right)\mathrm{d}\mu.$$

So now we know that $I_{a,b}$ is a convex function of h. Therefore if $L = Q_+ - I$ we have that

$$I_{a,b}(e^{tL}h) = I_{a,b}\left((1-t)h + t\left(Q_+(h) + o(t)\right)\right) \le (1-t)I_{a,b}(h) + tI_{a,b}(Q_+(h) + o(t)).$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}t} I_{a,b}(e^{tL}h) \leq I_{a,b}(Q_+(h)) - I_{a,b}(h).$$

Using our expression for $I_{a,b}(Q_+(h))$ this gives

$$\frac{\mathrm{d}}{\mathrm{d}t} I_{a,b}(e^{tL}h) \leq -\int_{\mathbb{R}^d\times\mathbb{T}^d} \left(\frac{b^2}{2}\frac{|\nabla_v h|^2}{h} + ab\frac{\nabla_x h\cdot\nabla_v h}{h}\right)\mathrm{d}\mu$$

□

Now we want to look at the case that

$$U(x) = \beta|x|^2 + \tilde{U}(x).$$

Let us define $T_1 = -v\cdot\nabla_x + \beta x\cdot\nabla_v$ and $T_2 = \nabla_x\tilde{U}\cdot\nabla_v$.

Lemma 5.13. *Suppose that $\|Hess\tilde{U}\| \leq \delta$ then we have that*

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{T_1} I^{a,b} =& -2ab\int\frac{|\nabla_x h|^2}{h}\mathrm{d}\mu + 2(a^2\beta - b^2)\int\frac{\nabla_x h\cdot\nabla_v h}{h}\mathrm{d}\mu + 2ab\beta\int\frac{|\nabla_v h|^2}{h}\mathrm{d}\mu,\\ \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{T_2} I^{a,b} \leq& 2a^2\delta\left(\int\frac{|\nabla_x h|^2}{h}\mathrm{d}\mu\int\frac{|\nabla_v h|^2}{h}\mathrm{d}\mu\right)^{1/2} + 2ab\delta\int\frac{|\nabla_v h|^2}{h}\mathrm{d}\mu.\end{aligned}$$

Proof. The first line follows simply from the chain rule as before. For the second we just use the chain rule and then split the mixed term using Cauchy-Schwartz. □

Theorem 5.4. *Suppose $f(t)$ is a solution to the linear Boltzmann equation as shown above (5.8) with*

$$U(x) = \beta\frac{|x|^2}{2} + \tilde{V}(x)$$

where $Hess\tilde{U} \leq \delta$ with δ satisfying

$$(9 + 8(\beta+\delta))\delta < 4\beta.$$

Then we have that

$$I_\mu(f(t)) \leq Ce^{-\lambda t}I_\mu(f(0)).$$

Proof. We define $J = I^{a,b} + I^c$ then using Lemmas 5.13 and 5.9 we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}J \leq -2abI^X + (2ab\beta + 2ab\delta - b^2/2 - c^2/2)I^V + (-ab + 2(a^2\beta - b^2 - c^2))I^M + 2a^2\delta\sqrt{I^X I^V}.$$

We can choose a, b, c such that

$$ab = 1, \quad b^2 + c^2 = 4(\beta+\delta) + 4, \quad a^2 = \frac{2\beta}{(9 + 8(\beta+\delta))}.$$

This gives

$$\frac{\mathrm{d}}{\mathrm{d}t}J \leq -2(I^X + I^V) + \frac{9 + 8(\beta+\delta)}{\beta}\sqrt{I^X I^V}.$$

Then provided that δ is small in terms of β as required above we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}J \leq -2\lambda J.$$

We do not keep track of the constants but they can be made explicit. Since J is equivalent to $I_\mu(f)$ and we get exponential convergence for J:

$$I_\mu(f(t)) \leq Ce^{-\lambda t} I_\mu(f(0)).$$

□

5.7 Proofs that the results extend beyond smooth functions

We show for h, being bounded above and below and having bounded derivatives of all orders is propagated by the equation (this is similar to what is shown in the appendix of [33]). In this set we can do all the calculations given in the main part of the paper. We then show for $h \in W^{1,1}(\mu)$ with finite Fisher information then we can make a density argument to show that the result still holds in this case.

Lemma 5.14. *The equation preserves bounded derivatives of all orders.*

Proof. We rewrite the equation for h in a mild formulation as follows

$$e^{\lambda t} h(t,x,v) = h(0, x - vt, v) + \lambda \int_0^t e^{\lambda s} \int h(s, x - v(t-s), u)\mathcal{M}(u)\mathrm{d}u\mathrm{d}s.$$

This leads to the following inequality

$$e^{\lambda t}\|D_x^\alpha h(t)\|_\infty \leq \|D_x^\alpha h(0)\| + \lambda \int_0^t e^{\lambda s}\|D_x^\alpha h(s)\|_\infty \mathrm{d}s.$$

Therefore by Grönwall's inequality we have that

$$\|D_x^\alpha h(t)\|_\infty \leq \|D_x^\alpha h(0)\|_\infty.$$

We also from this mild formulation that any mixed derivative can be written in terms of x derivative and derivatives of the initial data. Therefore, the derivatives will remain in L^∞ for all time. □

Lemma 5.15. *The equation preserves positivity and constants are a steady state of the equation therefore being bounded above and below is preserved.*

Proof. We can show that

$$\partial_t \left(e^{\lambda t} h(t, x+vt, v)\right) = \int \lambda e^{\lambda t} h(t, x+vt, u)\mathcal{M}(u)\mathrm{d}u.$$

Therefore if $e^{\lambda t}h(t,x+vt,v)$ is positive for all x and v then so is its derivative. Therefore it will remain positive for all time.

It is easy to check that constants are a steady state so if $h(0) - c$ is positive then since positivity is preserved so is $h(t) - c$ and similarly if $C - h(0)$ is positive then so is $C - h(t)$. □

Lemma 5.16. *Suppose that we have $h(0)$ is in $W^{1,1}(\mu)$ with bounded Fisher information, and also suppose we have a sequence $h_n(0)$ which has all our good properties and converges to $h(0)$ in $L^1(\mu)$ with*

$$H(h_n(t)) \leq Ae^{-\Lambda t} I(h_n(0)),$$

for every n then we have

$$H(h(t)) \leq Ae^{-\Lambda t} I(h(0)).$$

Proof. Convergence in L^1 implies that h_n tends to h a.e. along a subsequence. Also, suppose that h_1 and h_2 are two solutions to the equation then

$$\sup_{s \leq t} \|h_1(s) - h_2(s)\|_{L^1(\mu)} \leq e^{-\lambda t} \|h_1(0) - h_2(0)\|_{L^1(\mu)} + \sup_{s \leq t} \|h_1(s) - h_2(s)\|_{L^1(\mu)} (1 - e^{-\lambda t}).$$

Therefore,

$$\sup_{s \leq t} \|h_1(s) - h_2(s)\|_{L^1(\mu)} \leq \|h_1(0) - h_2(0)\|_{L^1(\mu)},$$

hence $h_n(t)$ tends to $h(t)$ in L^1 therefore $h_n(t)$ also converges to $h(t)$ almost everywhere along a subsequence.

Then since $h_n \log(h_n) - h_n + 1 \geq 0$ by Fatou's lemma we have

$$\int (h(t,x,v) \log(h(t,x,v)) - h(t,x,v) + 1) \mathrm{d}\mu \leq \liminf_n \int (h_n(t,x,v) \log(h_n(t,x,v)) - h_n(t,x,v) + 1) \mathrm{d}\mu.$$

Therefore, if we have h a solution to the equation with initial data $h(0)$ as defined above we have that

$$H(h(t)) \leq \liminf_n Ae^{-\Lambda t} I(h_n(0)).$$

So to prove our theorem holds in this larger set it remains to show that we can find a sequence $h_n(0)$ converging to $h(0)$ in $L^1(\mu)$ where for every n $h_n(0)$ is positive, integrates to 1 against μ, is bounded below and has derivatives bounded of all orders which also satisfies

$$\liminf_n I(h_n(0)) \leq I(h(0)).$$

To do this we make a very standard mollifier argument. Let χ be a smooth function on $\mathbb{R}_+$ with $\chi(x) = 1$ for $x < 1$ and $\chi(x) = 0$ for $x > 2$ and $|\chi'(x)|^2/\chi(x)$ integrable. Then define $\chi_R(x,v) = \chi(\|v\|/R)$. Also let ϕ be a molifier integrating to one and compactly supported in $B(0,1)$ then set $\phi_\epsilon(x,v) = \epsilon^{-2d}\phi((x,v)/\epsilon)$. Take some h in $W^{1,1}(\mu)$ with finite Fisher information. Let $h_R = h\chi_R$, then set $h_{\epsilon,R} = \phi_\epsilon \star h_R$ and then $h_{\eta,\epsilon,R} = (h_{\epsilon,R} + \eta)/(\|h_{\epsilon,R}\|_1 + \eta)$. So $h_{\eta,\epsilon,R}$ is bounded below and has derivatives bounded of all orders and fairly clearly converges to h in $L^1(\mu)$.

So first we try and get rid of η since $\nabla h_{\eta,\epsilon,R} = \nabla h_{\epsilon,R}/(\|h_{\epsilon,R}\|_1 + \eta)$ we get that

$$\frac{|\nabla h_{\eta,\epsilon,R}|^2}{h_{\eta,\epsilon,R}} \text{ increases to } \frac{|\nabla h_{\epsilon,R}|^2}{h_{\epsilon,R}}.$$

Therefore, by monotone convergence,

$$\lim_{\eta \to 0} I(h_{\eta,\epsilon,R}) = I(h_{\epsilon,R}).$$

Now we work on ϵ, we have that $\nabla h_{\epsilon,R} = \phi_\epsilon \star \nabla h_R$. We can now make a similar argument based on Jensen's inequality and the fact that $|x|^2/y$ is convex to get that

$$\frac{|\nabla h_{\epsilon,R}|^2}{h_{\epsilon,R}} \leq \phi_\epsilon \star \left(\frac{|\nabla h_R|^2}{h_R}\right).$$

Since, the mollification of an L^1 function converges in L^1 to that function we get that

$$\lim_{\epsilon\to 0} I(h_{\epsilon,R}) \leq I(h_R).$$

Now we work on R, we note that

$$\frac{|\nabla h_R|^2}{h_R} = \chi_R \frac{|\nabla h|^2}{h} + \frac{2}{R}\chi'(\|v\|/R)\frac{v}{\|v\|} \cdot \nabla h + h\frac{1}{R^2}\frac{(\chi'(\|v\|/R))^2}{\chi(\|v\|/R}.$$

Since, $h, \nabla h, |\nabla h|^2/h$ are all in $L^1(\mu)$ we can see that

$$\lim_{R\to\infty} I(h_R) = I(h).$$

□

5.8 General Entropies

In fact the proofs given above work for a more general class of entropies, Φ-entropies defined by

$$H^\Phi = \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi(h)\mathrm{d}\mu$$
$$I^\Phi = \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla h|^2\mathrm{d}\mu.$$

We work with Φ a positive function such that $\Phi(1) = 0, \Phi''(t) > 0 \ \forall t$, $1/\Phi''(t)$ a concave function and $\Phi(t)\Phi''(t) > 2\Phi'(t)^2 \ \forall t$. We use a new method of differentiating the entropies which allows us to extend the calculations to a more general class of Φ-entropy.

Let us define the entropy

$$J_\mu^\Phi(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)\left(a|\nabla_x h|^2 + 2b\nabla_x h\cdot\nabla_v h + c|\nabla_v h|^2\right)\mathrm{d}\mu.$$

Lemma 5.17. *Let Φ satisfy for all $t > 0$:*

- $\Phi(t) \geq 0$
- $\Phi''(t) \geq 0$
- $\Phi''(t)\Phi^{(4)}(t) > 2\Phi^{(3)}(t)^2$

Then if $b^2 \leq ac$ then J is a convex functional.

Proof. Since $b^2 < ab$ we can write J as the sum of functionals like

$$\tilde{J}(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\alpha\nabla_x h + \beta\nabla_v h|^2\mathrm{d}\mu.$$

Then if the function

$$\phi(x,y) = \Phi''(y)|x|^2$$

the whole functional will be convex. This is because if ϕ is convex then

$$\begin{aligned}\tilde{J}(th+(1-t)g) =& \int_{\mathbb{R}^d\times\mathbb{T}^d} \phi(t(\alpha\nabla_x h+\beta\nabla_v h)+(1-t)(\alpha\nabla_x g+\beta\nabla_v g), th+(1-t)g)\mathrm{d}\mu \\ \leq& \int_{\mathbb{R}^d\times\mathbb{T}^d} \left(t\phi(\alpha\nabla_x h+\beta\nabla_v h, h)+(1-t)\phi(\alpha\nabla_x g+\beta\nabla_v g, g)\right)\mathrm{d}\mu \\ =& t\tilde{J}(h)+(1-t)\tilde{J}(g).\end{aligned}$$

It remains to prove that ϕ is convex. We know that ϕ is the sum of functions $\tilde{\phi} = \Phi''(y)x^2$ where now x is one dimensional. So we only need to show that these are convex. The Hessian of $\tilde{\phi}$ is

$$\begin{pmatrix} 2\Phi''(y) & 2x\Phi^{(3)}(y) \\ 2x\Phi^{(3)}(y) & x^2\Phi^{(4)}(y) \end{pmatrix}.$$

This has positive trace as both diagonal terms are positive by our assumptions. It also has determinant $2x^2\Phi''(x)\Phi^{(4)}(x) - 4x^2\Phi^{(3)}(x)^2$ which is again positive due to he assumptions we made on Φ therefore the Hessian is positive definite so $\tilde{\phi}$ is convex. □

Theorem 5.5. *Let Φ satisfy the conditions in Lemma 5.17 and also let Φ be such that the uniform measure on the torus satisfies a Φ Sobolev inequality and $1/\Phi''$ is a concave function. If f is a solution to (5.1) with initial data f_0 such that*

$$\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h_0)|\nabla_{x,v}h_0|^2\mathrm{d}\mu < \infty, \qquad f_0 \in W^{1,1}(\mu),$$

then there exist constants $\Lambda > 0$ and $A > 0$ depending on λ and the constant in the Φ-Sobolev inequality, such that

$$I^\Phi_\mu(h_t) + H^\Phi_\mu(\Pi h_t) \leq A\exp(-\Lambda t)\left(I^\Phi_\mu(h_0) + H^\Phi_\mu(\Pi h_0)\right).$$

This implies that if the equilibrium measure satisfies a Φ-Sobolev inequality then for some γ,

$$H(h_t) \leq \gamma\exp(-\Lambda t)\, I(h_0).$$

We can take

$$\Lambda = \min\left\{1, \frac{C}{4(1+\lambda)}\right\}\min\{2, \lambda/2\}$$

and

$$A = 4\max\{2(1+1/\lambda)^2, (1+\lambda)\}.$$

Here C is the constant in the Φ-Sobolev inequality for the uniform measure on the torus.

