

ABSTRACT

Title of dissertation: MEAN FIELD LIMIT FOR
STOCHASTIC PARTICLE SYSTEMS
WITH SINGULAR FORCES
Zhenfu Wang, Doctor of Philosophy, 2017

Dissertation directed by: Professor Pierre-Emmanuel Jabin
Department of Mathematics and CSCAMM
University of Maryland, College Park

In this thesis, we systematically study the mean field limit for large systems of particles interacting through rough or singular kernels by developing a new statistical framework, based on controlling the relative entropy between the N -particle distribution and the limit law through identifying new Laws of Large Numbers.

We study both the canonical 2nd order Newton dynamics and the 1st order (kinematic) systems, leading to McKean-Vlasov systems in the large N limit. For the 2nd order case, we only require that the interactions K be bounded. The control of the relative entropy implies the mean field limit and the propagation of chaos through the strong convergence of all the marginals. For the 1st order case, with the help from noise we can even obtain the mean field limit for interactions $K \in W^{-1,\infty}$, *i.e.* the anti-derivatives of K are bounded (or even unbounded with weak singularity).

To our knowledge, this is the first time the relative entropy method applied to obtain the mean field limit. Compared to the classical framework with $K \in W^{1,\infty}$,

our results show another critical scale $K \in L^\infty$ for the mean field limit. Our results are quantitative: we can provide precise control of the relative entropy and hence the convergence of the marginals. We expect that the relative entropy method will be another standard tool in the study of the mean field limit.

This thesis resulted in the publications [93–95].

MEAN FIELD LIMIT FOR STOCHASTIC PARTICLE SYSTEMS
WITH SINGULAR FORCES

by

Zhenfu Wang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:

Professor Pierre-Emmanuel Jabin, Chair/Advisor

Professor Matei Machedon

Professor Antoine Mellet

Professor Konstantina Trivisa

Professor Christopher Jarzynski, Dean's Representative

© Copyright by
Zhenfu Wang
2017

Dedicated to my parents.

Acknowledgments

First and foremost I would like to express my deepest gratitude to my advisor, Prof. Pierre-Emmanuel Jabin for introducing me to this exciting project and for his consistent support for the last three years. He has always made himself available for help and advice. It has always been a pleasure to work with and learn from such an extraordinary mathematician. He is a great role model for me and I could not imagine having a better advisor.

I would like to thank Prof. Christopher Jarzynski, Prof. Matei Machedon and Prof. Antoine Mellet, Prof. Konstantina Trivisa, for serving on my thesis committee and for sparing their invaluable time reviewing the manuscript. I also want to thank Prof. Jacob Bedrossian for organizing several interesting reading courses, Prof. David Levermore for a lot of instructive comments on PDE RIT and Prof. Eitan Tadmor for first introducing me to CSCAMM and offering so nice work condition there. I also want to thank Prof. Konstantina Trivisa for her interests of my work and mentoring the teaching for Linear Algebra, Prof. Jian-Guo Liu for many interesting discussions and encouragement and Prof. Athanasios Tzavaras for the support during the job search.

My colleagues and friends in the Math Department have enriched my graduate life in many ways and deserve a special mention. Dr. Jinxin Xue and Ms. Chunting Lu helped me a lot when I just came to Maryland. I also have great experience with Dr. Guan Huang, Dr. Wujun Zhang, Dr. Dongmian Zou, Jacky Chong, Tao Zhang, Dr. Scott Smith, Sam Punshon-Smith, Dr. Wenqing Hu, Dr. Changhui

Tan, Siming He etc. from University of Maryland.

I would also want to thank many mathematicians I met during conferences, Prof. Jose Carrillo, Prof. Yao Yao, Prof. Qin Li, Prof. Chanwoo Kim, Prof. Jianfeng Lu, Dr. Zhennan Zhou, Dr. Lei Li, Dr. Xiaoqian Xu, Dr. Cheng Yu, Dr. Wai-Tong Fan, Dr. Ruoyu Wu, Dr. Young-Pil Choi, Dr. Tom Holding etc. The discussion with them is really helpful.

I would also like to acknowledge the help and support from some of the staff members, in particular Celeste from Math Department and Anil, Agi and Sara from CSCAMM.

My PhD project was partially supported by the NSF Grant 1312142 and Ann G. Wylie Dissertation Fellowship.

I owe my deepest thanks to my family - my father, my mother and my sisters who have always taken great pride of me and have always encouraged me to overcome difficulties at times. Words cannot express the gratitude I owe to them.

Table of Contents

List of Abbreviations	vii
1 Introduction	1
1.1 Large systems of particles: canonical models	1
1.2 The mean field limit: McKean-Vlasov PDEs	4
1.3 Examples of interaction kernels and some variant models	5
1.4 Classical mean field framework as introduced by Kac	11
1.4.1 The N -particle Liouville equation	11
1.4.2 Propagation of (Kac's) chaos	12
1.4.3 Formal derivation of the McKean-Vlasov system (1.3) from the BBGKY hierarchy	15
1.5 From relative entropy to propagation of chaos	16
1.5.1 Preliminary about relative entropy	17
1.5.2 Mean field limit for the 2nd order system	19
1.5.3 Consequence of the main result	20
1.5.4 Comparison with the literature	22
1.6 Relative entropy estimates: the need of combinatorics	24
1.6.1 An intuitive example	24
1.6.2 The need of combinatorics	28
1.7 Combinatorics: Laws of Large Numbers	31
1.7.1 Classical Laws of Large Numbers	31
1.7.2 Combinatorics for double multi-indices	34
2 Main results for the 1st order system and the comparison with the literature	40
2.1 Existence of weak solutions of the Liouville equations	40
2.1.1 The 2nd order case	41
2.1.2 The 1st order case	42
2.1.3 Remarks on Proposition 7 and Proposition 8	43
2.2 Main results: Mean field limit for the 1st order system	45
2.3 The difference between the 2nd order case and the 1st order case. . .	47
2.4 Comparison with the literature	50
2.5 Related problems	53

3	Proof of the main result: The 2nd order case	56
3.1	The Vlasov equation (1.3): Weak-strong uniqueness	56
3.2	Main Estimate: The need of combinatorics	59
3.3	From combinatorics and Theorem 5, to Theorem 1	61
4	Proof of the main results: The 1st order case	66
4.1	Main Estimates: Combinatorics results	66
4.2	The evolution of the relative entropy	69
4.3	Control of the K_1 part with $K_1 = \operatorname{div}V$	74
4.3.1	Control of \mathcal{J}_1^b : The bounded part of K_1	75
4.3.2	Control of \mathcal{J}_1^s : The singular part of K_1	79
4.4	Control of the K_2 part with $K_2 \in L^\infty$	83
4.5	Final step of the Proof of Theorem 3	85
5	Preliminary of combinatorics	87
6	Main estimates:The 2nd order case	91
6.1	Intuitive calculations: the scaling of R_N	91
6.2	Main Estimates: Proof of Theorem 5	96
6.2.1	The case $3k \leq N$: Proof of Proposition 15	97
6.2.2	The case $3k > N$: Proof of Proposition 16	103
7	Main Estimates: The 1st order case	105
7.1	Main estimate I : Proof of Theorem 6	105
7.2	Main estimate II: Proof of Theorem 7	109
7.2.1	The case $4k > N$: Proof of Proposition 17	111
7.2.2	The case $4 \leq 4k \leq N$: Proof of Proposition 18	112
A	Preliminary about entropy and relative entropy	121
A.1	Definitions	121
A.2	Monotonicity of the relative entropy	122
A.3	Csiszár-Kullback-Pinsker(CKP) inequality	124
A.4	Lower bounds for the entropy	124
B		126
B.1	Weak-strong uniqueness on Eq. (1.3) and the proof of Theorem 4 . .	126
B.2	Proof of Proposition 10	128
	Bibliography	133

List of Notations

E	As a space it denotes a Polish space
$\mathcal{P}(E)$	Space of probability measures on the Polish space E
$\mathcal{P}_{\text{Sym}(E^N)}$ or $\mathcal{P}_{\text{Sym}(E^k)}$	Space of symmetric probability measures on E^N or E^k
$X = (x_1, \dots, x_N)$	Short notation for space variables
$V = (v_1, \dots, v_N)$	Short notation for velocity variables
$z_i = (x_i, v_i)$	Short notations for (x_i, v_i)
$Z = (X, V)$	Short notation for all variables
$f_N(t, Z)$	Joint distribution of $(X_1(t), V_1(t), \dots, X_N(t), V_N(t))$
$\rho_N(t, X)$	Joint distribution of $(X_1(t), \dots, X_N(t))$
$\partial_t f$	The time derivative of function f
$f_t(\cdot)$ or $f(t, \cdot)$	The function f valued at time t
$\bar{f}_N(t, Z)$ or $f_t^{\otimes N}$	The full tensor product $\bar{f}_N(t, Z) = \prod_{i=1}^N f_t(z_i)$
$\bar{\rho}_N(t, X)$ or $\rho_t^{\otimes N}$	The full tensor product $\bar{\rho}_N(t, X) = \prod_{i=1}^N \rho_t(x_i)$
$f_t^{\otimes k}$	The k -tensor product of f as $\prod_{i=1}^k f_t(z_i)$
$\rho_t^{\otimes k}$	The k -tensor product of ρ as $\prod_{i=1}^k \rho_t(z_i)$
$f_{N,k}$	The k -marginal distribution of f_N
$\rho_{N,k}$	The k -marginal distribution of ρ_N

Chapter 1: Introduction

In this thesis, we rigorously derive mean field equations from large systems of interacting particles with singular or rough interaction kernels, focusing on the stochastic case where a large system of Stochastic Differential Equations (SDEs) converges to a McKean-Vlasov Partial Differential Equation (PDE) as the number N of particles goes to infinity. This is a longstanding open and challenging question, considered as part of Hilbert’s 6th problem, which has only a few recent successes. We refer to the book [133] and recent reviews [69, 91, 94] for detailed introduction of this subject.