Remark. *The conditions of Φ are satisfied when Φ is one of*

$$\Phi_1(t) := t\log(t) - t + 1$$

and

$$\Phi_p(t) := \frac{1}{p-1}(t^p - 1 - p(t-1)),$$

where $p \in (1,2]$ *which are introduced below.*

In order to prove our theorem we would like to study how a functional like J behaves under the action of the collision part of the operator. We write $L = \lambda(\Pi - I)$ and $T = -v \cdot \nabla_x$ and write $(\mathrm{d}/\mathrm{d}t)_O$ to write the derivative along the flow of the operator O. We have that

Lemma 5.18. *We can differentiate* J *along the flow of* L *to get that*

$$\begin{aligned}\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_L J_\mu^\Phi(h) \leq & a\left(\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)|\nabla_x \Pi h|^2 \mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi'(h)|\nabla_x h|^2 \mathrm{d}\mu\right) \\ & - 2b\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)\nabla_x h\cdot\nabla_v h \mathrm{d}\mu - c\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_v h|^2\mathrm{d}\mu.\end{aligned}$$

Proof. As J_μ^Φ is convex we can see by Taylor expanding that

$$\begin{aligned}J_\mu^\Phi(e^{Ls}h(t)) =& J_\mu^\Phi\left(h(t) + \lambda s(\Pi - I)h(t) + o(s)\right) \\ \leq & (1-\lambda s)J_\mu^\Phi\left(h(t) + o(s)\right) + \lambda s J_\mu^\Phi(\Pi h(t)).\end{aligned}$$

Now we calculate that

$$\begin{aligned}J_\mu^\Phi(\Pi h) =& \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)\left(a|\nabla_x \Pi h|^2 + 2b\nabla_x\Pi h\cdot\nabla_v \Pi h + c|\nabla_v\Pi h|^2\right)\mathrm{d}\mu \\ =& a\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)|\nabla_x\Pi h|^2\mathrm{d}\mu.\end{aligned}$$

This means that

$$\begin{aligned}J_\mu^\Phi(e^{sL}h(t)) - J_\mu^\Phi(h(t)) \leq & \lambda s a\left(\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)|\nabla_x\Pi h|^2\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_x h|^2\mathrm{d}\mu\right) \\ & - \lambda s b\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)\nabla_x h\cdot\nabla_v h\mathrm{d}\mu \\ & - \lambda s c\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_v h|^2\mathrm{d}\mu \\ & + J_\mu^\Phi(h(t)+o(s)) - J_\mu^\Phi(h(t)).\end{aligned}$$

Dividing by s and taking the limit as $s \to 0$ gives the result. $\square$

We now look at how J behaves under the flow of T.

Lemma 5.19. *We can differentiate* J *along the flow of* T *to get that*

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T J_\mu^\Phi(h) = -2b\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_x h|^2\mathrm{d}\mu - 2c\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)\nabla_x h\cdot\nabla_v h\mathrm{d}\mu.$$

Proof. We apply the chain rule. We have

$$\begin{aligned}
\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_T J_\mu^\Phi(h) =& - a\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi'''(h)(v\cdot\nabla_x h)|\nabla_x h|^2\mathrm{d}\mu - 2a\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)\nabla_x(v\cdot\nabla_x h)\cdot\nabla_x h\mathrm{d}\mu\\
&- 2b\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi'''(h)(v\cdot\nabla_x h)\nabla_x h\cdot\nabla_v h\mathrm{d}\mu - 2b\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)\nabla_x(v\cdot\nabla_x h)\cdot\nabla_v h\mathrm{d}\mu\\
&- 2b\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)\nabla_x h\cdot\nabla_v(v\cdot\nabla_x h)\mathrm{d}\mu - c\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi'''(h)(v\cdot\nabla_x h)|\nabla_v h|^2\mathrm{d}\mu\\
&- 2c\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)\nabla_v(v\cdot\nabla_x h)\cdot\nabla_v h\mathrm{d}\mu\\
=&\int_{\mathbb{R}^d\times\mathbb{T}^d} v\cdot\nabla_x\left(\Phi''(h)(a|\nabla_x h|^2 + 2b\nabla_x h\cdot\nabla_v h + c|\nabla_v h|^2\right)\mathrm{d}\mu\\
&- 2b\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)|\nabla_x h|^2\mathrm{d}\mu - 2c\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)\nabla_x h\cdot\nabla_v h\mathrm{d}\mu\\
=& - 2b\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)|\nabla_x h|^2\mathrm{d}\mu - 2c\int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)\nabla_x h\cdot\nabla_v h\mathrm{d}\mu.
\end{aligned}$$

□

Now we need to show our helpful lemma relating projected entropy to the mixed term. This result relates the quantities involving only Πh to quantities coming from the full Fisher information. For this we define the local average speed $U(x)$, of a solution to (5.1) by

$$U(x) := \int_{\mathbb{R}^d} vh(v,x)\mathcal{M}(v)\mathrm{d}v = \int_{\mathbb{R}^d} vf(v,x)\mathrm{d}v.$$

Lemma 5.20. *Suppose that the uniform measure on the torus satisfies a Φ-Sobolev inequality. Then for any h we have that*

$$I^{\Pi X}(h) = \int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(\Pi h)|\nabla_x\Pi h|^2\mathrm{d}\mu \le \int_{\mathbb{R}^d\times\mathbb{T}^d}\Phi''(h)|\nabla_x h|^2\mathrm{d}\mu.$$

This implies that for all h there exists a constant C such that

$$H_\Pi(h) = \int_{\mathbb{T}^d}\Phi(\Pi h)\mathrm{d}x \le C\int_{\mathbb{T}^d\times\mathbb{R}^d}\Phi''(h)|\nabla_x h|^2\mathrm{d}\mu.$$

Finally, if h is a solution to (5.2) then

$$\frac{\mathrm{d}}{\mathrm{d}t}H_\Pi(h(t)) = -\int_{\mathbb{T}^d}\Phi'(\Pi h)\nabla_x\cdot U(x)\mathrm{d}x.$$

Proof. We can see that the first inequality will follow from

$$\Phi''(\Pi h)|\nabla_x\Pi h|^2 \le \Pi\left(\Phi''(h)|\nabla_x h|^2\right).$$

Since Π is integrating against a probability measure we would like to use Jensen's inequality. Instead of looking at h we consider $H = (\nabla_x h, h)$ we have already shown the function $\phi(\mathbf{x}, y) = \Phi''(y)|\mathbf{x}|^2$ is convex so from Jensen's inequality we have

$$\phi(\Pi H) \le \Pi(\phi(H)).$$

This implies our desired result since Π commutes with ∇_x. (Here Π acts component wise on vectors).

Now since we have a Φ-Sobolev inequality for the uniform measure on the torus we have

$$\int_{\mathbb{T}^d} \Phi(\Pi h)\mathrm{d}x \leq C \int_{\mathbb{T}^d} \Phi''(\Pi h)|\nabla_x \Pi h|^2 \mathrm{d}x.$$

We can then conclude this part by the first inequality.

For the last part,

$$\begin{aligned}
\partial_t \Pi h =& -\int_{\mathbb{R}^d} v\nabla_x h \mathcal{M}(v)\mathrm{d}v + \lambda\Pi(\Pi h) - \lambda \Pi h \\
=& -\nabla_x \cdot U(x).
\end{aligned}$$

This implies that

$$\partial_t H_\Pi = \int_{\mathbb{T}^d} \Phi'(\Pi h)\partial_t \Pi h \mathrm{d}x = -\int_{\mathbb{T}^d} \Phi'(\Pi h)\nabla_x \cdot U(x)\mathrm{d}x.$$

□

We now need a lemma which will help us control the mixed derivative.

Lemma 5.21. *If $1/\Phi''(t)$ is a concave function then for any positive η we have*

$$\begin{aligned}
-\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)\nabla_x h\cdot\nabla_v h \mathrm{d}\mu \leq& \frac{\eta}{2}\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_v h|^2\mathrm{d}\mu \\
&+ \frac{1}{2\eta}\left(\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)|\nabla_x\Pi h|^2\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_x h|^2\mathrm{d}\mu\right) \\
&- \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi(\Pi h)\mathrm{d}\mu.
\end{aligned}$$

Proof. We need to rewrite the mixed term

$$\begin{aligned}
-\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)\nabla_x h\cdot\nabla_v h\mathrm{d}\mu =& -\int_{\mathbb{R}^d\times\mathbb{T}^d} \nabla_x U'(h)\cdot\nabla_v h\mathrm{d}\mu \\
=& -\int_{\mathbb{R}^d\times\mathbb{T}^d} (\nabla_x U'(h) - \nabla_x U'(\Pi h))\cdot\nabla_v h\mathrm{d}\mu \\
&-\int_{\mathbb{R}^d\times\mathbb{T}^d} \nabla_x U'(\Pi h)\cdot\nabla_v h\mathrm{d}\mu \\
\leq& \frac{\eta}{2}\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_v h|^2\mathrm{d}\mu \\
&+ \frac{1}{2\eta}\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x U'(h) - \nabla_x U'(\Pi h)|^2}{\Phi''(h)}\mathrm{d}\mu \\
&- \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi'(\Pi h)\nabla_x\cdot U(x)\mathrm{d}\mu.
\end{aligned}$$

We get the equality for the last term since

$$-\int \nabla_v h\mathcal{M}(v)\mathrm{d}v = -\int vh\mathcal{M}(v)\mathrm{d}v = U(x).$$

Then we can use the last part of lemma 5.20. Now we observe that

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x\Phi'(h)-\nabla_x\Phi'(\Pi h)|^2}{\Phi''(h)}\mathrm{d}\mu = &\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_x h|^2\mathrm{d}\mu \\ &- 2\int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)\nabla_x h\cdot\nabla_x\Pi h\mathrm{d}\mu \\ &+ \int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{\Phi''(\Pi h)^2}{\Phi''(h)}|\nabla_x\Pi h|^2\mathrm{d}\mu.\end{aligned}$$

Now we see in the second term the only part which depends on v is the $\nabla_x h$ so we can replace it by $\nabla_x\Pi h$. The last term is positive and the only term which depends on v is $1/\Phi''(h)$ since we have that $1/\Phi''(h)$ is a concave function we have

$$\Pi\left(\frac{1}{\Phi''(h)}\right) \leq \frac{1}{\Phi''(\Pi h)}.$$

Therefore we have that

$$\int_{\mathbb{R}^d\times\mathbb{T}^d} \frac{|\nabla_x\Phi'(h)-\nabla_x\Phi'(\Pi h)|^2}{\Phi''(h)}\mathrm{d}\mu \leq \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(h)|\nabla_x h|^2\mathrm{d}\mu - \int_{\mathbb{R}^d\times\mathbb{T}^d} \Phi''(\Pi h)|\nabla_x\Pi h|^2\mathrm{d}\mu.$$

This completes the proof of our lemma. □

Now we can prove the main theorem

Proof of Theorem 5.5. The proof of this is now exactly the same as for Boltzmann entropy. □

Chapter 6

Non-equilibrium steady states in Kac's model coupled to a thermostat

6.1 Introduction

Kac's model was introduced by Mark Kac in 1956 [85]. It is a stochastic N-particle model designed to mimic the dynamics of velocities of particles in a spatially homogeneous dilute gas. The dynamics are those of N particles with one dimensional velocities, these particles interact in a Markov process, where two particles "collide" resulting in a mixing of their velocities. The state of the system can be described by the vector of velocities of each of the particles. Kac derived an equation on the law of this system, this equation is usually called the Kac master equation and it is a linear integro-differential equation. Kac showed that, in a certain sense, as the number of particles goes to infinity the master equation tends to a Boltzmann like equation. This motivates estimates on the behaviour of the marginals of solutions which are uniform in the number of particles, which could then be used to show, or at least indicate, the same behaviour for the Boltzmann equation. In general a direct study of the Boltzmann equation has proved more fruitful, however the master equation has become an object of study in its own right. Convergence to equilibrium and spectral gaps have been studied in Kac's master equation in both entropy [37, 56] and L^2 [83, 35]. This paper studies convergence to equilibrium for solutions of the master equation coupled to a thermostat. More precisely, we study the master equation for a system of N particles who, as well as "colliding" with each other, can also "collide" with some infinite collection of other particles whose velocities lie in some fixed distribution. When this fixed distribution is not a Maxwellian this allows for the possibility of a non-equilibrium steady state. One possible more physical interpretation of this would be if the system was interacting with two different heat baths at different temperatures. Situations related to the existence and convergence to non-equilibrium steady states are studied in [24, 1, 62, 55, 107] and in particular looking at exponential convergence in [108, 53].

This chapter is fundamentally motivated by two others the first [26] studies a similar model but only in the situation where the thermal bath is a Maxwellian distribution. They show exponential convergence to equilibrium in both entropy and L^2. The second [39] studies the existence of non-equilibrium steady states in various coupled equations arising from mathematical physics including the non-linear spatially homogeneous Boltzmann equation. The paper [26] suggest as a further question, what would happen in the case of a non-Maxwellian reservoir and we adapt the techniques of [39] to study this situation. We also include a study of how our estimates on the first marginal behave as the number of particles $N \to \infty$. This allows us, in some sense, to commute the long time and $N \to \infty$ limit. The $N \to \infty$ limit is very similar to the equations studied in [39], they study a coupled Boltzmann equation where in our case the limit would be a coupled Boltzmann-Kac equation. The convergence, both in this paper and in the Maxwellian case studied in [26], is primarily driven by the external force and not by the Kac mixing part. However, the effect of the Kac part is more evident in this paper since it affects the form of the steady state. The work in [26] has been extended in [113, 25] to study how their thermostatted model relates to a partially thermostatted model and to the original Kac's model. In this second paper they make use of the GTW distance used in our work.

Following the strategy of [39] we study the problem of convergence to equilibrium in the Gabetta-Toscani-Wennberg metric . This metric is introduced in [59] and is

$$d_{GTW,N}(f,h) = \sup_{\xi\in\mathbb{R}^N,\xi\neq 0} \frac{|\hat{f}(\xi)-\hat{h}(\xi)|}{|\xi|^2},$$

where $\hat{f}$ represents the Fourier transform of f. This is a metric on the space of probability measures with finite second moment and the same finite first moment. We also study convergence in the metric

$$d_{T1,N}(f,h) = \sup_{\xi\in\mathbb{R}^N,\xi\neq 0} \frac{|\hat{f}(\xi)-\hat{h}(\xi)|}{|\xi|},$$

This is a metric on the space of probability distributions with finite mean.

If we choose g to be the distribution of the particles in the thermostat and we pick $g \in L^2$ such that g is a probability distribution function with zero mean and finite second moment K_g then the master equation for the system we study is

$$\partial_t F_N = -\lambda N(I-Q)[F_N] - \mu\sum_{j=1}^{N}(I-R_j)[F_N] = \mathcal{L}[F_N], \tag{6.1}$$

where

$$Q[F_N] = \frac{1}{\binom{N}{2}}\sum_{i<j}\fint_0^{2\pi} F_N(v_{i,j}(\theta))\mathrm{d}\theta,$$

and

$$R_j[F_N] = \int \mathrm{d}w\fint_0^{2\pi} \mathrm{d}\theta g(w_j^*)F_N(v_j(w,\theta)).$$

In these

$$\begin{aligned} v_{ij}(\theta) &= (v_1,\dots,v_i\cos(\theta)+v_j\sin(\theta),\dots,-v_i\sin(\theta)+v_j\cos(\theta),\dots,v_N),\\ v_j(w,\theta) &= (v_1,\dots,v_j\cos(\theta)+w\sin(\theta),\dots,v_N),\\ w_j^* &= w\cos(\theta)-v_j\sin(\theta). \end{aligned}$$

We show that

Theorem 6.1. *A steady state for the master equation exists, is unique and has the same moments up to order 2 as $g^{\otimes N}$.*

Theorem 6.2. *If we start with initial data F_N^0 and H_N^0 which are probability distributions on $\mathbb{R}^N$*

with finite first and second moments then we have the following possible situations:

1. If F^0 and H^0 have the same mean initially then the GTW distance between the solutions is finite for all time and we get the exponential convergence:

$$d_{GTW,N}(F_N(t), H_N(t)) \leq e^{-\mu t/2} d_{GTW,N}(F_N^0, H_N^0).$$

2. If F^0 and H^0 have different means then we can construct an altered distance in which the solutions still converge exponentially fast towards each other with rate $\mu/2$. We also have the estimate

$$d_{T1,N}(F_N(t), H_N(t)) \leq e^{-\mu t/4} d_{T1,N}(F_N^0, H_N^0).$$

Remark. *The altered distance involves adding a correction term and is defined in order to deal with the fact that the GTW distance cannot deal with initial data with non-zero mean. If the two solutions initially have the same mean this reduces to the GTW distance. We give the theorem in both distances which shows we can either sacrifice something in the dependence on initial data or in the rate. In the asymptotic study as $N \to \infty$ the two distances give the same dependence on N through different mechanisms which suggests that the dependence on N occurring here is in some way intrinsic to the problem.*

Remark. *Here $\mu/2$ is the rate found in [26] to be the L^2 spectral gap and the rate of convergence to equilibrium in relative entropy.*

Furthermore we wish to study how the N particle Kac's model behaves as $N \to \infty$ in the manner originally proposed by Kac to link it with the spatially homogeneous Boltzmann equation. In order to do this we study how the convergence results which we have obtained can be translated into convergence results on the first marginal. We prove properties of the GTW metric which are similar to subadditivity. If the initial data $(F_N(0))_{N \geq 2}$ forms a chaotic family then we can control the convergence rate of the first marginals to equilibrium uniformly in N. We formally define the notion of chaotic family later. Similarly to [26] we can prove propagation of chaos in exactly the same manner as Kac in [85]. This means that the first marginals of the solution to the master equation will limit to the solution of a Boltzmann like equation. This motivates our proof of uniform in N convergence rates for the first marginal.

Theorem 6.3. *Suppose that f and h are mean zero probability densities on $\mathbb{R}$. If $(F_N(0,v))_{N \geq 2}$ and $(H_N(0,v))_{N \geq 2}$ are respectively f, h-chaotic families with respect to the Gabetta-Toscani-Wennberg metric. If furthermore, the distance between $F_N(0,\cdot)$ and $f^{\otimes N}$, and between $H_N(0,\cdot)$ and $h^{\otimes N}$ are bounded uniformly in N, and F_N, H_N are the solution to the N-particle coupled Kac's master equation with this initial data then there exists a constant C independent of N such that*

$$d_{GTW,1}(\Pi_1(F_N), \Pi_1(H_N)) \leq (C + d_{GTW,1}(f,h)) e^{-\frac{\mu}{2} t}.$$

Here we say that a family is f-chaotic with respect to a family of metrics, (d_k), if

$$d_k(\Pi_k[F_N], f^{\otimes N}) \to 0,$$

as $N \to 0$ for every k. Here d_k is a metric on $\mathbb{R}^k$ and Π_k is a projection onto this subspace of $\mathbb{R}^N$.

This is the standard notion of chaoticity which was introduced by Kac. Here we write it in terms of a distance which metrises weak convergence of measures as it is more convenient for our set up.

Remark. *Our theorem is really designed to work in the case of tensorised initial data and can be extended slightly as we have shown. If we no longer wanted our estimates to depend on the first marginal of the initial data we could replace the assumption of being close to tensorised initial data with the weaker, but difficult to check, condition*

$$d_N(F_N, H_N) \leq C \quad \forall N.$$

We also have two theorems in the case where we have non-zero and non equal mean for f and h using each of the different metrics which we use to study this case.

Theorem 6.4. *Let F_N^0 and H_N^0 are respectively f and h chaotic families where the GTW distance between F_N^0 and $f^{\otimes N}$ (resp. for H_N^0 and $h^{\otimes N}$) is bounded uniformly in N. Furthermore if f and h are probability densities with finite first and second moments, then we can construct an altered distance $\tilde{d}$ so that*

$$\tilde{d}\left(\Pi_1[F_N], \Pi_1[H_N]\right)) \leq (C_1 + (C_2 + C_3)\sqrt{N} + \tilde{d}(f,h))e^{-\frac{\mu}{2}t}.$$

Theorem 6.5. *Suppose that f and h are probability densities on $\mathbb{R}$ with finite mean. If $(F_N(0,v))_{N\geq 2}$ and $(H_N(0,v))_{N\geq 2}$ are respectively f, h-chaotic families with respect to the $T1$ metric, and the $T1$ distance between $F_N(0,\cdot)$ and $f^{\otimes N}$, and between $H_N(0,\cdot)$ and $h^{\otimes N}$ are bounded uniformly in N. Furthermore, let F_N, H_N are the solution to the N-particle coupled Kac's master equation with this initial data, then there exists a C (the bound between the initial data and the tensorised form) of N such that*

$$d_{T1,1}(\Pi_1[F_N](t), \Pi_1[H_N](t)) \leq (C + \sqrt{N} d_{T1,1}(f,h))e^{-\mu t/4}.$$

We can also prove two similar theorems in Wasserstein distance on measures with finite second moment. The Wasserstein distance is given by

$$\mathcal{W}_{2,d}(\mu,\nu) = \inf_{\pi} \left(\int_{\mathbb{R}^{2d}} \|\mathbf{x} - \mathbf{y}\|^2 \pi(\mathrm{d}\mathbf{x}, \mathrm{d}\mathbf{y}) \right)^{1/2},$$

here π ranges over measures with marginals μ, ν.

Theorem 6.6. *If μ_N and ν_N are two solutions to the master equation with finite second moments then*

$$\mathcal{W}_2(\mu_N(t), \nu_N(t)) \leq e^{-\mu t/2} \mathcal{W}_2(\mu_N(0), \nu_N(0)).$$

Theorem 6.7. *Suppose that $\mu_N(t)$ and $\nu_N(t)$ are solutions to the master equation at time t, with initial data $\mu_0^{\otimes N}$ and $\nu_0^{\otimes N}$ then we have that for any N,*

$$\mathcal{W}_{2,1}(\Pi_1(\mu_N(t)), \Pi_1(\nu_N(t))) \leq e^{-\mu t/2} \mathcal{W}_{2,1}(\mu_0, \nu_0).$$

Lastly we have a section where we look at Kac's model without coupling to a thermostat. Recently there has been a lot of work on Kac's model in GTW-distance. We have the result in

[112] which gives rates of convergence to equilibrium and shows that initially the decay in Kac's model in GTW can be very small. We also have the paper [25] which looks at Kac's model when a large number of particles are already in equilibrium and shows this can be uniformly approximated by the system coupled to a Gaussian thermostat. Convergence with uniform rates in N has been shown for the 1D Kac's model in Wasserstein-4 by Maxime Hauray [75]. We show that this result has an analogy in the GTW setting. The original goal of showing this was to see if reinterpreting the results in this form might allow us to extend this result to non-uniform collision kernels or to the 3D Kac's model.

Definition 6.1. *The seminorm ρ is given by*

$$\rho(\mu,\nu) = \sup_{\xi\neq 0, \Sigma\xi=0} \frac{|\mathcal{R}(\mu-\nu)(\xi)|}{|\xi|^2}.$$

Here μ and ν are two probability measures on $\mathbb{R}^N$ with the same second moment (energy) and finite fourth order moments. Also, $\mathcal{R}$ is a linear functional from the space of finite signed measures with finite second and fourth moments to the space of continuous functions which are differentiable up to order two at zero. This is given by

$$\mathcal{R}(\mu)(\xi) = \int \exp\left(-i\left(\xi_1 v_1^2 + \cdots + \xi_N v_N^2\right)\right)\mu(\mathrm{d}\mathbf{v}).$$

Remark. *This is in fact the GTW distance between the laws of the V_i^2.*

Theorem 6.8. *If $\mu(t)$ and $\nu(t)$ are two solutions to the Kac master equation that are supported on the sphere $\sqrt{N}\mathbb{S}^{N-1}$ then we have the following convergence*

$$\rho(\mu(t),\nu(t)) \leq \exp\left(-\left(1+\frac{1}{2(N-1)}\right)t\right)\rho(\mu,\nu).$$

Further we have that $\rho(\mu,\nu)=0$ iff the probability densities for the V_i^2 corresponding to μ and ν are the same. Further we show that solutions converge towards the set where the signs of the V_i are uniformly distributed on $\{-1,1\}$ with bound $2Ne^{-t}$. I.e. if S is the set of measures μ such that $\mu(\{sgn(V_i)=\sigma_i \forall i\}) = 1/(2^N)$ for every σ_i a vector of $+1s$ and $-1s$, then

$$\|\mu(t) - S\|_{TV} \leq 2Ne^{-t}.$$

6.2 Behaviour of the Moments

In this section we prove some basic lemmas on how the moments of a solution behave. We recall that K_g is the second moment of g our fixed distribution.

Lemma 6.1. *The kinetic energy of a solution to the coupled master equation converges exponentially fast to NK_g with rate $\mu/2$.*

Proof. Let

$$K(t) = \int_{\mathbb{R}^n} \|v\|^2 F_N(v)\mathrm{d}v.$$

Differentiating under the integral and recalling that radial functions are in the kernel of $(I-Q)$

and that $(I-Q)$ is self adjoint we get,

$$\partial_t K = \mu \sum_{j=1}^{N} \int_{\mathbb{R}^N} \mathrm{d}v \int \mathrm{d}w \fint_0^{2\pi} \mathrm{d}\theta g(w_j^*) F_N(v_j(w,\theta)) \|v\|^2 - \mu N K.$$

The Jacobian of the change of variables $(v_j(w,\theta), w_j^*) \leftrightarrow (v,w)$ is 1. Also we have that $\|v\|^2 + w^2 = \|v_j(w,\theta)\|^2 + w_j^{*2}$. Using these we have

$$\begin{aligned}
\partial_t K =& \mu \sum_{j=1}^{N} \int_{\mathbb{R}^N} \mathrm{d}v \int \mathrm{d}w \fint_0^{2\pi} \mathrm{d}\theta g(w) F_N(v)(\|v\|^2 + w^2) \\
&- \mu \sum_{j=1}^{N} \int_{\mathbb{R}^N} \mathrm{d}v \int \mathrm{d}w \fint_0^{2\pi} \mathrm{d}\theta g(w) F_N(v) w_j^{*2} - \mu N K, \\
=& \mu N K + \mu N K_g - \mu N K \\
&- \mu \sum_{j=1}^{N} \int_{\mathbb{R}^N} \mathrm{d}v \int \mathrm{d}w \fint_0^{2\pi} \mathrm{d}\theta g(w) F_N(v)(w^2 \cos^2\theta - 2 w v_j \cos\theta \sin\theta + v_j^2 \sin^2\theta), \\
=& \mu N K_g - \mu N \frac{1}{2} K_g - \frac{\mu}{2} K, \\
=& -\frac{\mu}{2}(K - N K_g).
\end{aligned}$$

□

Lemma 6.2. *The first moments of a solution to the coupled master equation converge to* 0 *with rate greater than* $\mu/2$. *Also the second order moments*

$$d_{k,l} = \int_{\mathbb{R}^N} F_N(v) v_k v_l \mathrm{d}v,$$

($k \neq l$) converge to 0 *with rate greater than* $\mu/2$.