1.1 Large systems of particles: canonical models

Large systems of interacting particles are now fairly ubiquitous. They are usually formulated by first-principle (for instance Newton’s 2nd law) individual based models which are conceptually simple. For instance, in physics particles can represent ions and electrons in plasmas [144], or molecules in a fluid [90] or even galaxies [1] in some cosmological models; in biosciences they typically model the collective behavior of animals or micro-organisms (cell or bacteria) [29, 42, 118]; in economics or social sciences particles are individual “agents” or “players” [99, 119, 145].

Large systems of particles are usually (at least in the classical regime) modeled by systems of Ordinary Differential Equations (ODE) or SDEs. In this thesis, we focus on two canonical models of large systems of particles formulated below.

The most classical model is the Newton dynamics for N indistinguishable point particles driven by 2-body interaction forces and Brownian motions. Denote by $X_i \in \mathbb{D}$ and $V_i \in \mathbb{R}^d$ the position and velocity of particle number i . The evolution of the system is given by the following SDEs,

$$dX_i = V_i dt, \quad dV_i = \frac{1}{N} \sum_{j \neq i} K(X_i - X_j) dt + \sqrt{2\sigma_N} dW_t^i, \quad (1.1)$$

where $i = 1, 2, \dots, N$. The W^i are N independent Brownian motions or Wiener processes, which may model various types of random phenomena: For instance random collisions against a given background. The stochastic term here and later in (1.2) should be understood in the Itô sense. If $\sigma_N \equiv 0$, the system (1.1) reduces to the classical deterministic Newton dynamics. Here vector valued kernels K model the interaction forces between two particles. Detailed discussions on various choices of K will appear in Section 1.3. We use the convention that $K(0) = 0$, *i.e.* there is no self-interaction.

The space domain \mathbb{D} may be the whole space \mathbb{R}^d , the flat torus \mathbb{T}^d or some bounded domain. The analysis of a bounded, smooth domain is strongly dependent on the type of boundary conditions but can sometimes be handled in a similar manner with some adjustments. Thus for simplicity we typically limit ourselves to $\mathbb{D} = \mathbb{R}^d, \mathbb{T}^d$. Even if \mathbb{D} is bounded, there is no hard cap on velocities so that the actual domain in position and velocity, $\mathbb{D} \times \mathbb{R}^d$ is always unbounded.

The critical scaling in (1.1) (and later in (1.2)) is the factor $\frac{1}{N}$ in front of the interaction terms. This is *the mean field scaling* and it keeps, at least formally, the total strength of the interaction of order 1. For more detailed discussion on the mean field scaling and other type scalings, we refer to the discussion in Section 1.1 in the review [91].

As the companion of (1.1), we also consider the 1st order stochastic system

$$dX_i = F(X_i) dt + \frac{1}{N} \sum_{j \neq i} K(X_i - X_j) dt + \sqrt{2\sigma_N} dW_t^i, \quad (1.2)$$

where $i = 1, \dots, N$, F models the exterior forces and other assumptions follows the 2nd order system (1.1).

In the deterministic regime, *i.e.* $\sigma_N \equiv 0$, the 1st order system (1.2) comprises the 2nd system (1.1) as a special case. Indeed, by setting that

$$Z_i = (X_i, V_i), \quad F(Z_i) = (V_i, 0), \quad \tilde{K}(Z_i, Z_j) = (0, K(X_i - X_j)),$$

the 1st order system (1.2) with \tilde{K} defined above reduces to the 2nd order system (1.1).

However, in the stochastic case when $\sigma_N > 0$, we have a full diffusion in (1.2) while only a degenerate diffusion (only on the velocity variables) in (1.1). This will have several important consequences. See the discussions in Section 2.1.3.

We focus on the canonical models (1.1) and (1.2) simply because with various kernels K they are enough for many interesting applications and capture the essential difficulties of the mean field limit problem. We believe that our method have implications well beyond them: models with friction, self-compelled terms,

multi-species, even with space dependent strength σ_N of noises and various models in biophysics or in quantum mechanics settings...

1.2 The mean field limit: McKean-Vlasov PDEs

Due to the large number N of particles, it is extremely complicated and costly to study or simulate the microscopic systems (1.1) or (1.2) directly. The number N of particles can be as large as 10^{25} for typical physical settings and 10^9 in typical bioscience settings. Even for $N = 4, 5$, the dynamics of certain ODE systems (let alone SDE systems) can be so chaotic [139–141] that it is impossible to trace the trajectories of particles exactly. Fortunately, the large scale dynamics (for instance the statistical information or the average behavior) can usually be approximated by a continuous PDE model, thanks to the very famous critical mechanism known as *Laws of Large Numbers*, in which people are most interested for practical purposes.

The basic but fundamental idea to reduce this complexity by deriving a mesoscopic or macroscopic system dates back to Maxwell and Boltzmann in their work on the later called Boltzmann equation. For the derivation of the Boltzmann equation, we only refer the readers to [37, 65, 101]. Here we work on a different regime: the collision-less regime under the mean field scaling.

For the 2nd order system (1.1), for very large N , one expects to approximate the system (1.1) by the following Vlasov equation or McKean-Vlasov equation (if diffusion is present)

$$\partial_t f + v \cdot \nabla_x f + K \star \rho \cdot \nabla_v f = \sigma \Delta_v f, \quad \rho(t, x) = \int_{\mathbb{R}^d} f(t, x, v) dv \quad (1.3)$$

where the unknown $f = f(t, x, v)$ is the phase space density or 1-particle distribution and $\rho = \rho(t, x)$ is the spatial (macroscopic) density and $\sigma_N \rightarrow \sigma \geq 0$. Our central problem is then to *show the mean field limit of the system (1.1) towards McKean-Vlasov equation (1.3) and in particular to quantify how close they are for a given N .*

Similarly, for 1st order system (1.2), one expects that as the number N of particles goes to infinity the system (1.2) will converge to the following PDE

$$\partial_t \rho + \operatorname{div}_x (\rho [F + K \star \rho]) = \sigma \Delta_x \rho, \quad (1.4)$$

where the unknown $\rho = \rho(t, x)$ is the spatial density and again $\sigma_N \rightarrow \sigma \geq 0$.

1.3 Examples of interaction kernels and some variant models

In this section, we list some examples of K and discuss variant models of (1.1) and (1.2). The references that are cited have no pretension to be exhaustive but hopefully indicate that it is critical to consider the mean field limit for systems with singular or rough kernels.

- *The Poisson kernel.* For the 2nd order system (1.1), the best known example of interaction kernel is the Poisson kernel, that is

$$K(x) = \pm C_d \frac{x}{|x|^d}, \quad d = 2, 3, \dots,$$

where $C_d > 0$ is a constant depending on the dimension and the physical parameters of the particles (mass, charges...). This corresponds to particles under gravitational interactions for the case with a minus sign and electrostatic interactions (ions in a

plasma for instance) for the case with a positive sign. See [96, 144] for the original modelings and [66, 67] for particle methods for the Vlasov-Poisson system (1.3).

The 1st order model (1.2) can be regarded as the zero inertia limit (Smoluchowski-Kramers approximation) (see for instance [56, 136]) of Langevin equations in statistical physics. However, the model (1.2) has its own important applications.

- *The Biot-Savart kernel.* The most famous example is the stochastic vortex model (1.2) with $F = 0$ in fluid dynamics with the Biot-Savart kernel

$$K(x) = \frac{1}{2\pi} \left(\frac{-x_2}{|x|^2}, \frac{x_1}{|x|^2} \right),$$

which is widely used to approximate the 2D Navier-Stokes equation written in vorticity form. See for instance [27, 28, 63, 113, 124] and (random) point vortex method [41, 72, 110].

One important class of the kernels are given in the gradient form $K = -\nabla W$, where W are interaction potential functions. This class includes the Poisson kernels as discussed above. Indeed, one chooses $W(x) = \pm C_d/|x|^{d-2}$ for $d \geq 3$ and $W(x) = \mp \frac{1}{2\pi} \log|x|$ for $d = 2$, where $C_d > 0$. The positive sign in $d \geq 3$ and the minus sign in $d = 2$ correspond to repulsive forces. However, we have more examples of K in the gradient form.

- *The 2nd order system (1.1) with kernels $K = -\nabla W$.* For the 2nd order case, the interaction potential W can model the short-range repulsion and long-range attraction mechanism in bioscience or physical applications. For instance W might be

$$W(x) = -C_A e^{-|x|/l_A} + C_R e^{-|x|/l_R},$$

where C_A, C_R and l_A, l_R are the strengths and the typical lengths of attraction and repulsion respectively. See [50] for the modeling and [17] for the mean field limit.

- *The 1st order system (1.2) with kernels $K = -\nabla W$.* For the 1st order system, the kernels K can also be the Poisson kernels, in particular $W(x) = \frac{1}{2\pi} \log|x|$ in 2D, the system (1.4) corresponds to the famous Keller-Segel equation of chemotaxis, a canonical model for the collective motion of micro-organisms. The corresponding microscopic model (1.2) is usually used as a particle model to approximate (1.4). We refer mainly to [64] for the mean field limit, together with [68, 108].

In general, we consider aggregation models (1.2) with an exterior force $F(X_i) = -\nabla V(X_i)$. Mathematically well-investigated models typically require that W and V are (quasi-)convex and with polynomial or exponential growth at infinity, with the help of gradient flow structures. They are widely used in many settings such as in biology, ecology and in study of space homogeneous granular media [9]. See for instance [19, 20, 36, 43, 111, 112] for the mathematical study of the particle system (1.2) and more recently the mean field limit [10, 11, 33, 35, 51] using the gradient flow techniques as in [5]. Similar to the 2nd order case above, certain choices of W can model the short-range repulsion and long-range attraction mechanism. For instance one can choose

$$W(x) = \frac{1}{|x|^{d-2}} + \frac{1}{2}|x|^2$$

for $d \geq 3$ as in [32] (see the references therein for a more detailed modeling discussion).

In these cases the kernels K are usually only locally Lipschitz or even singular.

See also the examples in [51] where $K = -\nabla W$ and W can be chosen as the s -Riesz functions as

$$W(x) = \begin{cases} \frac{1}{C_{d,s}} \frac{1}{|x|^s}, & \text{if } 0 < s < d, \\ -\frac{1}{C_{d,0}} \log |x|, & \text{if } s = 0. \end{cases} \quad (1.5)$$

where $C_{d,s}$ are certain normalized constants depending on the dimension d .

The gradient flow structure for the 1st order system (1.2) with $K = -\nabla W$ shall be compared to the Hamiltonian structure for the 1st order systems with $\operatorname{div}_x K = 0$ (one example of K is the Biot-Savart kernel) and the 2nd order systems (1.1). The Hamiltonian structure, *i.e.* the velocity fields $K \star \rho$ in the 1st order system and $(v, K \star \rho(x))$ in the 2nd order system are divergence free, enjoys a special attention in this thesis.

In the following, we discuss some variant models of (1.1) or (1.2).

- *Fokker-Planck equation.* One can add extra terms like friction or self-propulsion in the acceleration dV_i in (1.1). For example, the expected limit (1.3) with an extra term $-\kappa \operatorname{div}_v(vf)$ in the left-hand side, correspondingly the particle system (1.1) with an extra friction term $-\kappa V_i dt$ in the acceleration dV_i , is usually called the Fokker-Planck (Vlasov-Poisson-Fokker-Planck if K is the Poisson kernel) equation in the physics literature. See [84] for the mean field limit in 1D case.

- *Alignment models.* Since the pioneering works in [42, 137] and later in [118], Newton like systems (variants of (1.1)) have been used to model flocks of birds, schools of fish, swarms of insects... One can see [29, 34, 76] and the references therein for a more detailed discussion of flocking or swarming models in the literature. In

the Cucker-Smale model [42] the evolution of particle number i reads

$$dX_i = V_i dt, \quad dV_i = \frac{1}{N} \sum_{j \neq i} k(|X_i - X_j|)(V_j - V_i)$$

where $i = 1, \dots, N$. Or similarly, one can also consider the corresponding variant of the 1st order model as

$$dX_i = \frac{1}{N} \sum_{j \neq i} k(|X_i - X_j|)(X_j - X_i),$$

where $i = 1, \dots, N$. These alignment models are also quite popular in modeling opinion dynamics [99, 119] and synchronization [100] for instance.

Here k is a scalar function now modeling the strength of the alignment, which typically in the form of $1/(1 + |x|)^\alpha$ in [29, 42] or singular $1/|x|^\alpha$ in [31]. In [118] the strength is normalized as

$$\frac{k(|X_i - X_j|)}{\sum_{k=1}^N k(|X_i - X_k|)}.$$

Hence the force acting on each particle i is automatically bounded.

- *Why stochastic models?* Sometimes the presence of the noise in the models is important since we cannot expect animals to interact with each other or the environment in a completely deterministic way. We in particular refer to [75] for stochastic Cucker-Smale model with additive white noise as in (1.2) and to [3] for multiplicative white noise in velocity variables respectively. The rigorous proof of the mean field limit was given in [17] for systems similar to (1.1) with locally Lipschitz vector fields; the mean-field limit for stochastic Vicsek model where the speed is fixed is given in [18].

- *Rough kernels or kernels with discontinuities/jumps at critical distances.*

In the above examples, K can be *singular* at the origin, *i.e.* $|K(x)| \rightarrow \infty$ or $|\nabla K(x)| \rightarrow \infty$ as $|x| \rightarrow 0$, but they are usually smooth outside any neighborhood of 0. This does not hold for many applications.

For instance, in typical social science or bioscience settings, it is natural to have *discontinuous* kernels, which means that the interaction between two particles (a prey and a predator, a buyer and a seller, two birds in a flock...) could change abruptly at certain critical distances. For instance, birds or mammals only have limited vision abilities [30]: outside a visible region the interaction might suddenly vanish. We can thus only expect localized interactions, for instance $K(x) = h(|x|) = 0$ if $|x| > R$ where h is function measuring the vision ability and $R > 0$ is the maximum distance an animal can see or the minimum distance to take action for instance run away from predators. Here h can be discontinuous or only in L^∞ as in [30]. See also [85, 86] for modeling discussions.

- *General collision models.* In collision models, particles only interact when they collide. A canonical example is the famous Boltzmann equation describing the evolution of dilute gases [14, 15, 37, 101]. More general, one can consider particles with non-smooth shapes (for instance cells or micro-organisms) in fluids which only interact when they collide. For instance $K(X_i - X_j)$ is more or less related to $\nabla \mathbf{1}_{C_j}$ in (1.2), where C_j is the region occupied by the j -th particle. In this case K can be chosen to be a measure in an appropriate way on a sphere or even not a measure. For fixed N , this general dynamics may even not be well-defined. But the large scale dynamics similar to (1.3) or (1.4) might be clarified mathematically.

As in [12, 78], one can also consider collective dynamics in the sense of Cucker-Smale but driven by rank-based interactions. For instance, each particle (bird) can only be influenced by the nearest m particles. See [46, 105] for examples from evolutionary game theory and economics respectively. In the large N limit, this will lead to a Boltzmann type PDE. See also another rank-based model called competing Brownian particles in [127], with possible applications in stock markets for instance.

1.4 Classical mean field framework as introduced by Kac

We introduce the classical setting for the mean field limit introduced by Kac [97], focusing on the simple but significant 2nd order system (1.1), leading to the McKean-Vlasov equation (1.3) in the large N limit.

1.4.1 The N -particle Liouville equation

The starting point of our statistical framework is the joint distribution/law

$$f_N(t, Z) = \text{Law}(X_1(t), V_1(t), \dots, X_N(t), V_N(t)) \in \mathcal{P}(E^N),$$

where $E = \mathbb{D} \times \mathbb{R}^d$. The evolution of the joint law f_N is governed by the N -particle Liouville equation (or the Master equation)

$$\partial_t f_N + \sum_{i=1}^N v_i \cdot \nabla_{x_i} f_N + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i}^N K(x_i - x_j) \cdot \nabla_{v_i} f_N = \sigma_N \sum_{i=1}^N \Delta_{v_i} f_N, \quad (1.6)$$

corresponding to (1.1) and usually coupled with initial data f_N^0 . It can be derived by applying Itô's formula to $d\phi(X_1(t), V_1(t), \dots, X_N(t), V_N(t))$, where ϕ is a test function. See Proposition 7 on the existence of weak solutions of (1.6) and related issues in Section 2.1.

The fact that particles are indistinguishable implies that $f_N \in \mathcal{P}_{\text{Sym}}(E^N)$, a symmetric probability measure on the space E^N . That is for any permutation of indices $\tau \in S_N$,

$$f_N(t, z_1, \dots, z_N) = f_N(t, z_{\tau(1)}, \dots, z_{\tau(N)}).$$

We can then define the k -marginals of f_N as

$$f_{N,k}(z_1, \dots, z_k) = \int_{E^{N-k}} f_N(t, z_1, \dots, z_N) dz_{k+1} \cdots dz_N,$$

where $z_i = (x_i, v_i)$, $Z = (z_1, \dots, z_N)$ and $E = \mathbb{D} \times \mathbb{R}^d$. It is easy to check that the k -marginal distribution is also symmetric $f_{N,k} \in \mathcal{P}_{\text{Sym}}(E^k)$ for $2 \leq k \leq N$.

The joint distribution $f_N \in \mathcal{P}(E^N)$ contains all the information of the particle system (1.1) but is not experimentally observable. Instead the observable statistical information (temperature, pressure or other macroscopic quantities) of the system is contained in the marginals $f_{N,k}$: Usually it is enough to know the behavior of the marginals for practical reasons.

1.4.2 Propagation of (Kac's) chaos

The original notion of propagation of chaos goes as far back as Maxwell and Boltzmann. The classical notion of propagation of chaos was formalized by Kac in [97].