Proof. Let $d_k = \int \mathrm{d}v F_N(v) v_k$ then we get the equation

$$\begin{aligned}
\partial_t d_k &= -N(\lambda + \mu) d_k + \lambda(N-2) d_k + \mu(N-1) d_k, \\
&= -(2\lambda + \mu) d_k.
\end{aligned}$$

For the second set we can calculate

$$\partial_t d_{k,l} = \left(-4\lambda - 2\mu + \frac{2\lambda}{N-1} \right) d_{k,l}$$

□

6.3 Existence, Uniqueness and Convergence to a Steady State

We wish to show existence and uniqueness of a steady state via the Banach fixed point theorem in the space of probability measures with zero mean and finite second moment with the GTW distance. In order to do this we write the steady state equation for F_N as a fixed point theorem.

We set $\gamma = \lambda/(\lambda+\mu)$ to mirror the notation in [39].

$$F_N = \gamma Q[F_N] + (1-\gamma)\frac{1}{N}\sum_{j=1}^{N} R_j[F_N] = \Phi[F_N].$$

We want to show that Φ is a contraction in the Gabetta-Toscani-Wennberg metric. We first need to show that Φ preserves the metric space that we are working in.

Lemma 6.3. *Suppose F_N has mean zero and finite second moment then $\Phi[F_N]$ has mean zero and finite second moment.*

Proof.

$$\begin{aligned}\int_{\mathbb{R}^N} Q[F_N]v_k \mathrm{d}v =& \frac{N-2}{N}\int_{\mathbb{R}^N} F_N(v)v_k\mathrm{d}v + \frac{1}{\binom{N}{2}}\sum_{i<k}\int_{\mathbb{R}^N}\fint_0^{2\pi} F_N(v)(v_i\cos\theta + v_k\sin\theta)\mathrm{d}\theta\mathrm{d}v\\ &+ \frac{1}{\binom{N}{2}}\sum_{k<j}\int_{\mathbb{R}^N}\fint_0^{2\pi} F_N(v)(-v_k\sin\theta + v_j\cos\theta)\mathrm{d}\theta\mathrm{d}v,\\ =& \frac{N-2}{N}\int_{\mathbb{R}^N} F_N(v)v_k\mathrm{d}v = 0.\end{aligned}$$

It is immediate that $\int R_j[F_N](v)v_k\mathrm{d}v = 0$ for $j \neq k$. So it remains to look at

$$\begin{aligned}\int_{\mathbb{R}^N}\mathrm{d}v R_k[F_N](v)v_k &= \int_{\mathbb{R}^N}\int \mathrm{d}w \fint_0^{2\pi}\mathrm{d}\theta g(w_j^*)F_N(v_j(w,\theta)v_k\\ &= \fint_0^{2\pi}\mathrm{d}\theta\int_{\mathbb{R}^N}\int \mathrm{d}v\mathrm{d}w g(w)F_N(v)(v_k\cos\theta - w\sin\theta) = 0.\end{aligned}$$

The fact that $\Phi[F_N]$ has finite second moments is clear since Q^*, R_j^* acting on $\|v\|^2$ or similar produces a finite linear combination of other functions to make second moments. □

Further we would like to calculate how Q and R_j act in Fourier space.

Lemma 6.4.

$$\widehat{Q[F_N]}(\xi) = \frac{1}{\binom{N}{2}}\sum_{k<j}\fint_0^{2\pi}\widehat{F_N}(\xi_{k,j})\mathrm{d}\theta,$$

where $\xi_{k,j} = (\xi_1,\dots,\xi_k\cos\theta + \xi_j\sin\theta,\dots,-\xi_k\sin\theta+\xi_j\cos\theta,\dots,\xi_N)$. Also,

$$\widehat{R_j[F_N]}(\xi) = \fint_0^{2\pi}\widehat{F_N}(\xi_j(\theta))\hat{g}(\xi_j\sin\theta)\mathrm{d}\theta,$$

where $\xi_j(\theta) = (\xi_1,\dots,\xi_j\cos\theta,\dots,\xi_N)$.

Proof.

$$\begin{aligned}\int_{\mathbb{R}^N} Q[F_N]e^{-iv\cdot\xi}\mathrm{d}v &= \frac{1}{\binom{N}{2}}\sum_{k<j}\fint_0^{2\pi}\mathrm{d}\theta\int_{\mathbb{R}^N}\mathrm{d}v F_M(v_{kj}(\theta))e^{-iv\cdot\xi},\\ &= \frac{1}{\binom{N}{2}}\sum_{k<j}\fint_0^{2\pi}\mathrm{d}\theta\int_{\mathbb{R}^N}\mathrm{d}v F_N(v)e^{-iv_{k,j}(\theta)\cdot\xi},\\ &= (2\pi)^{N/2}\frac{1}{\binom{N}{2}}\sum_{k<j}\fint_0^{2\pi}\mathrm{d}\theta\widehat{F_N}(\xi_{k,j}).\end{aligned}$$

Where $\xi_{k,j} = (\xi_1, \dots, \xi_k \cos\theta + \xi_j \sin\theta, \dots, -\xi_k \sin\theta + \xi_j \cos\theta, \dots, \xi_N)$.

$$\begin{aligned}\int_{\mathbb{R}^N} \mathrm{d}v R_j[F_N] e^{-iv\cdot\xi} &= \fint_0^{2\pi} \mathrm{d}\theta \int \mathrm{d}w \int_{\mathbb{R}^N} \mathrm{d}v g(w_j^*) F_N(v_j(w,\theta)) e^{-iv\cdot\xi} \\ &= \fint_0^{2\pi} \mathrm{d}\theta \int \mathrm{d}w \int_{\mathbb{R}^N} \mathrm{d}v g(w) F_N(v) e^{-iv_j(w,\theta)\cdot\xi} \\ &= (2\pi)^{N/2} \fint_0^{2\pi} \mathrm{d}\theta \widehat{F_N}(\xi_j(\theta)) \hat{g}(\xi_j \sin\theta).\end{aligned}$$

Where $\xi_j(\theta) = (\xi_1, \dots, \xi_j \cos\theta, \dots, \xi_N)$. □

Now we can show existence and uniqueness.

Proof of Theorem 6.1. Calculating we have

$$\widehat{\Phi[F_N]}(\xi) = \frac{1}{(2\pi)^{N/2}} \left(\gamma \int_{\mathbb{R}^N} Q[F_N](v) e^{-v\cdot\xi} \mathrm{d}v + (1-\gamma) \frac{1}{N} \sum_{j=1}^N \int_{\mathbb{R}^N} R_j[F_N] e^{-iv\cdot\xi} \mathrm{d}v \right).$$

Using the results of 6.4 we have

$$\widehat{\Phi[F_N]} = \fint_0^{2\pi} \mathrm{d}\theta \left(\frac{\gamma}{\binom{N}{2}} \sum_{i<j} \widehat{F_N}(\xi_{i,j}(\theta)) + \frac{1-\gamma}{N} \sum_{j=1}^N \widehat{F_N}(\xi_j(\theta)) \hat{g}(\xi_j \sin\theta) \right).$$

Therefore

$$\begin{aligned}&\sup_{\xi\neq 0} \frac{|\widehat{\Phi[F_N]}(\xi) - \widehat{\Phi[H_N]}(\xi)|}{|\xi|^2} \\ &\leq \sup_{\xi\neq 0} \frac{|\widehat{F_N}(\xi) - \widehat{H_N}(\xi)|}{|\xi|^2} \fint_0^{2\pi} \mathrm{d}\theta \left(\frac{\gamma}{\binom{N}{2}} \sum_{i<j} \frac{|\xi_{i,j}(\theta)|^2}{|\xi|^2} + \frac{1-\gamma}{N} \sum_{j=1}^N \hat{g}(\xi_j \sin\theta) \frac{|\xi_j(\theta)|^2}{|\xi|^2} \right) \\ &\leq \left(\gamma + \frac{1-\gamma}{N} \left(N - \frac{1}{2} \right) \right) d_{GTW}(F_N, H_N) \\ &\leq \left(1 - \frac{1-\gamma}{2N} \right) d_{GTW}(F_N, H_N).\end{aligned}$$

Here to go between the second and third line we used

$$\begin{aligned}\sum_{j=1}^N \hat{g}(\xi_j \sin\theta) \frac{|\xi_j(\theta)|^2}{|\xi|^2} &\leq \sum_{j=1}^N \frac{|\xi_j(\theta)|^2}{|\xi|^2} \\ &= \sum_{j=1}^N \frac{|\xi|^2 - \xi_j^2 \sin^2\theta}{\|\xi\|^2} \\ &= N - \sin^2\theta.\end{aligned}$$

So we have the required contraction property for any fixed N. Which shows existence and uniqueness of a steady state thanks to the contraction mapping theorem. The moments being the same up to order 2 as g follow from the lemmas on the behaviour of moments in the previous section. □

We also want to prove a contraction estimate in the $T1$ distance.

Lemma 6.5.
$$d_{T1,N}(\Phi[F_N],\Phi[H_N]) \le \left(1-\frac{1-\gamma}{4N}\right) d_{T1,N}(F_N,H_N).$$

Proof. The proof is the same as for the GTW distance but here it is necessary to use

$$(1-x^2)^{1/2} \le 1-\frac{1}{2}x^2,$$

when bounding $|\xi_j(\theta)|/|\xi|$. This time we have

$$\begin{aligned}\sum_{j=1}^{N}\hat{g}(\xi_j\sin\theta)\frac{|\xi_j(\theta)|}{|\xi|} &\le \sum_{j=1}^{N}\sqrt{\frac{|\xi|^2-\xi_j^2\sin^2\theta}{|\xi|^2}}\\ &\le \sum_{j=1}^{n}\left(1-\frac{1}{2}\frac{\xi_j^2\sin^2\theta}{|\xi|^2}\right)\\ &= N-\frac{1}{2}\sin^2\theta.\end{aligned}$$

□

Using these estimates we can also show convergence to equilibrium.

Proof of Theorem 6.2. Suppose initially that $F_N(t)$ and $H_N(t)$ both have zero mean. From the above calculation we have

$$F_N(t+s)-H_N(t+s) = (1-s(\lambda+\mu)N)(F_N(t)-H_N(t))+s(\lambda+\mu)N(\Phi[F_N(t)]-\Phi[H_N(t)])+o(s).$$

Therefore

$$\begin{aligned}d_{GTW}(F_N(t+s),H_N(t+s)) \le&(1-s(\lambda+\mu)N)d_{GTW}(F_N(t),H_N(t))\\ &+s(\lambda+\mu)Nd_{GTW}(\Phi[F_N],\Phi[H_N])+o(s)\\ \le&(1-s(\lambda+\mu))d_{GTW}(F_N(t),H_N(t))\\ &+s(\lambda+\mu)N\left(1-\frac{1-\gamma}{2N}\right)d_{GTW}(F_N(t),H_N(t))+o(s)\\ =&\left(1-\frac{\mu}{2}s\right)d_{GTW}(F_N(t),H_N(t))+o(s).\end{aligned}$$

Hence,
$$\frac{\mathrm{d}}{\mathrm{d}t}d_{GTW}(F_N(t),H_N(t)) \le -\frac{\mu}{2}d_{GTW}(F_N(t),H_N(t)).$$
So that we have exponential decrease with the stated rate. Since in 6.2 we showed that if we start the dynamics with two distribution which have zero mean then this property will be preserved, we see that if we start the dynamics with a zero mean distribution then it will converge exponentially fast towards the steady state. Now we would like to add a correction term so that we can deal with a wider class of initial data as in [39]. We define

$$\widehat{\mathcal{M}[F_N]} := \chi(\xi)\sum_{k=1}^{N}\left(\int_{\mathbb{R}^N}v_kF_N(v)\mathrm{d}v\right)i\xi_k,$$

where χ is a smooth, compactly supported function which is 1 in some neighbourhood of 0. Therefore, if $D_N = F_N - H_N - \mathcal{M}[F_N - H_N]$ we will have that

$$\widehat{D_N} = \int_{\mathbb{R}^N} \mathrm{d}v \, (F_N(v) - H_N(v)) \left(e^{-iv\cdot\xi} - \chi(\xi) \sum_{j=1}^N v_j \xi_j \right).$$

This means that

$$\sup_{\xi \neq 0} \frac{\widehat{D_N}(\xi)}{|\xi|^2} < \infty.$$

We calculate that

$$\begin{aligned}
\partial_t D_N =& \partial_t F_N - \partial_t H_N - \partial_t \mathcal{M}[F_N - H_N] \\
=& \lambda N (I - Q)[D_N] - \mu \sum_{j=1}^N (I - R_j)[D_N] \\
& - \lambda (I - Q)[\mathcal{M}[F_N - H_N]] - \mu \sum_{j=1}^N (I - R_j)[\mathcal{M}[F_N - H_N]] - \partial_t \mathcal{M}[F_N - H_N].
\end{aligned}$$