Let us begin with the the simplest definition

Definition 1 *Assume E is a Polish space. A law $f_N \in \mathcal{P}(E^N)$ is tensorized/chaotic if there exists a probability measure $f \in \mathcal{P}(E)$ such that*

$$f_N(z_1, \dots, z_N) = \prod_{i=1}^N f(z_i).$$

We can simply denote as $f_N = f^{\otimes N}$.

The chaotic initial data $f_N^0 = f_0^{\otimes N}$ for the Liouville Eq. (1.6) means that the initial phase space positions $Z_i(0) = (X_i(0), V_i(0))$ ($i = 1, \dots, N$) are independent and identically distributed according to the common law f_0 . This is a usual assumption, in particular in the probability community, for the initial distributions. It is reasonable since in general the initial condition can be the result of a different dynamics which could have an ergodic or mixing property.

But the chaotic initial condition is too strong. It is more realistic to have the independence or chaos in the large N limit, instead of finite N , for the marginals $f_{N,k}$ in fixed dimension k , instead of the joint law f_N in dimension N .

This leads to Kac's chaos, which is an asymptotic chaos in the large N limit.

Definition 2 (Kac's chaos) *Let E be a Polish space (In this section $E = \mathbb{D} \times \mathbb{R}^d$).*

A sequence $(f_N)_{N \geq 2}$ of symmetric probability measures on E^N is said to be f -chaotic for a probability measure f on E , if one of the following equivalent properties holds:

i) For any fixed $k = 1, 2, 3, \dots$, the k -marginal $f_{N,k}$ of f_N converges weakly to $f^{\otimes k}$ as N goes to infinity, i.e. $f_{N,k} \rightharpoonup f^{\otimes k}$;

ii) The second marginal $f_{N,2}$ converges weakly to $f^{\otimes 2}$ as N goes to infinity: $f_{N,2} \rightharpoonup f^{\otimes 2}$;

iii) The empirical measure (random probability measure valued in $\mathcal{P}(E)$) associated with f_N , that is

$$\mu_N(z) = \frac{1}{N} \sum_{i=1}^N \delta(z - Z_i) \quad (1.7)$$

with $f_N = \text{Law}(Z_1, \dots, Z_N)$ where $(Z_1, \dots, Z_N) \in E^N$ are exchangeable random vari-

ables, converges in law to the deterministic measure f as N goes to infinity.

Here the weak convergence $f_{N,k} \rightharpoonup f^{\otimes k}$ simply means that for any test functions $\phi_1, \dots, \phi_k \in C_b(E)$,

$$\lim_{N \rightarrow 0} \int_{E^k} \phi_1(z_1) \cdots \phi_k(z_k) f_{N,k}(z_1, \dots, z_k) dz_1 \cdots dz_k = \prod_{i=1}^k \int_E f(z_i) \phi_i(z_i) dz_i,$$

and μ_N converges in law to f means for any test function $\phi \in C_b(E)$,

$$\begin{aligned} & E_{f_N} \left| \int_E \phi(z) d\mu_N(z) - \int_E \phi(z) f(z) dz \right|^2 \\ &= E_{f_N} \left| \frac{1}{N} \sum_{i=1}^N \phi(Z_i) - \int_E \phi(z) f(z) dz \right|^2 \rightarrow 0 \end{aligned}$$

as $N \rightarrow \infty$, where E_{f_N} means the expectation is taken according to the law f_N . In this section, we chose $Z_i = (X_i(t), V_i(t)) \in E = \mathbb{D} \times \mathbb{R}^d$.

We refer to [135] for the classical proof of equivalence between the three properties. A version of the equivalence has recently been obtained in [83], quantified by the 1 Monge-Kantorovich-Wasserstein (MKW) distance between the laws.

Even with the very strong chaotic initial condition $f_N^0 = f_0^{\otimes N}$, we can only expect that the solution $f_N(t)$ to the Liouville Eq. (1.6) is f_t -chaotic in the asymptotic or Kac's sense as Definition 2. Indeed, for fixed N the solution $f_N(t)$ cannot be chaotic. There are correlations between particles simply because they are interacting with each other through the force term K and hence strict independence is only possible asymptotically as $N \rightarrow \infty$ as an effect of Laws of Large Numbers.

We are more interested in propagating (Kac's) chaos, *i.e.* considering whether or not the initial asymptotically chaotic condition can be propagated for certain time if we run the dynamics (1.1) or (1.6). This corresponds to the notion of propagation of chaos.

Definition 3 (Propagation of (Kac's) chaos) *Assume that the sequence of initial data $(f_N^0)_{N \geq 2}$ is f_0 -chaotic. Then propagation of chaos holds for systems (1.1) (or (1.6)) up to time $T > 0$ iff for any $t \in [0, T]$, the sequence $(f_N(t))_{N \geq 2}$ is also f_t -chaotic, where f_t is the solution to the limit (1.3) with initial data f_0 .*

1.4.3 Formal derivation of the McKean-Vlasov system (1.3) from the BBGKY hierarchy

Propagation of chaos is the key concept to obtain the mean field limit (the classical form is given in Def. 2 iii) for instance). Assuming propagation of chaos, in particular for any k fixed $f_{N,k}(t) \rightarrow f_t^{\otimes k}$ up to time $t \leq T$, one can formally derive the McKean-Vlasov equation (1.3) from the Liouville equation (1.6) through the famous BBGKY hierarchy.

We only give a sketch of the formal calculations here. The readers are encouraged to see [69] for a more detailed discussion.

From the Liouville equation (1.6), it is easy to deduce equations on each marginal $f_{N,k}$. Applying the fact $f_N \in \mathcal{P}_{\text{Sym}}(E^N)$ and using the appropriate permutation, one obtains the BBGKY hierarchy

$$\begin{aligned} \partial_t f_{N,k} + \sum_{i=1}^k v_i \cdot \nabla_{x_i} f_{N,k} + \frac{1}{N} \sum_{i=1}^k \sum_{j=1, j \neq i}^k K(x_i - x_j) \cdot \nabla_{v_i} f_{N,k} \\ + \frac{N-k}{N} \sum_{i=1}^k \int_{\mathbb{D} \times \mathbb{R}^d} K(x_i - x) \cdot \nabla_{v_i} f_{N,k+1}(t, z_1, \dots, z_k, z) dz = \sigma_N \sum_{i=1}^k \Delta_{v_i} f_{N,k}, \end{aligned} \quad (1.8)$$

where $z = (x, v)$ and $z_i = (x_i, v_i)$.

Writing the formal limit of $f_{N,k}$ as $f_{\infty,k}$, formally one obtains the Vlasov hier-

archy or the mean field hierarchy

$$\begin{aligned} & \partial_t f_{\infty,k} + \sum_{i=1}^k v_i \cdot \nabla_{x_i} f_{\infty,k} + \\ & + \sum_{i=1}^k \int_{\mathbb{D} \times \mathbb{R}^d} K(x_i - x) \cdot \nabla_{v_i} f_{\infty,k+1}(t, z_1, \dots, z_k, z) dz = \sigma \sum_{i=1}^k \Delta_{v_i} f_{\infty,k}, \end{aligned} \tag{1.9}$$

where $\sigma = \lim_{N \rightarrow \infty} \sigma_N \geq 0$. Taking the tensorized form of $f_{\infty,k} = f^{\otimes k}$ and also $f_{\infty,k+1} = f^{\otimes(k+1)}$, given by the propagation of chaos, all (1.9) reduce to the Vlasov equation (1.3).

The BBGKY hierarchy is more like a formal tool to get the right mean field equations. The best rigorous result [132] to obtain the mean field limit through the BBGKY hierarchy up to now still requires $K \in W^{1,\infty}$. We now switch to the original Liouville equation (1.6), which contains exactly the same information as in the BBGKY hierarchy.

1.5 From relative entropy to propagation of chaos

Our method works in the level of the Liouville equation which has several advantages. First, we only need the existence of weak solutions to the Liouville equation (1.6), which is possible under very weak assumptions of K , for instance $K \in L^\infty$. Second, working on the Liouville equation is conceptually easy: We do not need to consider certain technical issues of stochastic processes.

The main difficulty of the mean field limit is the lack of an appropriate norm which can measure the distance between the particle system (1.1) and its limit (1.3). In the following, we will show that the (scaled) relative entropy is the right norm to obtain the mean field limit.

1.5.1 Preliminary about relative entropy

Here we only define the key concepts in this chapter. For a more complete discussion on entropy and relative entropy, we refer the readers to Appendix A.

Let us define the (scaled) relative entropy of the joint distribution $f_N \in \mathcal{P}_{\text{Sym}}(E^N)$ with respect to the full tensor product of $f \in \mathcal{P}(E)$, *i.e.* $\bar{f}_N := f^{\otimes N}$ or $\bar{f}_N(Z) = \prod_{i=1}^N f(z_i)$, as the following

$$H_N(f_N|\bar{f}_N) = \frac{1}{N} \int_{E^N} f_N \log \frac{f_N}{\bar{f}_N} dZ$$

where $Z = (z_1, \dots, z_N)$ and $z_i \in E$ and $E = \mathbb{D} \times \mathbb{R}^d$ if we choose f_N as a weak solution to the Liouville (1.6), and f as the strong solution to the limit (1.3).

Similarly, one can define that the k dimensional relative entropy of $f_{N,k}$ *w.r.t.* the k -tensor product $f^{\otimes k}$ of f as

$$H_k(f_{N,k}|f^{\otimes k}) := \frac{1}{k} \int_{E^k} f_{N,k} \log \frac{f_{N,k}}{f^{\otimes k}} dz_1 \cdots dz_k.$$

It is easy to check that any relative entropy (once well-defined) must be non-negative. However, a more important observation is the monotonicity of the (scaled) relative entropy as per

Proposition 1 (Monotonicity of the scaled relative entropy) *For each $1 \leq k \leq N$, one has*

$$0 \leq H_k(f_{N,k}|f^{\otimes k}) \leq H_N(f_N|f^{\otimes N}).$$

The proof of a slightly stronger version is given in Appendix A.

Prop.1 indicates that the relative entropy estimate for the joint law f_N can be transferred to its marginals. This is really crucial, usually absent for other norms for instance L^p norm.

The k dimensional relative entropy $H_k(f_{N,k}|f^{\otimes k})$ can in turn control $\|f_{N,k} - f^{\otimes k}\|_{L^1}$ thanks to the very famous

Lemma 1 (Classical Csiszár-Kullback-Pinsker inequality) *Assume that E is a Polish space. Let $F, G \in \mathcal{P}(E) \cap L^1(E)$, then*

$$\|F - G\|_{L^1} \leq \sqrt{2H(F|G)},$$

where the relative entropy is not scaled, that is

$$H(F|G) = \int_E F(Z) \log \frac{F(Z)}{G(Z)} dZ.$$

The proof of this lemma can be found in Chapter 22 in [138]. We refer the readers to Appendix A for a baby version and its elementary proof.

Combining Prop. 1 and Lemma 1, one reaches the following crucial estimate.

Proposition 2 (From relative entropy to chaos) *Assume that E is a Polish space, $f_N \in \mathcal{P}_{Sym}(E^N)$ and $f \in \mathcal{P}(E)$. Then one has*

$$\|f_{N,k} - f^{\otimes k}\|_{L^1} \leq \sqrt{2kH_k(f_{N,k}|f^{\otimes k})} \leq \sqrt{2kH_N(f_N|f^{\otimes N})}.$$

Proof Applying Lemma 1 for two probability measures $f_{N,k}$ and $f^{\otimes k}$ on E^k , one has the first inequality. The last inequality follows Prop. 1. \square

The joint distribution f_N itself is not very interesting and more like an intermediate object to get the physical relevant quantities for instance the marginals

$f_{N,k}$. Prop. 2 indicates the possibility to work directly at the level of the Liouville equation but hopefully to recover the information of the marginals. This is the main idea of our framework: we directly compare the joint law f_N to the limit through the relative entropy, which in turn implies the propagation of chaos.

1.5.2 Mean field limit for the 2nd order system

With an asymptotically chaotic initial condition in the sense of relative entropy, *i.e.* $H_N(0) := H_N(f_N^0 | f_0^{\otimes N}) \rightarrow 0$ as $N \rightarrow \infty$ and an expected evolution bound $\frac{d}{dt} H_N(f_N(t) | f_t^{\otimes N}) \leq \frac{C}{N}$, we obtain the propagation of chaos for the 2nd order system (1.1) as per

Theorem 1 (Propagation of Chaos for the 2nd order system) *Assume $K \in L^\infty$ and that the limiting solution $f(t, x, v) \in L^\infty([0, T], L^1(\mathbb{D} \times \mathbb{R}^d) \cap W^{1,p})$ for every $1 \leq p \leq \infty$ solves the Vlasov Eq. (1.3) with the bound*

$$\theta_f = \sup_{t \in [0, T]} \int_{\mathbb{D} \times \mathbb{R}^d} e^{\lambda_f |\nabla_v \log f|} f \, dx \, dv < \infty, \quad (1.10)$$

for some $\theta_f, \lambda_f > 0$. For the case of vanishing randomness, that is in the case $\sigma_N \rightarrow \sigma = 0$, we further assume that

$$\sup_{t \in [0, T]} |\nabla_{(x,v)} \log f(t, x, v)| \leq C(1 + |x|^k + |v|^k). \quad (1.11)$$

Assume that the initial data f_N^0 of the Liouville equation (1.6) satisfies

$$\sup_{N \geq 2} \frac{1}{N} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N^0 \log f_N^0 \, dZ < \infty, \quad \sup_{N \geq 2} \frac{1}{N} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} \sum_{i=1}^N (1 + |z_i|^2) f_N^0 \, dZ < \infty, \quad (1.12)$$

as well as

$$H_N(f_N^0 | f_0^{\otimes N}) = \frac{1}{N} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N^0 \log\left(\frac{f_N^0}{f_0^{\otimes N}}\right) dZ \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

In the case $\sigma_N \rightarrow \sigma = 0$, we also assume that

$$\sup_{N \geq 2} \frac{1}{N} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} \sum_{i=1}^N (1 + |x_i|^{2k} + |v_i|^{2k}) f_N^0 dZ < \infty. \quad (1.13)$$

There exists a universal constant C s.t. for any corresponding weak solution f_N to the Liouville Eq. (1.6) as given by Proposition 7 and for any $t \leq T$

$$H_N(f_N(t) | f_t^{\otimes N}) \leq e^{Ct \|K\|_{L^\infty} \theta_f / \lambda_f} \left(H_N(f_N^0 | f_0^{\otimes N}) + \alpha_N + \frac{C}{N} \right) \rightarrow 0, \quad \text{as } N \rightarrow \infty,$$

where $\alpha_N = C(\sigma - \sigma_N)^2 / (\sigma \sigma_N)$ if $\sigma > 0$ and $\alpha_N = C \sigma_N$ if $\sigma = 0$.

Hence for any fixed k , the k -marginal $f_{N,k}$ of f_N converges to the k -tensor product of f in L^1 as $N \rightarrow \infty$, i.e.

$$\|f_{N,k} - f^{\otimes k}\|_{L^1} \rightarrow 0, \quad \text{as } N \rightarrow \infty. \quad (1.14)$$

1.5.3 Consequence of the main result

Mean Field Limit. The main theorem above is a *Propagation of Chaos* result but in a stronger form. In particular, any marginal $f_{N,k}$ converges towards $f^{\otimes k}$ in L^1 norm with an explicit rate. Propagation of chaos implies the *classical Mean Field limit*. Firstly note that the 1-particle distribution $f_{N,1}$ converges to f in L^1 .

Secondly, assume that one can obtain solutions to the SDE (1.1) system (at least for a short time independent of N) for almost all initial data. Consider now

a solution to (1.1) with random initial data determined according to the law f_N^0 ; the solution $(X_1(t), V_1(t), \dots, X_N(t), V_N(t))$ is hence random as well (even the deterministic system (1.1) with $\sigma_N = 0$ propagates any initial randomness). Then the empirical measure as defined by (1.7) satisfies $\mu_N(t) \rightarrow f_t$ in law in $\mathcal{P}(\mathcal{P}(\mathbb{D} \times \mathbb{R}^d))$ for $t \leq T$ and also that with probability 1, μ_N will converge to f for the *weak* $*$ topology of measures. We refer to [69, 71, 91, 135] for a more precise presentation of this connection between the various concepts of Mean Field limit.

Some other stronger notions of propagation of chaos have recently been more thoroughly investigated and some of the connections between them elucidated in [83, 116, 117].

Weak-strong argument. Our main results are quite demanding on the expected limit f , in particular through assumption (1.10). They are essentially weak-strong type results: Weak requirements on $f_N(0)$ and K are replaced by strong assumptions on the limit. In Theorem 1 the assumption (1.10) is satisfied if f has Gaussian or any kind of exponential decay: $f \sim e^{-\nu|v|^\alpha}$. In general C^k functions with compact support cannot satisfy (1.10) though Gevrey-like regularity seems to be possible.

Relative entropy turns out to be a very convenient norm for studying the mean field limit. But the major restriction is the existence of smooth solution (and also uniqueness) of the mean field PDE. How to extend the relative entropy method to the case where discontinuity could occur in the limit PDE is critical. But major difficulties might arise.

The validity of the time interval. All the theorems here are really conditional

results: They hold on any time interval $[0, T]$ for which one has existence of appropriate solutions f to the McKean-Vlasov Eq. (1.3). In particular, Prop. 10 guarantees that such a time interval will exist in Theorem 1 but T could be larger than what is given by Prop. 10. One may have $T = +\infty$ for some initial data or if additional regularity is known for K .

1.5.4 Comparison with the literature

The first proofs of the mean field limit for deterministic systems such as (1.1) with $\sigma_N = 0$ were performed in [24, 49, 121] (see also [133]) and for stochastic systems (1.1) and (1.2) in [114] (see also [115, 135]). Those now classical results have introduced the main concepts and questions for the mean field limit and propagation of chaos. They demand that $K \in W^{1,\infty}$ and rely on the corresponding Gronwall estimates for systems of ODEs (extended to infinite dimensional settings).

Classical results of the mean field limit need the kernel $K \in W^{1,\infty}$. One possible way to overcome the singularity is to regularize or truncate the kernel K . Since in many settings (like Poisson kernel), K is only singular at the origin, this leads to working with a smooth K_N s.t. $K_N(x) = K(x)$ for $|x| \geq \varepsilon_N$, ε_N being small parameter which typically vanishes when $N \rightarrow \infty$. The accuracy of the method depends on how small the scale ε_N can be taken; one critical scale is $\varepsilon_N = N^{-1/d}$ which would be the minimal distance in physical space of N particles over a grid.