So if we let

$$W = -\lambda N (I - Q)[\mathcal{M}[F_N - H_N]] - \mu \sum_{j=1}^N (I - R_j)[\mathcal{M}[F_N - H_N]] - \partial_t \mathcal{M}[F_N - H_N],$$

then D_N is a zero momentum, zero integral function and we have the equation

$$\partial_t D_N = -(\lambda + \mu) N (D_N - \Phi[D_N]) + W.$$

So if we want to show that

$$\sup_{\xi \neq 0} \frac{|\widehat{D_N}|}{|\xi|^2},$$

converges to zero exponentially fast it is sufficient to show that,

$$\sup_{\xi \neq 0} \frac{|\widehat{W}(\xi)|}{|\xi|^2},$$

converges to zero exponentially fast. Since ∂_t commutes with Fourier transform and χ is compactly supported we know that

$$\widehat{\mathcal{M}}[F_N - H_N] = \chi(\xi) \sum_{k=1}^N (m_f(0) - m_h(0)) e^{-(2\lambda + \mu)t} i\xi_k,$$

So ignoring χ and looking near 0 we have, after Taylor expanding and using the formula from lemma 6.4

$$
\begin{aligned}
&-\lambda N\widehat{(I-Q)}[\mathcal{M}]-\mu\sum_{j=1}^{N}\widehat{(I-R_j)}[\mathcal{M}]=\\
&-(2\lambda+\mu)(m_f(0)-m_h(0))e^{-(2\lambda+\mu)t}\sum_{k=1}^{N}\xi_k\\
&-\frac{1}{2}\mu K_g(m_f(0)-m_h(0))e^{-(2\lambda+\mu)t}|\xi|^2\sum_{k=1}^{N}\xi_k+o(|\xi|^3).
\end{aligned}
$$

Therefore near $\xi=0$, we have

$$
\frac{\widehat{W}(\xi)}{|\xi|^2}=-\frac{1}{2}\mu K_g\sum_{k=1}^{N}\xi_k+\frac{1}{2}\mu K_g\frac{\sum_{k=1}^{N}\xi_k^3}{|\xi|^2}+o(\xi).
$$

This is because the lower order terms cancel. So in particular we have that

$$
\lim_{\xi\to 0}\frac{\widehat{W}(\xi)}{|\xi|^2}=0.
$$

Therefore, since $\widehat{W}$ has compact support we can bound

$$
\frac{\widehat{W}(\xi)}{|\xi|^2}\leq Ce^{-(2\lambda+\mu)t}
$$

where C may increase with N. At 0 the gradient of

$$
w(\xi)=\frac{\hat{W}(\xi)}{|\xi|^2}
$$

is $C\sqrt{N}\mu K_g/2$ so the gradient of w cannot be bounded uniformly in N. Since we can calculate $w(\xi)$ explicitly if χ is always radial as

$$
\mu\left(1-\sum_{j=1}^{N}(1-\alpha_j(\xi))\right)\frac{\mathcal{M}}{|\xi|^2}
$$

where

$$
\alpha_j(\xi)=\int_0^{2\pi}\left(1-\frac{\xi_j(1-\cos\theta)}{\sum_k\xi_k}\right)\hat{g}(\xi_j\sin\theta)\frac{\chi(\xi_j(\theta))}{\chi(\xi)}\mathrm{d}\theta.
$$

This can be bounded uniformly provided we can bound the ration of the χs. Therefore under these additional assumptions we see that w increases no faster than $\sqrt{N}$. This will give that

$$
\sup_{\xi\neq 0}\frac{|\widehat{D_N(t)}|}{|\xi|^2}\leq\left(C\sqrt{N}+\frac{|\widehat{D_N}(0)|}{|\xi|^2}\right)e^{-\frac{\mu}{2}t}.
$$

Therefore if we define a new distance

$$
\tilde{d}_N(F_N,H_N)=\sup_{\xi\neq 0}\frac{|\widehat{D_N}|}{|\xi|^2}+\sup_{\xi\neq 0}\frac{|\widehat{W}|}{|\xi|^2},
$$

we will get the inequality

$$\tilde{d}_N(F_N(t), H_N(t)) \leq Ce^{-\frac{\mu}{2}t}.$$

For the exponential convergence in the $T1$ distance we use the same argument as for the GTW distance with the same mean and the contraction estimate in Lemma 6.5. □

Remark. *If it were possible to get a bound on $|\nabla w(\xi)|$ in terms of $\sqrt{N}$ then it might in fact allow us to choose χ for each N such that we didn't get the increase with N by letting the radius of the support of χ decrease with $\sqrt{N}$. However, since the goal is to control the behaviour as $N \to \infty$ then in the case of different marginals working with the correction term would introduce an error of at least $\sqrt{N}$ when trying to control the initial data by its first marginal. In general because of having to choose a χ for each N the altered distance is not well adapted to asymptotic analysis. We include it to show that for each N we can get the rate $\mu/2$ and to compare with the limit equation case which is studied using this method in [39].*

6.4 Convergence Rate of the First Marginal

It is shown in [26] that propagation of chaos holds for this type of coupled Kac's model. The argument is very similar to Kac's original argument therefore is not repeated here. Since we have propagation of chaos we know that the first marginal of $F_N(t)$ will converge weakly towards a solution of the Boltzmann-Kac equation. In some sense we would like to be able to understand the two limits $t \to \infty$ and $N \to \infty$ simultaneously. For this reason we prove a bound on convergence to equilibrium for the first marginal which is uniform in N. Unfortunately, the GTW distance and our correction term W behave differently as $N \to \infty$ so it was only possible to get these estimates when the initial data has zero mean.

The functions we work with will be invariant under permutations of variables so we can define the k^{th} marginal for $k \leq N$

$$\Pi_k[F_N] := \int_{\mathbb{R}^{N-k}} F_N(v_1, \dots, v_N) \mathrm{d}v_{i_1} \dots \mathrm{d}v_{i_{N-k}}$$

for any choice of $1 \leq i_1 < i_2 < \cdots < i_{N-k} \leq N$. Many of the distances in which we could study Kac's model, typically weighted L^2 distances will not behave well as the number of particles tends to infinity so will not give convergence of the first marginal to an equilibrium in entropy, here the subadditivity property of entropy in the number of variables is crucial. We wish to show that the GTW and related distances will possess similar subadditivity properties, which will allow us to control things in a similar way.

Lemma 6.6.

$$d_{GTW,k}(\Pi_k[F_N], \Pi_k[H_N]) \leq d_{GTW,N}(F_N, H_N),$$

$$\tilde{d}_k(\Pi_k[F_N], \Pi_k[H_N]) \leq \tilde{d}_k(F_N, H_N),$$

and

$$d_{T1,k}(\Pi_k[F_N], \Pi_k[H_N]) \leq d_{T1,N}(F_N, H_N).$$

Proof. The proof is the same for all the distances so we only do it in the case of GTW. We can notice that

$$\widehat{\Pi_k[F_N]}(\xi_1, \ldots, \xi_k) = \widehat{F_N}(\xi_1, \ldots, \xi_k, 0, \ldots, 0).$$

Using this we have that

$$\begin{aligned} d_{GTW,k}(\Pi_k[F_N], \Pi_k[H_N]) = & \sup_{\xi \neq 0, \xi_{k+1} = \cdots = \xi_N = 0} \frac{|\widehat{F_N}(\xi) - \widehat{H_N}(\xi)|}{|\xi|^2} \\ \leq & \tilde{d}(F_N, H_N). \end{aligned}$$

□

Lemma 6.7. *If f, h have the same first moments*

$$d_{GTW,N}(f^{\otimes N}, h^{\otimes N}) = d_{GTW,1}(f, h)$$

where $d_{GTW,k}$ is the GTW distance on probability densities with k-variables.

Proof.

$$\begin{aligned} d_{GTW}(f^{\otimes N}, h^{\otimes N}) = & \sup_{\xi \neq 0} \frac{|\hat{f}(\xi_1) \ldots \hat{f}(\xi_N) - \hat{h}(\xi_1) \ldots \hat{h}(\xi_N)|}{|\xi|^2} \\ \leq & \sup_{\xi \neq 0} \frac{\sum_{i=1}^{N} |\hat{f}(\xi_1) \ldots \hat{f}(\xi_{i-1})(\hat{f}(\xi_i) - \hat{h}(\xi_i))\hat{h}(\xi_{i+1}) \ldots \hat{h}(\xi_N)|}{|\xi|^2} \\ \leq & \sup_{\xi \neq 0} \sum_{i=1}^{N} \frac{\hat{f}(\xi_i) - \hat{h}(\xi_i)}{\xi_i^2} \frac{\xi_i^2}{|\xi|^2} \\ \leq & \sup_{\xi \neq 0} \sum_{i=1}^{N} d_{GTW,1}(f, h) \frac{\xi_i^2}{|\xi|^2} = d_{GTW,1}(f, h). \end{aligned}$$

Since f, h are the first marginals of $f^{\otimes N}, h^{\otimes N}$ respectively we have by the earlier lemma that

$$d_{GTW,1}(f, h) \leq d_{GTW,N}(f^{\otimes N}, h^{\otimes N})$$

putting the two inequalities together gives the required result. □

We have already seen that

$$\frac{\widehat{W}(\xi)}{|\xi|^2},$$

may increase with N so this will cause us problems if we wished to try and control $\tilde{d}_N(f^{\otimes N}, h^{\otimes N})$ by $\tilde{d}_1(f, h)$. Even given this it would be good to be able to push the control by first marginals to general functions. However, the next lemma shows that this is not possible.

Lemma 6.8. *There exist f, g with finite second moment such that f, g are symmetric and mean zero and they have the same marginals but f, g are not the same. This means we cannot control the GTW distance between f and g in terms of the GTW distance between their first marginals.*

Proof. Let ϕ be a density function on $\mathbb{R}$ which is mean zero but not even. Define

$$f(v_1, v_2) := \frac{1}{2}(\phi(v_1)\phi(-v_2) + \phi(-v_1)\phi(v_2)),$$

and

$$g(v_1, v_2) = \frac{1}{2}(\phi(v_1)\phi(v_2) + \phi(-v_1)\phi(-v_2)).$$

Then it is easy to see that f and g have the required properties. $\square$

We wish to combine these lemmas in such a way as to get uniform control on the first marginal. Given the restriction shown by Lemma 6.8 we want to choose 'good' initial data in order that the distance between the initial data is controlled by the distance between the first marginals.

Proof of Theorem 6.3. Since f, h have mean zero and the GTW distance between $F_N(0)$ and $f^{\otimes N}$ is finite, we have that F_N and H_N have zero mean initially. By 6.2 this holds for all time. Therefore we have by Lemma 6.6

$$d_{GTW,1}(\Pi_1[F_N], \Pi_1[H_N]) \leq d_{GTW,N}(F_N, H_N).$$

Furthermore, by Theorem 6.2

$$d_{GTW,N}(F_N(t), H_N(t)) \leq d_{GTW,N}(F_N(0), H_N(0))e^{-\frac{\mu}{2}t}.$$

Now we use the chaoticity property and our control on tensorised functions form Lemma 6.7 to get

$$\begin{aligned} d_{GTW,N}(F_N(0), H_N(0)) &\leq d_{GTW,N}(F_N(0), f^{\otimes N}) + d_{GTW,N}(f^{\otimes N}, h^{\otimes N}) + d_{GTW,N}(h^{\otimes N}, H_N(0)) \\ &= C_1 + d_{GTW,1}(f, h). \end{aligned}$$

Here C_1 only depends on how close the initial data is to tensorised. Putting this together gives

$$d_{GTW,1}(\Pi_1[F_N](t), \Pi_1[H_N](t)) \leq (d_{GTW,1}(f, h) + C_1)e^{-\frac{\mu}{2}t}.$$

We do not have from our conditions that C_1 will decrease to 0 as $N \to \infty$, but since in this situation the real interest is just to choose any f-chaotic family we may as well have that $F_N(0) = f^{\otimes N}$ and similarly with H which would dispense with the C_1 altogether. $\square$

Now we would like to prove a theorem in the spirit of Theorem 6.3 when we do not have f and h having zero mean initially. We cannot recover uniform estimates in N but we can control the growth with N. We have from lemma 6.6 control of marginals by the function for the $\tilde{d}$ distance so we have

$$\tilde{d}_k(\Pi_k[F_N], \Pi_k[H_N]) \leq \tilde{d}(F_N, H_N).$$

Following this we would like to prove something in the spirit of lemma 6.7 in order to control in the other direction.