For Poisson kernels, $K = Cx/|x|^d$, the mean field limit was obtained for particles initially on a regular mesh in [66, 142] for $\varepsilon_N \gg N^{-1/d}$. When the particles

are not initially regularly distributed, propagation of chaos was obtained in [67] but only for $\varepsilon_N \sim (\log N)^{-1}$. Those results were recently improved in [104] with much smaller truncation scales $\varepsilon_N \sim N^{-1/d+\varepsilon}$. See also [102, 103] for more detailed discussion of the derivation of the Vlasov-Poisson (or Vlasov-Maxwell) system.

The only results for deterministic 2nd systems with singular, non-Lipschitz, kernels without truncation are [81] and the more recent [82] for the propagation of chaos. Those require that K satisfies for some $\alpha < 1$

$$|K(x)| \leq \frac{C}{|x|^\alpha}, \quad |\nabla K(x)| \leq \frac{C}{|x|^{\alpha+1}}.$$

Theorem 1 does not require any bound on $|\nabla K|$ but does not allow K to be unbounded either. It is therefore not directly comparable. In fact Theorem 1 is interesting precisely because it introduced a new and unexpected critical scale, $K \in L^\infty$.

Notice that the 2nd system (1.1) has a degenerate stochastic part (there is no diffusion in the x variable) which may in addition vanish at the limit if $\sigma_N \rightarrow 0$. Theorem 1 is *the only result* that we are aware of in such a degenerate setting for non Lipschitz force terms.

Using a quite different method, i.e. viewing the ODEs as a differential inclusion system, the article [30] also deals with some flocking models with rough but bounded influence functions, which shall be compared to our assumption that $K \in L^\infty$. We also refer to [84] for the Vlasov-Poisson-Fokker-Planck system with Coulomb forces in 1D (hence the forces are bounded) and [80] for the earlier Dobrusin type estimate [80] for the 1D Vlasov-Poisson system.

The relative entropy method is widely used in the context of “diffusion limit” or

“scaling/hydrodynamic limit” context, see for instance [143] and earlier the entropy method [74]. It has also recently been applied to SDEs, for example in [59]. But to our best knowledge, it is the first time to be applied in the mean field limit in our work [93, 95].

1.6 Relative entropy estimates: the need of combinatorics

Our results show relative entropy *is* the right norm to obtain the mean field limit. The proof of our main results relies on the study of the evolution of the relative entropy $H_N(t) := H_N(f_N(t)|f_t^{\otimes N})$. Since initially $H_N(0)$ is small ($H_N(0) \rightarrow 0$ as $N \rightarrow \infty$), to make the relative entropy method work, we only need to show $\frac{d}{dt}H_N(t) \leq o(1)$ when $N \rightarrow \infty$.

In this chapter, we only prove an special case of Theorem 1: We focus on the deterministic case $\sigma_N = \sigma = 0$ and assume a stronger assumption for the limit f , that is

$$\nabla_v \log f \in L^\infty. \tag{1.15}$$

All other assumptions follow exactly Theorem 1. The complete proof of Theorem 1 will appear in Chapter 3.

1.6.1 An intuitive example

In order to illustrate under what conditions we can expect that $\frac{d}{dt}H_N(t) \leq o(1)$ when N is large, we consider the following more general questions.

Consider the following two PDEs for $g_N(t) \in \mathcal{P}(E^N)$ and $h_N(t) \in \mathcal{P}(E^N)$

$$\partial_t g_N + L_N g_N = 0 \quad (1.16)$$

and

$$\partial_t h_N + L_N h_N = Q_N h_N, \quad (1.17)$$

where iL_N is a self-adjoint linear operator, while $\sup_N \|Q_N\|_{L^\infty} \leq C < \infty$.

Under this setting, one has

Lemma 2 *Assume that the system (1.16) dissipates the entropy*

$$\int_{E^N} g_N(t) \log g_N(t) \, dZ \leq \int_{E^N} g_N^0 \log g_N^0 \, dZ$$

and the system (1.17) with $\sup_N \|Q_N\|_{L^\infty} < \infty$ admits a strong solution $h_N(t)$ with corresponding smooth initial condition h_N^0 for $t \in [0, T]$, $T > 0$. Then for $t < T$,

$$\frac{d}{dt} \|g_N - h_N\|_{L^1} \leq \int_{E^N} |Q_N| h_N \, dZ \leq \sup_N \|Q_N\|_{L^\infty}, \quad (1.18)$$

and

$$\frac{d}{dt} \frac{1}{N} \int_{E^N} g_N \log \frac{g_N}{h_N} \, dZ \leq -\frac{1}{N} \int_{E^N} g_N Q_N \, dZ \leq \frac{\sup_N \|Q_N\|_{L^\infty}}{N}. \quad (1.19)$$

Proof For the L^1 distance, one has

$$\partial_t |g_N - h_N| + L_N |g_N - h_N| \leq |Q_N| h_N.$$

Taking integrals on both sides and then integration by parts gives

$$\frac{d}{dt} \|g_N - h_N\|_{L^1} \leq \int |Q_N| h_N \, dZ \leq \|Q_N\|_{L^\infty}.$$

For the relative entropy, since we have already scaled it with the factor $1/N$, we expect its time derivative is in the order $1/N$, provided that $\sup_N \|Q_N\|_{L^\infty} < \infty$. Since g_N dissipates the entropy, written in a formal way, that is

$$(\partial_t + L_N)(g_N \log g_N) \leq 0.$$

Since h_N is a strong solution, one has

$$\partial_t \log h_N + L_N \log h_N = Q_N.$$

Hence the evolution of the relative entropy can be computed formally as

$$\begin{aligned} \frac{d}{dt} \int g_N \log \frac{g_N}{h_N} dZ &= \frac{d}{dt} \int g_N \log g_N - \int \partial_t g_N \log h_N dZ - \int g_N \partial_t \log h_N \\ &\leq - \int g_N (\partial_t + L_N) \log h_N = - \int g_N Q_N \end{aligned}$$

that is

$$\frac{d}{dt} \frac{1}{N} \int g_N \log \frac{g_N}{h_N} dZ \leq - \frac{1}{N} \int g_N Q_N dZ.$$

Under the assumption $\sup_N \|Q_N\|_{L^\infty} < \infty$, one obtains (1.19). This completes the proof. \square

Consequence of Lemma 2. We can conclude that given $\sup_N \|Q_N\|_{L^\infty} < \infty$ in (1.17),

$$\frac{1}{N} \int_{E^N} g_N^0 \log \frac{g_N^0}{h_N^0} \leq C/N \implies \frac{1}{N} \int_{E^N} g_N(t) \log \frac{g_N(t)}{h_N(t)} \leq C/N$$

for $t \in [0, T]$. Usually T cannot be arbitrarily large since the strong solution h_N might cease to be smooth (develop shock for instance) after a period of time.

If we can establish (1.19) in Lemma 2 for $g_N = f_N$, a weak solution to the Liouville equation (1.6) and $h_N = \bar{f}_N = f^{\otimes N}$, where f is the strong solution to the

Vlasov equation (1.3), then combining with Prop. 2, the relative entropy estimate (1.19) implies propagation of chaos. In particular, in the following of this chapter let us write

$$H_N(t) := H_N(f_N(t)|\bar{f}_N(t)), \quad H_k(t) := H_k(f_{N,k}(t)|f^{\otimes k}(t))$$

in short. If $\frac{d}{dt}H_N(t) \leq \frac{C}{N}$ holds true as (1.19) in Lemma 2, then combining with Prop. 2, we would obtain

$$\|f_{N,k}(t) - f_t^{\otimes k}\|_{L^1} \leq \sqrt{2kH_k(t)} \leq \sqrt{2kH_N(t)} \rightarrow 0$$

as $N \rightarrow \infty$ given the asymptotic initial condition $H_N(0) \rightarrow 0$ as $N \rightarrow \infty$. The relative entropy estimate for the joint law f_N can be transferred to the counterpart of its marginals $f_{N,k}$, which then implies the propagation of chaos. Next subsection is devoted to show why we can expect $\frac{d}{dt}H_N(t) \leq \frac{C}{N}$ even though the essential $\sup_N \|Q_N\|_{L^\infty} < \infty$ is not satisfied.

Why L^1 norm does not work. The L^1 norm does not work well simply because it does not have the tensorization properties as for the entropy and the relative entropy, in particular Prop. 1. Set $g_N = f_N$ and $h_N = \bar{f}_N = f^{\otimes N}$. Then under the assumption that $\sup_N \|Q_N\|_{L^\infty} < \infty$, the L^1 estimate (1.18) in Lemma 2 becomes

$$\frac{d}{dt}\|f_N - \bar{f}_N\|_{L^1} \leq \int |Q_N| \bar{f}_N dZ \leq \sup_N \|Q_N\|_{L^\infty} < C, \quad (1.20)$$

where C is a universal constant.

However, we do not have a counterpart of Prop. 1 for L^1 (or L^p) distance. In particular, for $f_N \in \mathcal{P}_{\text{Sym}}(E^N)$ and $f_{N,1}$ its 1-marginal, we *cannot* control $\|f_{N,1} - f\|_{L^1}$ by any normalization of $\|f_N - \bar{f}_N\|_{L^1}$ in the form $\frac{1}{\lambda_N}\|f_N - \bar{f}_N\|_{L^1}$

with $\lambda_N \rightarrow \infty$ as $N \rightarrow \infty$. Indeed, by choosing $f_N = g^{\otimes N}$, where $g \in \mathcal{P}(E)$ and $\|g - f\|_{L^1} = \frac{1}{2}$, one has

$$\frac{1}{2} = \|g - f\|_{L^1} = \|f_{N,1} - f\|_{L^1} \leq \|f_N - \bar{f}_N\|_{L^1} \leq 2,$$

while $\frac{1}{\lambda_N} \|f_N - \bar{f}_N\|_{L^1} \rightarrow 0$ as $N \rightarrow \infty$.

As a result of Lemma 2, in particular (1.18) or (1.20), even though

$$\|f_N^0 - \bar{f}_N^0\|_{L^1} \leq C \implies \|f_N(t) - \bar{f}_N(t)\|_{L^1} \leq C$$

for any $t \leq T$, we cannot recover any useful information for the marginals, except for the trivial bound

$$\|f_{N,k}(t) - f_t^{\otimes k}\|_{L^1} \leq \|f_N(t) - \bar{f}_N(t)\|_{L^1} \leq C$$

which tells us nothing about the mean field limit or propagation of chaos.

1.6.2 The need of combinatorics

By Prop. 2, it seems tempting to apply Lemma 2 to the Liouville equation (1.6) and a variant Liouville equation for $\bar{f}_N = f^{\otimes N}$. Let us first write down two concrete examples for systems (1.16) and (1.17).

We now take L_N as the Liouville operator of (1.6) with $\sigma_N = 0$. That is

$$L_N = \sum_{i=1}^N v_i \cdot \nabla_{x_i} + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i}^N K(x_i - x_j) \cdot \nabla_{v_i}.$$

Therefore the Liouville equation (1.6) with $\sigma_N = 0$ can be written shortly as

$$\partial_t f_N + L_N f_N = 0.$$

Recall $\bar{f}_N(t, Z) = \prod_{i=1}^N f(t, z_i)$. It is easy to check that \bar{f}_N solves

$$\partial_t \bar{f}_N + \sum_{i=1}^N v_i \cdot \nabla_{x_i} \bar{f}_N + \sum_{i=1}^N K \star \rho(x_i) \cdot \nabla_{v_i} \bar{f}_N = 0.$$

Or using the Liouville operator L_N ,

$$\partial_t \bar{f}_N + L_N \bar{f}_N = R_N \bar{f}_N \tag{1.21}$$

where we define

$$R_N = \frac{1}{N} \sum_{i,j=1}^N \nabla_{v_i} \log f(x_i, v_i) \cdot \{K(x_i - x_j) - K \star \rho(x_i)\}. \tag{1.22}$$

The R_N is an example of Q_N as in Lemma 2.

However, the basic but important assumption $\sup_N \|Q_N\|_{L^\infty} \leq C$ does not hold for our R_N defined in (1.22). Indeed, a priori $R_N = O(N)$ since it is in a double summation form but only normalized by $1/N$. This indicates the need of combinatorics which is the main technical difficulty of our method.

Recalling the calculations in Lemma 2, one has

$$H_N(t) \leq H_N(0) - \frac{1}{N} \int_0^t \int_{E^N} f_N R_N \, dZ \, ds \tag{1.23}$$

where R_N is the double summation defined in (1.22). We can complete the relative estimate (1.23) by showing that

$$-\frac{1}{N} \int_{E^N} f_N R_N \, dZ \leq \frac{C}{N}.$$

However, a priori $R_N = O(N)$ indicates that this is only possible if we can have certain cancellation rules. Recall that in the classical Laws of Large Numbers, the independence of N random variables (or their joint distribution is tensorized)

plays a key rule in obtaining the very famous convergence rate $1/\sqrt{N}$. We thus seek to replace f_N by a tensorized object, hopefully $\bar{f}_N = f^{\otimes N}$. We have the following lemma

Lemma 3 (Evolution of the relative entropy) *Consider any weak solution f_N to the Liouville equation (1.6) and f the strong solution to the Vlasov (1.3) with initial data f_N^0 and f_0 respectively. Recall that $\bar{f}_N(t) = f_t^{\otimes N}$. Then the evolution of the relative entropy $H_N(t) = H_N(f_N(t)|\bar{f}_N(t))$ reads*

$$H_N(t) \leq H_N(0) + \frac{1}{\nu} \int_0^t H_N(s) ds + \frac{1}{\nu} \frac{1}{N} \int_0^t \int_{E^N} \bar{f}_N \exp(\nu|R_N|) dZ ds,$$

where R_N is defined in (1.22) and ν is an positive parameter.

Proof Applying the Fenchel's inequality to the function $u(x) = x \log x$, that is for any $x, y \geq 0$, $x \log x \leq x \log y + e^{y-1}x$, we obtain

$$\begin{aligned} -\frac{1}{N} \int f_N R_N dZ &\leq \frac{1}{\nu} \frac{1}{N} \int \bar{f}_N \left(\frac{f_N}{\bar{f}_N} \nu |R_N| \right) \\ &\leq \frac{1}{\nu} \frac{1}{N} \int f_N \log \frac{f_N}{\bar{f}_N} dZ + \frac{1}{\nu} \frac{1}{N} \int \bar{f}_N \exp(\nu|R_N|) dZ. \end{aligned}$$

Therefore

$$-\frac{1}{N} \int_{E^N} f_N R_N dZ \leq \frac{1}{\nu} H_N(f_N|\bar{f}_N) + \frac{1}{\nu} \frac{1}{N} \int_{E^N} \bar{f}_N \exp(\nu|R_N|) dZ.$$

Combining with (1.23) or integrating over time t completes the proof. \square

We have changed the reference measure f_N to a tensorized one \bar{f}_N but as a compensation $|R_N|$ becomes $\exp(|R_N|)$.

1.7 Combinatorics: Laws of Large Numbers

By the previous Lemma 3, we can conclude the relative entropy estimate by Gronwall's inequality if we can show under proper assumptions

$$\int \bar{f}_N \exp(\nu |R_N|) dZ \leq C < \infty$$

where C does not depend on N . This is the place where the combinatorics, in spirit of Laws of Large Numbers, comes in. In this section, we will establish this main estimate under a stronger assumption $\nabla_v \log f \in L^\infty$ or (1.15), concluding the proof of this special case of Theorem 1.

1.7.1 Classical Laws of Large Numbers

Recall the very famous Laws of Large Numbers in probability. To make it adaptable to our framework, we consider the L^2 , L^4 or more general L^{2k} convergence of the experimental average Ξ_N towards the mean value μ .

Proposition 3 (Law of Large Numbers in L^2) *Assume that a sequence of independent and identically distributed (real) random variable ξ, ξ_1, ξ_2, \dots in $L^2(\Omega, P, \mathcal{F})$ with the same law $g \in \mathcal{P}(\mathbb{R})$. Define $\Xi_N = \frac{1}{N} \sum_{i=1}^N \xi_i$. Then*

$$\|\Xi_N - \int_{\mathbb{R}} xg(x) dx\|_{L^2(\Omega, P)} \leq \frac{C}{\sqrt{N}} \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Proof For simplicity we let $\mu = E\xi = \int xg(x) dx = 0$. Then one simply calculates

$$E(\Xi_N)^2 = \frac{1}{N^2} \sum_{i_1, i_2=1}^N E\xi_{i_1}\xi_{i_2} = \frac{1}{N^2} \sum_{i=1}^N E\xi_i^2 = \frac{1}{N} E\xi^2 \rightarrow 0 \quad (1.24)$$

as $N \rightarrow \infty$. The crucial fact here is the independence of each pair (ξ_{i_1}, ξ_{i_2}) with $i_1 \neq i_2$, leading to the vanishing of off-diagonal terms

$$E\xi_{i_1}\xi_{i_2} = E\xi_{i_1} E\xi_{i_2} = 0, \quad \text{if } i_1 \neq i_2. \quad (1.25)$$

Therefore one has the L^2 convergence with the very famous convergence rate $1/\sqrt{N}$.

□

Let us translate the expectation into the integral against the joint law $\bar{g}_N = g^{\otimes N}$ of ξ_1, \dots, ξ_N . Indeed, the cancellation rule (1.25) reduces to the following

$$\int_{\mathbb{R}^N} x_{i_1} x_{i_2} \bar{g}_N(X) \, dX = \left(\int_{\mathbb{R}} x_{i_1} g(x_{i_1}) \, dx_{i_1} \right) \left(\int_{\mathbb{R}} x_{i_2} g(x_{i_2}) \, dx_{i_2} \right) = 0$$

provided that $i_1 \neq i_2$.

In this trivial case, we have N^2 multi-indices $I_2 = (i_1, i_2)$ in the summation (1.24) while only N diagonal terms $I_2 = (i, i)$ will *not* vanish after taking expectation or integral against the tensorized joint law \bar{g}_N . Note that $N = \sqrt{N^2}$, which corresponds to the critical convergence rate $1/\sqrt{N}$.

Of course here the combinatorics is almost trivial, but the counterpart for the convergence in L^{2k} needs more advanced combinatorics, which motivates our work.