Lemma 6.9. *Suppose we have f and h probability distributions on $\mathbb{R}$ with differentiable Fourier transforms. If we define*

$$n_f = \int |v| f(v) \mathrm{d}v,$$

and let $M = \max\left\{\frac{n_f}{|m_f|}, \frac{n_h}{|m_h|}\right\}$ then we have the following control by the first marginals for the $\tilde{d}$ distance on tensorised functions.

$$\tilde{d}_N(f^{\otimes N}, h^{\otimes N}) \leq \tilde{d}_1(f, h) + M|m_f - m_h|\sqrt{N}.$$

Proof. Using the same bridging argument as before we see that

$$\begin{aligned}
&\hat{f}(\xi_1)\dots\hat{f}(\xi_N)-\hat{h}(\xi_1)\dots\hat{h}(\xi_N)-(m_f-m_h)\chi_N(\xi)\sum_k i\xi_k\\
&=\sum_k \hat{f}(\xi_1)\dots\hat{f}(\xi_{k-1})(\hat{f}(\xi_k)-\hat{h}(\xi_k)-\chi_1(\xi_k)(m_f-m_h)i\xi_k)\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)\\
&+\sum_k \hat{f}(\xi_1)\dots\hat{f}(\xi_{k-1})(m_f-m_h)\chi_1(\xi_k)i\xi_k\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)-\chi_N(\xi)\sum_k(m_f-m_h)i\xi_k.
\end{aligned}$$

In order to complete the proof we want to bound the last term by something of the form

$$M|m_f-m_h|\sqrt{N}|\xi|^2.$$

Provided the radius of the set in which the χ are 1 is sufficiently large this will be true. So if we look at the last term where the χ are 1, we have

$$(m_f-m_h)i\sum_k \xi_k\left(\hat{f}(\xi_1)\dots\hat{f}(\xi_{k-1})\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_k)-1\right).$$

If instead we try and bound

$$A=\frac{\hat{f}(\xi_1)\dots\hat{f}(\xi_{k-1})\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)-1}{m_f\sum_{j<k}i\xi_j+m_h\sum_{k<j}i\xi_j}\le M$$

then we would have the bound

$$\begin{aligned}
&\left|\frac{\sum_k(\hat{f}(\xi_1)\dots\hat{f}(\xi_{k-1})\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)-1)\xi_k(m_f-m_h)}{|\xi|^2}\right|\\
&\le M\frac{\left|\sum_{k=1}^N(m_f\sum_{j<k}i\xi_j+m_h\sum_{k<j}i\xi_j)\xi_k i(m_f-m_h)\right|}{|\xi|^2}\le M|m_f-m_h|\sqrt{N}.
\end{aligned}$$

Therefore it remains to prove the bound on A, we do this first by noting that by Taylor expanding we can see that as $|\xi|\to 0, A\to 1$ and that as $|\xi|\to\infty, A\to 0$. A is differentiable everywhere except possibly 0. Now we differentiate to get that at any stationary point of A and for every $l<k$ we have

$$\begin{aligned}
&\hat{f}(\xi_1)\dots\hat{f}'(\xi_l)\dots\hat{f}(\xi_{k-1})\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)\left(m_f\sum_{j<k}i\xi_j+m_h\sum_{k<j}i\xi_j\right)\\
&=im_f\left(\hat{f}(\xi_1)\dots\hat{f}(\xi_{k-1})\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)-1\right).
\end{aligned}$$

Substituting this into our expression for A shows that at a stationary point

$$A=\frac{1}{im_f}\hat{f}(\xi_1)\dots\hat{f}'(\xi_l)\dots\hat{f}(\xi_{k-1})\hat{h}(\xi_{k+1})\dots\hat{h}(\xi_N)\le M.$$

This gives the claimed bound. It seems like there will be a problem if $m_f=0$ but if so we can always choose to differentiate in a direction so that we will get m_h rather than m_f and the cannot both be 0. Here C_1, in the statement, only depends on the distance between the initial data and the tensorised functions, C_2 only depends on g and χ and C_3 is a constant times $M|m_f-m_h|$

where M is the maximum of $\int |v|f(v)\mathrm{d}v$ with the same quantity for h. ☐

We can now prove the theorem

Proof of Theorem 6.4. This is found by putting together the convergence theorems and lemmas on distance control in exactly the same way as Theorem 6.2. ☐

If we move on to looking at the $T1$ distance we again have the bound on the $T1$ distance between marginals by the distance between the full function from Lemma 6.6. We would like to be able to control the distance between tensorised functions by the marginals in order to give similar arguments to Theorem 6.3 and Theorem 6.4.

Lemma 6.10.

$$d_{T1,N}(f^{\otimes N}, h^{\otimes N}) \leq \sqrt{N} d_{T1,1}(f,h).$$

Furthermore, the square root dependence is the best possible if f, h have different means.

Proof. This follows a similar argument to the others

$$\begin{aligned}
\sup_{\xi \neq 0} \frac{|\hat{f}(\xi_1)\ldots\hat{f}(\xi_N) - \hat{h}(\xi_1)\ldots\hat{h}(\xi_N)|}{|\xi|} &\leq \sup_{\xi \neq 0} \frac{\sum_k |\hat{f}(\xi_1)\ldots\hat{f}(\xi_{k-1}(\hat{f}(\xi_k) - \hat{h}(\xi_k))\hat{h}(\xi_{k+1})\ldots\hat{h}(\xi_N)|}{|\xi|} \\
&\leq \sup_{\xi \neq 0} \sum_k \frac{|\hat{f}(\xi_k) - \hat{h}(\xi_k)|}{|\xi_k|} \frac{|\xi_k|}{|\xi|} \\
&\leq sup_{\xi \neq 0} \frac{|\hat{f}(\xi) - \hat{h}(\xi)|}{|\xi|} \sum_k \frac{|\xi_k|}{|\xi|} \\
&\leq \sqrt{N} \sup_{\xi \neq 0} \frac{|\hat{f}(\xi) - \hat{h}(\xi)|}{|\xi|}.
\end{aligned}$$

The fact that the square root dependence is necessary for functions with different means can be seen by Taylor expanding

$$\frac{\hat{f}(\xi_1)\ldots\hat{f}(\xi_N) - \hat{h}(\xi_1)\ldots\hat{h}(\xi_N)}{|\xi|}$$

around $\xi = 0$ then we can see that the limit as $\xi \to 0$ of this expression has modulus $\sqrt{N}|m_f - m_h|$. ☐

Proof of Theorem 6.5. Again we combine the convergence theorem that we have for the $T1$ distance with the control on distances as in Theorem 6.2. ☐

6.5 Contraction in Wasserstein-2

We can also show contraction of this model in Wasserstein distances using a simple coupling of two different systems. This coupling involves taking two of the coupled Kac's models and giving them simultaneous collisions with the same angle if it is an internal collision and the same angle and velocity of the external particle if it is an external collision. We can represent the stochastic process as an integral against several Poisson point processes. This is done in [75] and is helpful here to prove contraction for the energy process in Kac's model.

$$V_{i,t} = V_{i,0} + \lambda \sum_{j \neq i} \int_0^t \int_0^{2\pi} \left(V_{i,s^-} \cos\theta + V_{j,s^-} \sin\theta - V_{i,s^-} \right) \Pi_{i,j}(\mathrm{d}s, \mathrm{d}\theta) \tag{6.2}$$

$$+ 2\mu \int_0^t \int_{-\infty}^{\infty} \int_0^{2\pi} \left(V_{i,s^-} \cos\theta + w \sin\theta - V_{i,s^-} \right) \nu_i(\mathrm{d}s, \mathrm{d}w, \mathrm{d}\theta). \tag{6.3}$$

Here $\Pi_{i,j}$ is a Poisson point process on $[0, \infty) \times [0, 2\pi]$ with intensity measure being $1/2\pi(N-1)$ times Lebesgue measure, and ν_i is a Poisson point process with intensity measure g tensored with $1/2\pi(N-1)$ times Lebesgue. Using this representation we can prove contraction in Wasserstein-2.

Proof of Theorem 6.6. Using the representation above we can write out a similar formula for the difference between two solutions coupled by giving them the same driving Poisson processes. If we call this difference in the i^{th} variable $\Delta_{i,t}$ then we can write

$$\Delta_{i,t}^2 = \Delta_{i,0}^2 + \lambda \sum_{j \neq i} \int_0^t \int_0^{2\pi} \left(\Delta_{i,s^-}^2 (\cos^2\theta - 1) + \Delta_{j,s^-}^2 \sin^2\theta + 2\cos\theta \sin\theta \Delta_{i,s^-} \Delta_{j,s^-} \right) \Pi_{i,j}(\mathrm{d}s, \mathrm{d}\theta)$$

$$+ 2\mu \int_0^t \int_{-\infty}^{\infty} \int_0^{2\pi} \left(\Delta_{i,s^-}^2 (cos^2\theta - 1) + 2\Delta_{i,s^-} w \sin\theta \cos\theta \right) \nu(\mathrm{d}s, \mathrm{d}w, \mathrm{d}\theta).$$

Summing over i and taking expectations gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left(\sum_{i=1}^n \Delta_{i,t}^2 \right) =& 2\lambda(N-1) \frac{1}{2\pi} \int_0^{2\pi} (\cos^2\theta + \sin^2\theta - 1) \mathrm{d}\theta \mathbb{E}\left(\sum_{i=1}^n \Delta_{i,t}^2 \right) \\ &+ 2\mu \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} g(w)(\cos^2\theta - 1) \mathrm{d}\theta \mathrm{d}w \mathbb{E}\left(\sum_{i=1}^n \Delta_{i,t}^2 \right) \\ =& -\mu \mathbb{E}\left(\sum_{i=1}^n \Delta_{i,t}^2 \right). \end{aligned}$$

Which gives the result after taking the infimum over possible couplings. □

We can also prove a similar controls over how Wasserstein distances behave in as the dimension goes to infinity. Here we write $\mathcal{W}_{p,d}$ to be the Wasserstein-2 distance related to the euclidean distance on $\mathbb{R}^d$.

Lemma 6.11. *If μ, ν are measures on $\mathbb{R}$ with finite second moment then*

$$\mathcal{W}_{2,N}(\mu^{\otimes N}, \nu^{\otimes N}) = \sqrt{N} \mathcal{W}_{2,1}(\mu, \nu).$$

Proof. We know that there exists an optimal coupling, π_1 so that

$$\mathcal{W}_{2,1}(\mu, \nu) = \left(\int_{\mathbb{R}^2} (x-y)^2 \pi_1(\mathrm{d}x, \mathrm{d}y) \right)^{1/2}$$

and an optimal coupling, π_N, such that

$$\mathcal{W}_{2,N}(\mu^{\otimes N}, \nu^{\otimes N}) \left(\int_{\mathbb{R}^{2N}} \|\mathbf{x} - \mathbf{y}\|^2 \pi_N(\mathrm{d}\mathbf{x}, \mathrm{d}\mathbf{y}) \right)^{1/2}.$$

Suppose that $\pi_N \neq \pi_1^{\otimes N}$ then we have that

$$\begin{aligned}
&\int \left((x_1-y_1)^2+\cdots+(x_N-y_N)^2\right)\pi_N(\mathrm{d}\mathbf{x},\mathrm{d}\mathbf{y}) \\
&< \int \left((x_1-y_1)^2+\cdots+(x_N-y_N)^2\right)\pi_1(\mathrm{d}x_1,\mathrm{d}y_1)\dots\pi_1(\mathrm{d}x_N,\mathrm{d}y_N) \\
=&N\int (x-y)^2\pi_1(\mathrm{d}x,\mathrm{d}y).
\end{aligned}$$

Therefore, there exists some k such that

$$\int_{\mathbb{R}^{2N}} (x_k-y_k)^2\pi_N(\mathrm{d}\mathbf{x},\mathrm{d}\mathbf{y}) < \int_{\mathbb{R}^2}(x-y)^2\pi_1(\mathrm{d}x,\mathrm{d}y).$$

Since the integrand on the left hand side only depends on x_k, y_k π_N induces a coupling of μ and ν by projection onto the k^{th} variables. The cost under this measure is strictly less that the optimal cost which is a contradiction. Hence, the optimal coupling is achieved by $\pi_1^{\otimes N}$. This gives that,

$$\begin{aligned}
\mathcal{W}_{2,N}(\mu^{\otimes N},\nu^{\otimes N}) &= \left(\int \left((x_1-y_1)^2+\cdots+(x_N-y_N)^2\right)\pi_1(\mathrm{d}x_1,\mathrm{d}y_1)\dots\pi_1(\mathrm{d}x_N,\mathrm{d}y_N)\right)^{1/2} \\
&= \left(N\int (x-y)^2\pi_1(\mathrm{d}x,\mathrm{d}y)\right)^{1/2} \\
&= \sqrt{N}\mathcal{W}_{2,1}(\mu,\nu).
\end{aligned}$$

□

Lemma 6.12. *If μ_N and ν_N are symmetric probability distributions on $\mathbb{R}^N$ with finite second moment then*

$$\mathcal{W}_{2,1}(\Pi_1(\mu_N),\Pi_1(\nu_N)) \leq \frac{1}{\sqrt{N}}\mathcal{W}_{2,N}(\mu_N,\nu_N).$$

Proof. Suppose that π_N is a coupling of μ_N and ν_N then the marginals of π_N induce couplings of the marginals of μ_N and ν_N.

$$\begin{aligned}
&\left(\int \left((x_1-y_1)^2+\cdots+(x_N-y_N)^2\right)\pi_N(\mathrm{d}\mathbf{x},\mathrm{d}\mathbf{y})\right)^{1/2} \\
=&\left(\int (x_1-y_1)^2\pi_N(\mathrm{d}\mathbf{x},\mathrm{d}\mathbf{y})+\cdots+\int (x_N-y_N)^2\pi_N(\mathrm{d}\mathbf{x},\mathrm{d}\mathbf{y})\right)^{1/2} \\
\geq&\left(N\mathcal{W}_{2,1}(\Pi_1(\mu_N),\Pi_1(\nu_N))^2\right)^{1/2} = \sqrt{N}\mathcal{W}_{2,1}(\Pi_1(\mu_N),\Pi(\nu_N)).
\end{aligned}$$

□

Like with the earlier sections we can combine this behaviour with our contraction estimates to show uniform behaviour of the first marginal. For simplicity we only looked at tensorised initial data.

Proof of Theorem 6.7.

$$\begin{aligned}\mathcal{W}_{2,1}(\Pi_1(\mu_N(t)),\Pi_1(\nu_N(t))) &\le \frac{1}{\sqrt{N}}\mathcal{W}_{2,N}(\mu_N(t),\nu_N(t))\\ &\le \frac{1}{\sqrt{N}}e^{-\mu t/2}\mathcal{W}_{2,N}(\mu_0^{\otimes N},\nu_0^{\otimes N})\\ &= e^{-\mu t/2}\mathcal{W}_{2,1}(\mu_0,\nu_0).\end{aligned}$$

□

Remark. *These uniform estimates in N combined with propagation of chaos means that the limit Boltzmann-Kac equation will also show exponential convergence to equilibrium in Wasserstein-2. This is very similar to the result shown in [39] in the Toscani distance.*

6.6 Kac's Model in the GTW-like Metric

Lastly we look the uncoupled Kac's model in the GTW-metric. We translate the results of Hauray in [75] from a Wasserstein-2 result into a GTW result. We study the behaviour of this model under a semi-norm. We wish to show contraction in this seminorm to the set on which it is zero, between any two solutions to the Kac master equation. In order to do this we need to look at

$$\rho(Q\mu, Q\nu),$$

where ρ is this seminorm to be defined shortly. Let us work in the case where μ and ν are absolutely continuous with respect to the indicator function of the sphere of radius $\sqrt{N}$. Here we can write out how Q acts on the measures explicitly. Furthermore observe that here the restriction in the range of ξ that $\Sigma\xi = 0$ will make no difference to the supremum. Let us write F and G for the Radon-Nikodym derivatives of μ and ν respectively against the uniform measure on the Kac sphere. Then since Q does not act on the uniform measure on the Kac sphere, we write

$$Q[\mu] = Q[F] = \frac{1}{\binom{N}{2}}\sum_{i<j}\frac{1}{2\pi}\int_0^{2\pi} F(v_{ij}(\theta))\mathrm{d}\theta,$$

where

$$v_{ij}(\theta) = (v_1,\dots,v_i\cos\theta + v_j\sin\theta,\dots,-v_i\sin\theta+v_j\cos\theta,\dots,v_N).$$

Lemma 6.13. *We have that*

$$R[Q[F]](\xi) = \frac{1}{\binom{N}{2}}\sum_{i<j}\frac{1}{2\pi}\int_0^{2\pi} R[F](\xi_{i,j}(\theta))\mathrm{d}\theta.$$

Where

$$\xi_{i,j}(\theta) = (\xi_1,\dots,\xi_i\cos^2\theta+\xi_j\sin^2\theta,\dots,\xi_i\cos^2\theta+\xi_j\sin^2\theta,\dots,\xi_N).$$

Proof. We calculate

$$\begin{aligned}
\mathcal{R}\left[Q[F]\right](\xi) &= \int_{\mathbb{R}^N} \frac{1}{2\pi\binom{N}{2}} \sum_{i<j} \int_0^{2\pi} F(v_{ij}(\theta)) \exp\left(-i\left(\xi_1 v_1^2 + \cdots + \xi_n v_N^2\right)\right) \mathrm{d}v\mathrm{d}\theta, \\
&= \int_{\mathbb{R}^N} \frac{1}{2\pi\binom{N}{2}} \sum_{i<j} \int_0^{2\pi} F(v) \exp(-i(\xi_1 v_1^2 + \cdots + \xi_i(v_i\cos\theta + v_j \sin\theta)^2 + \\
&\cdots + \xi_j(-v_i \sin\theta + v_j\cos\theta)^2 + \cdots + \xi_n v_N^2))\mathrm{d}v\mathrm{d}\theta \\
&= \int_{\mathbb{R}^N} \frac{1}{2\pi\binom{N}{2}} \sum_{i<j} \int_0^{2\pi} F(v) \exp(-i(\xi_1 v_1^2 + \cdots + (v_i^2+v_j^2)(\xi_i\cos^2\theta + \xi_j\sin^2\theta) + \cdots + \xi_N v_N^2))\mathrm{d}v\mathrm{d}\theta \\
&= \frac{1}{\binom{N}{2}} \sum_{i<j} \frac{1}{2\pi} \int_0^{2\pi} \mathcal{R}\left[F\right](\xi_1,\ldots,\xi_i\cos^2\theta + \xi_j\sin^2\theta,\ldots,\xi_i\cos^2\theta+\xi_j\sin^2\theta,\ldots,\xi_N) \\
&= \frac{1}{\binom{N}{2}} \sum_{i<j} \frac{1}{2\pi} \int_0^{2\pi} \mathcal{R}\left[F\right](\xi_{ij}(\theta)).
\end{aligned}$$

Here we first made a change of variables $v \leftrightarrow v_{ij\theta}$ which has Jacobian 1. Then we notice, as in the work of Hauray [75], that there exists α such that $v_i = \sqrt{v_i^2+v_j^2}\cos\alpha$ and that after the rotation we go to $\sqrt{v_i^2+v_j^2}\cos(\alpha+\theta)$. Then we make a new change of variables $\theta+\alpha \to \theta$ which also has Jacobian 1. □

Lemma 6.14. *Following this we have*

$$\rho(Q\mu, Q\nu) \leq \left(1 - \frac{1}{N} - \frac{1}{2N(N-1)}\right)\rho(\mu,\nu).$$

Proof. We calculate that

$$\begin{aligned}
\rho(Q\mu,Q\nu) &\leq \frac{1}{\binom{N}{2}} \sum_{i<j} \sup_{\xi\neq 0, \sum\xi=0} \frac{\left|\fint_0^{2\pi} \mathcal{R}(\mu-\nu)(\xi_{ij}(\theta))\mathrm{d}\theta\right|}{|\xi|^2} \\
\leq& \rho(\mu,\nu)\frac{1}{\binom{N}{2}} \sup_{\xi\neq 0,\sum\xi=0} \fint_0^{2\pi} \frac{|\xi_{ij}(\theta)|^2}{|\xi|^2}\mathrm{d}\theta \\
\leq& \rho(\mu,\nu) \sup_{\xi\neq 0,\sum\xi=0} \left(1 - \frac{1}{\binom{N}{2}}\sum_{i<j}\fint_0^{2\pi} \frac{(2\cos^4\theta-1)\xi_i^2 + (2\sin^4\theta-1)\xi_j^2 + 4\xi_i\xi_j\sin^2\theta\cos^2\theta}{|\xi|^2}\right) \\
\leq& \rho(\mu,\nu) \sup_{\xi\neq 0,\sum\xi=0} \left(1 - \frac{1}{\binom{N}{2}}\sum_{i<j} \frac{-1/4(\xi_i^2+\xi_j^2) + 1/2\xi_i\xi_j}{|\xi|^2}\right) \\
\leq& \rho(\mu,\nu) \sup_{\xi\neq 0,\sum\xi=0} \left(1 - \frac{1}{N} + \frac{1}{4}\frac{1}{\binom{N}{2}} \frac{\left(\sum_k\xi_k\right)^2 - |\xi|^2}{|\xi|^2}\right) \\
\leq& \rho(\mu,\nu)\left(1 - \frac{1}{N} - \frac{1}{2N(N-1)}\right).
\end{aligned}$$

□

Before we begin the main proof we have one more definition.

Definition 6.2. *Let σ be a vector in $\{-1,1\}^N$. Then let us define A_σ to be the subset of $\mathbb{R}^N$ where* *the v_i has sign σ_i. Then we can write the operator*

$$S[F](\sigma) = \int_{\mathbb{R}^N} F(v) 1_{A_\sigma} \mathrm{d}v.$$

Then $S[F]$ is a probability density on the set $\{-1, 1\}^N$.

Now we can prove the theorem.

Proof of Theorem 6.8. We can use this to show exponential decay of the seminorm. Suppose μ and ν are solutions to the Kac master equation supported on the Kac sphere then

$$\begin{aligned}\rho(\mu(t+s), \nu(t+s)) =& \rho(\mu(t), \nu(t))(1 - Ns) + Ns\rho(Q\mu, Q\nu) + o(s) \\ \leq& \rho(\mu(t), \nu(t)) \left(1 - Ns + Ns - s - \frac{s}{2(N-1)}\right) + o(s).\end{aligned}$$

Therefore,

$$\frac{\mathrm{d}^+}{\mathrm{d}t} \rho(\mu(t), \nu(t)) \leq -\left(1 + \frac{1}{2(N-1)}\right) \rho(\mu(t), \nu(t)).$$

Given this we would like to know about the measures supported on the Kac sphere such that $\rho(\mu, \nu) = 0$. We know that it means that

$$\mathcal{R}(\mu - \nu)(\xi) = 0, \quad \forall \xi.$$

We know split the domain of integration defining $\mathcal{R}$ into sections where the v_i have constant sign. We represent the signs by ϵ which is a string of ± 1s. Then in each of these sections we use the change of variables $v_i = \epsilon_i \sqrt{k_i^2}$.

$$\begin{aligned}&\sum_\epsilon \int_{sgn(v)=\epsilon} F(v) \exp\left(-i\left(\xi_1 v_1^2 + \dots \xi_N v_N^2\right)\right) \mathrm{d}\mathbf{v} \\ =& \sum_\epsilon \int_{sgn(v)=\epsilon} F(\epsilon_1 \sqrt{k_1}, \dots, \epsilon_N \sqrt{k_N}) e^{-i\xi \cdot k} \frac{\epsilon_1 \dots \epsilon_N}{2^N \sqrt{k_1 \dots k_N}} \mathrm{d}\mathbf{k}.\end{aligned}$$

We know that the Fourier transform in invertible on the chosen domain so this shows that when $\rho(\mu - \nu) = 0$ then

$$\sum_\epsilon F(\epsilon_1 \sqrt{k_1}, \dots, \epsilon_N \sqrt{k_N}) \frac{\epsilon_1 \dots \epsilon_N}{2^N \sqrt{k_1 \dots k_N}} = \sum_\epsilon G(\epsilon_1 \sqrt{k_1}, \dots, \epsilon_N \sqrt{k_N}) \frac{\epsilon_1 \dots \epsilon_N}{2^N \sqrt{k_1 \dots k_N}}.$$

We can see by a similar change of variables that this sum is the probability density of the energies of the individual particles.

Following this it is interesting to look at the process on the signs of the velocities of the particles. After a collision between particles i and j, with uniform collision kernel the probability that the sign of v_i is positive is $1/2$ and similarly for v_j this is easiest to see with the

$$v_i \to \sqrt{v_i + v_j} \cos(\theta + \alpha)$$

representation. Therefore the probability of a uniform distributions of signs is greater than the probability that all of the particles have been involved in a collision.

With the definition of the operator S, we have that

$$\|S[F] - S[G]\|_{TV} \leq 2Ne^{-t}$$

This is very similar to a lazy random walk on a hypercube which is well studied. We couple the processes as in [75] as described above. Then we just need to wait for every coordinate to be jump. Each particle collides with some other particle with rate 1 and there are N particles so the probability that at least one particle hasn't collided is bounded by $2Ne^{-t}$. $\square$

Bibliography

[1] D. A. Heat transport in low-dimensional systems. *Adv. Phys.*, 57:457–537, 2008.

[2] F. Achleitner, A. Arnold, and E. A. Carlen. On linear hypocoercive BGK models. In *From particle systems to partial differential equations. III*, volume 162 of *Springer Proc. Math. Stat.*, pages 1–37. Springer, [Cham], 2016.

[3] F. Achleitner, A. Arnold, and D. Stürzer. Large-time behavior in non-symmetric Fokker-Planck equations. *Riv. Math. Univ. Parma (N.S.)*, 6(1):1–68, 2015.

[4] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[5] A. Arnold, J.-P. Bartier, and J. Dolbeault. Interpolation between logarithmic Sobolev and Poincaré inequalities. *Commun. Math. Sci.*, 5(4):971–979, 2007.

[6] A. Arnold, A. Einav, and T. Wöhrer. On the rates of decay to equilibrium in degenerate and defective Fokker–Planck equations. *J. Differential Equations*, 264(11):6843–6872, 2018.

[7] A. Arnold and J. Erb. Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift. *ArXiv e-prints*, Sept. 2014.

[8] A. Arnold, P. Markowich, G. Toscani, and A. Unterreiter. On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations. *Comm. Partial Differential Equations*, 26(1-2):43–100, 2001.

[9] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

[10] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[11] F. Baudoin. Wasserstein contraction properties for hypoelliptic diffusions. *ArXiv e-prints*, Feb. 2016.

[12] F. Baudoin. Bakry–Émery meet Villani. *J. Funct. Anal.*, 273(7):2275–2291, 2017.

[13] W. Beckner. A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.*, 105(2):397–400, 1989.

[14] P. G. Bergmann and J. L. Lebowitz. New approach to nonequilibrium processes. *Phys. Rev. (2)*, 99:578–587, 1955.

[15] E. Bernard and F. Salvarani. On the convergence to equilibrium for degenerate transport problems. *Arch. Ration. Mech. Anal.*, 208(3):977–984, 2013.

[16] E. Bernard and F. Salvarani. On the exponential decay to equilibrium of the degenerate linear Boltzmann equation. *J. Funct. Anal.*, 265(9):1934–1954, 2013.

[17] M. Bisi, J. A. Cañizo, and B. Lods. Entropy dissipation estimates for the linear Boltzmann operator. *J. Funct. Anal.*, 269(4):1028–1069, 2015.

[18] A. V. Bobylev and C. Cercignani. On the rate of entropy production for the Boltzmann equation. *J. Statist. Phys.*, 94(3-4):603–618, 1999.

[19] T. Bodineau, J. Lebowitz, C. Mouhot, and C. Villani. Lyapunov functionals for boundary-driven nonlinear drift-diffusion equations. *Nonlinearity*, 27(9):2111–2132, 2014.

[20] F. Bolley and I. Gentil. Phi-entropy inequalities and Fokker-Planck equations. In *Progress in analysis and its applications*, pages 463–469. World Sci. Publ., Hackensack, NJ, 2010.

[21] F. Bolley, I. Gentil, and A. Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.

[22] F. Bolley, A. Guillin, and F. Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *M2AN Math. Model. Numer. Anal.*, 44(5):867–884, 2010.

[23] L. Boltzmann. *Lectures on gas theory.* Translated by Stephen G. Brush. University of California Press, Berkeley-Los Angeles, Calif., 1964.

[24] F. Bonetto, J. L. Lebowitz, and L. Rey-Bellet. Fourier's law: a challenge to theorists. In *Mathematical physics 2000*, pages 128–150. Imp. Coll. Press, London, 2000.

[25] F. Bonetto, M. Loss, H. Tossounian, and R. Vaidyanathan. Uniform approximation of a Maxwellian thermostat by finite reservoirs. *Comm. Math. Phys.*, 351(1):311–339, 2017.

[26] F. Bonetto, M. Loss, and R. Vaidyanathan. The Kac model coupled to a thermostat. *J. Stat. Phys.*, 156(4):647–667, 2014.

[27] E. Bouin, F. Hoffmann, and C. Mouhot. Exponential decay to equilibrium for a fiber lay-down process on a moving conveyor belt. *SIAM J. Math. Anal.*, 49(4):3233–3251, 2017.

[28] M. Briant. Instantaneous exponential lower bound for solutions to the Boltzmann equation with Maxwellian diffusion boundary conditions. *Kinet. Relat. Models*, 8(2):281–308, 2015.

[29] M. Briant. Instantaneous filling of the vacuum for the full Boltzmann equation in convex domains. *Arch. Ration. Mech. Anal.*, 218(2):985–1041, 2015.

[30] K. Burdzy and W. S. Kendall. Efficient Markovian couplings: examples and counterexamples. *Ann. Appl. Probab.*, 10(2):362–409, 2000.

[31] J. A. Cañizo, A. Einav, and B. Lods. On the rate of convergence to equilibrium for the linear Boltzmann equation with soft potentials. *J. Math. Anal. Appl.*, 462(1):801–839, 2018.

[32] J. A. Cañizo and H. Yoldaş. Asymptotic behaviour of neuron population models structured by elapsed-time. *ArXiv e-prints*, Mar. 2018.

[33] M. J. Cáceres, J. A. Carrillo, and T. Goudon. Equilibration rate for the linear inhomogeneous relaxation-time Boltzmann equation for charged particles. *Comm. Partial Differential Equations*, 28(5-6):969–989, 2003.

[34] C. Cao. The kinetic Fokker-Planck equation with weak confinement force. *ArXiv e-prints*, Jan. 2018.

[35] E. Carlen, M. C. Carvalho, and M. Loss. Many-body aspects of approach to equilibrium. In *Séminaire: Équations aux Dérivées Partielles, 2000–2001*, Sémin. Équ. Dériv. Partielles, pages Exp. No. XIX, 12. École Polytech., Palaiseau, 2001.

[36] E. A. Carlen and M. C. Carvalho. Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. *J. Statist. Phys.*, 67(3-4):575–608, 1992.

[37] E. A. Carlen, M. C. Carvalho, J. Le Roux, M. Loss, and C. Villani. Entropy and chaos in the Kac model. *Kinet. Relat. Models*, 3(1):85–122, 2010.

[38] E. A. Carlen, R. Esposito, J. L. Lebowitz, R. Marra, and C. Mouhot. Approach to the steady state in kinetic models with thermal reservoirs at different temperatures. *ArXiv e-prints*, Sept. 2016.

[39] E. A. Carlen, J. L. Lebowitz, and C. Mouhot. Exponential approach to, and properties of, a non-equilibrium steady state in a dilute gas. *Braz. J. Probab. Stat.*, 29(2):372–386, 2015.

[40] P. Cattiaux and A. Guillin. Functional inequalities via Lyapunov conditions. In *Optimal transportation*, volume 413 of *London Math. Soc. Lecture Note Ser.*, pages 274–287. Cambridge Univ. Press, Cambridge, 2014.

[41] P. Cattiaux, A. Guillin, P. Monmarché, and C. Zhang. Entropic multipliers method for langevin diffusion and weighted log sobolev inequalities. *ArXiv e-prints*, Aug. 2017.

[42] M. Chen. Optimal Markovian couplings and applications. *Acta Math. Sinica (N.S.)*, 10(3):260–275, 1994. A Chinese summary appears in Acta Math. Sinica **3**8 (1995), no. 4, 575.

[43] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[44] F. A. D. Die mittlere energie rotierender elektrischer dipole im strahlungsfeld. *Annalen der Physik*, 348(5):810–820.

[45] L. Desvillettes. Entropy dissipation rate and convergence in kinetic equations. *Comm. Math. Phys.*, 123(4):687–702, 1989.

[46] L. Desvillettes and C. Villani. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation. *Comm. Pure Appl. Math.*, 54(1):1–42, 2001.

[47] L. Desvillettes and C. Villani. On the trend to global equilibrium for spatially inhomogeneous kinetic systems: the Boltzmann equation. *Invent. Math.*, 159(2):245–316, 2005.

[48] H. Dietert, J. Evans, and T. Holding. Convergence to equilibrium for the kinetic Fokker-Planck equation on the torus. *ArXiv e-prints*, June 2015.

[49] J. Dolbeault and X. Li. Phi-entropies for Fokker-Planck and kinetic Fokker-Planck equations. *ArXiv e-prints*, Dec. 2017.

[50] J. Dolbeault, C. Mouhot, and C. Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.*, 367(6):3807–3828, 2015.

[51] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3):897–923, 2009.

[52] A. Eberle, A. Guillin, and R. Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *ArXiv e-prints*, Mar. 2017.

[53] J.-P. Eckmann and M. Hairer. Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.*, 212(1):105–164, 2000.

[54] J.-P. Eckmann and M. Hairer. Spectral properties of hypoelliptic operators. *Comm. Math. Phys.*, 235(2):233–253, 2003.

[55] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3):657–697, 1999.

[56] A. Einav. On Villani's conjecture concerning entropy production for the Kac master equation. *Kinet. Relat. Models*, 4(2):479–497, 2011.

[57] J. Evans. Non-equilibrium Steady States in Kac's Model Coupled to a Thermostat. *Journal of Statistical Physics*, 164(5):1103–1121, Sep 2016.

[58] J. Evans. Hypocoercivity in Phi-entropy for the Linear Relaxation Boltzmann Equation on the Torus. *ArXiv e-prints*, Feb. 2017.

[59] G. Gabetta, G. Toscani, and B. Wennberg. Metrics for probability distributions and the trend to equilibrium for solutions of the Boltzmann equation. *J. Statist. Phys.*, 81(5-6):901–934, 1995.

[60] P. Gabriel. Measure solutions to the conservative renewal equation. *ArXiv e-prints*, Apr. 2017.

[61] S. Gadat and L. Miclo. Spectral decompositions and $\mathbb{L}^2$-operator norms of toy hypocoercive semi-groups. *Kinet. Relat. Models*, 6(2):317–372, 2013.

[62] P. E. Goldstein S., Lebowitz J. L. Mechanical systems with stochastic boundaries. *Coll. Math. Soc. J. Bolyai*, 27:401–419, 1978.

[63] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[64] M. P. Gualdani, S. Mischler, and C. Mouhot. Factorisation of non-symmetric operators and exponential h-theorem. *Mem. d. l. SMF*, 153, 2017.

[65] Y. Guo. The Landau equation in a periodic box. *Comm. Math. Phys.*, 231(3):391–434, 2002.

[66] Y. Guo. The Vlasov-Poisson-Boltzmann system near Maxwellians. *Comm. Pure Appl. Math.*, 55(9):1104–1135, 2002.

[67] Y. Guo. Classical solutions to the Boltzmann equation for molecules with an angular cutoff. *Arch. Ration. Mech. Anal.*, 169(4):305–353, 2003.

[68] Y. Guo. The Vlasov-Maxwell-Boltzmann system near Maxwellians. *Invent. Math.*, 153(3):593–630, 2003.

[69] Y. Guo. The Boltzmann equation in the whole space. *Indiana Univ. Math. J.*, 53(4):1081–1094, 2004.

[70] M. Hairer. Lecture notes: P@w course on the convergence of markov processes. 2010.

[71] M. Hairer. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011.

[72] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.

[73] M. Hairer and J. C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.

[74] T. E. Harris. The existence of stationary measures for certain Markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II*, pages 113–124. University of California Press, Berkeley and Los Angeles, 1956.

[75] M. Hauray. Uniform Contractivity in Wasserstein Metric for the Original 1D Kac's Model. *Journal of Statistical Physics*, 162:1566–1570, Mar. 2016.

[76] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[77] F. Hérau. Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation. *Asymptot. Anal.*, 46(3-4):349–359, 2006.

[78] F. Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications. *J. Funct. Anal.*, 244(1):95–118, 2007.

[79] F. Hérau. Introduction to hypocoercive methods and applications for simple linear inhomogeneous kinetic models. *ArXiv e-prints*, Oct. 2017.

[80] F. Hérau and F. Nier. Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.*, 171(2):151–218, 2004.

[81] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[82] A. Iacobucci, S. Olla, and G. Stoltz. Convergence rates for nonequilibrium Langevin dynamics. *ArXiv e-prints*, Feb. 2017.

[83] E. Janvresse. Spectral gap for Kac's model of Boltzmann equation. *Ann. Probab.*, 29(1):288–304, 2001.

[84] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[85] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley and Los Angeles, 1956.

[86] W. S. Kendall. Coupling, local times, immersions. *Bernoulli*, 21(2):1014–1046, 2015.

[87] A. Kolmogoroff. über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, 104(1):415–458, 1931.

[88] S. Kullback. *Information theory and statistics*. Dover Publications, Inc., Mineola, NY, 1997. Reprint of the second (1968) edition.

[89] K. Kuwada. Characterization of maximal Markovian couplings for diffusion processes. *Electron. J. Probab.*, 14:no. 25, 633–662, 2009.

[90] V. Letizia and S. Olla. Nonequilibrium isothermal transformations in a temperature gradient from a microscopic dynamics. *Ann. Probab.*, 45(6A):3987–4018, 2017.

[91] T.-P. Liu, T. Yang, and S.-H. Yu. Energy method for Boltzmann equation. *Phys. D*, 188(3-4):178–192, 2004.

[92] P. Malliavin. Stochastic calculus of variation and hypoelliptic operators. In *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)*, pages 195–263. Wiley, New York-Chichester-Brisbane, 1978.

[93] D. Matthes and G. Toscani. Variation on a theme by Bobylëv and Villani. *C. R. Math. Acad. Sci. Paris*, 350(1-2):107–110, 2012.

[94] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[95] H. P. McKean, Jr. Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Arch. Rational Mech. Anal.*, 21:343–367, 1966.

[96] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.

[97] S. Mischler and C. Mouhot. Exponential stability of slowly decaying solutions to the kinetic-Fokker-Planck equation. *Arch. Ration. Mech. Anal.*, 221(2):677–723, 2016.

[98] P. Monmarché. Generalized Γ calculus and application to interacting particles on a graph. *ArXiv e-prints*, Oct. 2015.

[99] P. Monmarché. A note on Fisher Information hypocoercive decay for the linear Boltzmann equation. *ArXiv e-prints*, Mar. 2017.

[100] P. Mörters and Y. Peres. *Brownian Motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2010.

[101] C. Mouhot. Quantitative lower bounds for the full Boltzmann equation. I. Periodic boundary conditions. *Comm. Partial Differential Equations*, 30(4-6):881–917, 2005.

[102] C. Mouhot and L. Neumann. Quantitative perturbative study of convergence to equilibrium for collisional kinetic models in the torus. *Nonlinearity*, 19(4):969–998, 2006.

[103] J. Norris. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, volume 1204 of *Lecture Notes in Math.*, pages 101–130. Springer, Berlin, 1986.

[104] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[105] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated and edited by Amiel Feinstein. Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964.

[106] M. Planck. *Uber einen Satz der statistischen Dynamik und sein Erweiterung in der Quantentheorie*, volume Jan-Dec 1917. Berlin :Deutsche Akademie der Wissenschaften zu Berlin,, 1917.

[107] L. Rey-Bellet and L. E. Thomas. Asymptotic behavior of thermal nonequilibrium steady states for a driven chain of anharmonic oscillators. *Comm. Math. Phys.*, 215(1):1–24, 2000.

[108] L. Rey-Bellet and L. E. Thomas. Exponential convergence to non-equilibrium stationary states in classical statistical mechanics. *Comm. Math. Phys.*, 225(2):305–329, 2002.

[109] F. Rezakhanlou and C. Villani. *Entropy methods for the Boltzmann equation*, volume 1916 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008. Lectures from a Special Semester on Hydrodynamic Limits held at the Université de Paris VI, Paris, 2001, Edited by François Golse and Stefano Olla.

[110] R. M. Strain and Y. Guo. Almost exponential decay near Maxwellian. *Comm. Partial Differential Equations*, 31(1-3):417–429, 2006.

[111] G. Toscani and C. Villani. Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Comm. Math. Phys.*, 203(3):667–706, 1999.

[112] H. Tossounian. Equilibration in the Kac model using the GTW metric d_2. *J. Stat. Phys.*, 169(1):168–186, 2017.

[113] H. Tossounian and R. Vaidyanathan. Partially thermostated Kac model. *Journal of Mathematical Physics*, 56(8):083301, Aug. 2015.

[114] S. Ukai. On the existence of global solutions of mixed problem for non-linear Boltzmann equation. *Proc. Japan Acad.*, 50:179–184, 1974.

[115] C. Villani. Fisher information estimates for Boltzmann's collision operator. *J. Math. Pures Appl. (9)*, 77(8):821–837, 1998.

[116] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[117] C. Villani. *Optimal transport: Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[118] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, USA, 2001.