Proposition 4 *In the setting in Prop. 3, we further assume that*

$$E|\xi|^i = \int_{\mathbb{R}} |x|^i g(x) \, dx \leq 1 \quad (1.26)$$

for $i = 1, 2, \dots$. Then for any integer k , one has the L^{2k} convergence

$$\|\Xi_N - \int_{\mathbb{R}} xg(x) \, dx\|_{L^{2k}(\Omega, P)} \leq \frac{C\sqrt{k}}{\sqrt{N}}.$$

Proof The assumption (1.26) for $g \in \mathcal{P}(\mathbb{R})$ trivially holds true if the support $\text{supp}(g) \subset [0, 1]$. As usual, we assume the mean $\mu = 0$ for simplicity. We can expand the expectation as

$$E(\Xi)^{2k} = \frac{1}{N^{2k}} \sum_{1 \leq i_1, \dots, i_{2k} \leq N} E(\xi_{i_1} \cdots \xi_{i_{2k}}). \quad (1.27)$$

We summarize the cancellation rule as follows. Any term in the summation with index $I_{2k} = (i_1, \dots, i_{2k})$ vanishes, *i.e.*

$$E(\xi_{i_1} \cdots \xi_{i_{2k}}) = \int_{\mathbb{R}^N} x_{i_1} \cdots x_{i_{2k}} \bar{g}_N \, dx_1 \cdots dx_N = 0$$

provided that there exists i_α such that

$$i_\alpha \notin \{i_1, \dots, i_{\alpha-1}, i_{\alpha+1}, \dots, i_{2k}\}.$$

Consequently, the multi-indices I_{2k} for non-vanishing terms all belong to the effective set $\mathcal{E}_{N,2k}$, which is defined as

$$\mathcal{E}_{N,2k} = \{I_{2k} = (i_1, \dots, i_{2k}) \mid 1 \neq a_\nu := |\{1 \leq \alpha \leq 2k \mid i_\alpha = \nu\}|, \nu = 1, \dots, N\}.$$

In other words, any multi-index $I_{2k} \in \mathcal{E}_{N,2k}$ has no *singleton*: any integer in I_{2k} must be repeated.

Hence one has

$$E(\Xi_N)^{2k} = \frac{1}{N^{2k}} \sum_{I_{2k} \in \mathcal{E}_{N,2k}} E(\xi_{i_1} \cdots \xi_{i_{2k}}).$$

For each non-vanishing term, one has the trivial estimate

$$E(\xi_{i_1} \cdots \xi_{i_{2k}}) = E(\xi_1^{a_1} \cdots \xi_N^{a_N}) \leq (E|\xi|^{a_1}) \cdots (E|\xi|^{a_N}) \leq 1,$$

where a_ν is the multiplicity of the integer ν in I_{2k} .

Applying the Lemma 5.5 in Chapter 5, we can bound the cardinality of $\mathcal{E}_{N,2k}$ as

$$|\mathcal{E}_{N,2k}| \leq ke^k N^k k^k.$$

Combining all the estimates above, one finally obtains

$$E(\Xi_N)^{2k} \leq \frac{1}{N^{2k}} \sum_{I_{2k} \in \mathcal{E}_{N,2k}} 1 \leq \frac{1}{N^k} (ke^k k^k).$$

Therefore, we can still get the usual convergence rate $1/\sqrt{N}$. \square

In this advanced case, the total number of multi-indices I_{2k} is N^{2k} , while the cardinality of the effective set $\mathcal{E}_{N,2k}$ is in the order of $C^k k^k N^k$ which is again roughly $\sqrt{N^{2k}}$. Here the effective set $\mathcal{E}_{N,2k}$ can be regarded as a set of *General Diagonal Multi-indices*.

1.7.2 Combinatorics for double multi-indices

Now we go back to show that under proper assumptions

$$\int \bar{f}_N \exp(\nu |R_N|) dZ \leq C < \infty$$

where C does not depend on N . Our main estimates as in Theorem 5, Theorem 6 and Theorem 7 are all written in this form.

We here present a basic main estimate result, which can be proved by some similar but advanced combinatorics arguments in the spirit of Laws of Large Numbers, in particular Prop. 4.

Theorem 2 Assume that $\|K\|_{L^\infty} < \infty$ and $\|\nabla_v \log f\|_{L^\infty} < \infty$. For parameter $\nu > 0$ with $\nu\|K\|_{L^\infty}\|\nabla_v \log f\|_{L^\infty} < 1/C$, one has

$$\int_{E^N} \bar{f}_N \exp(\nu|R_N|) dZ \leq C < \infty$$

where C is a universal constant and $\bar{f}_N = f^{\otimes N}$ and R_N is defined in (1.22).

The preparation of the proof of Theorem 2. For simplicity we set $\nu = 1$. The classical Laws of Large Numbers such as Prop. 3 and Prop. 4 cannot directly apply to the exponential function. Therefore, by Taylor expansion for $y = \exp(x)$ and Cauchy-Schwarz inequality, one expand the above integral as a series

$$\int_{E^N} \bar{f}_N \exp(|R_N|) dZ \leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} \int \bar{f}_N |R_N|^{2k} dZ. \quad (1.28)$$

If we can show the above series converges, the job is done. There is no better way but to expand R_N by its definition. Therefore the k -th term

$$\frac{1}{(2k)!} \int_{E^N} \bar{f}_N |R_N|^{2k} dZ$$

can be expanded as

$$\frac{1}{(2k)!} \frac{1}{N^{2k}} \int \sum_{1 \leq i_1, j_1 \leq N} \cdots \sum_{1 \leq i_{2k}, j_{2k} \leq N} (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N dZ, \quad (1.29)$$

where we define

$$F_i = \nabla_{v_i} \log f(x_i, v_i), \quad \delta K^{i,j} = K(x_i - x_j) - K \star \rho(x_i).$$

Under the assumption $\|\nabla_v \log f\|_{L^\infty} < \infty$, the k -th term can be estimated with

$$\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ \leq \frac{1}{(2k)!} N^{2k} (2\|K\|_{L^\infty}\|\nabla_v \log f\|_{L^\infty})^{2k}. \quad (1.30)$$

We divide the estimates for (1.28) into two cases. For the case when $k \ll N$ (actually we will choose $4k \leq N$), the right hand side of (1.30) will blow up if we fix k but let $N \rightarrow \infty$. Therefore in this case we should go back to (1.29) and make use of the combinatorics in spirit of Prop. 4 to complete the estimates. For the case when k is large (actually we choose $4k > N$), the trivial bound (1.30) is sufficient.

Now Theorem 2 is a natural consequence of the following two propositions.

Proposition 5 (The case $4k > N$) *In the expanding (1.28), for $4k > N$, the k -th term can be estimated as*

$$\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ \leq (4e \|K\|_{L^\infty} \|\nabla_v \log f\|_{L^\infty})^{2k}.$$

Proof of Prop. 5 This proposition is a direct consequence of the trivial bound (1.30). Indeed, since $4k > N$,

$$\frac{1}{(2k)!} N^{2k} \leq (2k)^{-2k} e^{2k} (4k)^{2k} = (2e)^{2k},$$

where we use the inequality $p^p \leq p!e^p$ in Chapter 5 as a consequence of Stirling's formula. Inserting it back to (1.30) completes the proof. \square

Proposition 6 (The case $4k \leq N$) *In the expanding (1.28), for $4k \leq N$, the k -th term can be estimated as*

$$\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ \leq (C \|K\|_{L^\infty} \|\nabla_v \log f\|_{L^\infty})^{2k},$$

where C is a universal constant which does not depend on N .

Proof of Theorem 2 is trivial now: by assuming Prop. 5 and Prop. 6 and setting ν (or $\|K\|_{L^\infty}\|\nabla_v \log f\|_{L^\infty}$) be small enough, the series in (1.28) will converge. In the following, we focus on the proof of Prop. 6, which is the place where combinatorics plays a crucial role.

In the case $4k \leq N$, we shall go back to the complete expanding (1.29) and try to find the cancellation rules. More careful treatment will be given in Chapter 6 and Chapter 7. Here we give a general framework under which Theorem 2 or Prop. 6 shall be expected.

The general cancellation rules for the 2nd order case is simple. The exact formulation is Lemma 17 in Chapter 6. Here we summarize the essences as the following lemma

Lemma 4 *Assume that $4k \leq N$. Consider double multi-indices (I_{2k}, J_{2k}) with $I_{2k} = (i_1, \dots, i_{2k})$ and $J_{2k} = (j_1, \dots, j_{2k})$, where each component is chosen from $\{1, 2, \dots, N\}$. Therefore the term with multi-indices (I_{2k}, J_{2k}) in the expanding (1.29) will vanish provided that one of the following statements is satisfied:*

- 1) *there exists one i_α , such that $i_\alpha \notin \{i_1, \dots, i_{\alpha-1}, i_\alpha + 1, \dots, i_{2k}\}$;*
- 2) *there exists one j_β , such that $j_\beta \notin \{i_1, i_2, \dots, i_{2k}\} \cup \{j_1, \dots, j_{\beta-1}, j_{\beta+1}, \dots, j_{2k}\}$.*

Lemma 4 can be easily proved by Fubini's Theorem and the following two type cancellation rules (2.15)

$$\int F_i \cdot \delta K^{i,j} f(x_i, v_i) dx_i dv_i = 0, \quad \int F_i \cdot \delta K^{i,j} f(x_j, v_j) dx_j dv_j = 0.$$

The complete proof will be given in Chapter 6. Since a typical non-vanishing term

in (1.29) can be estimated as

$$\int (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N \, dZ \leq (2\|K\|_{L^\infty} \|\nabla_v \log f\|_{L^\infty})^{2k}, \quad (1.31)$$

the crucial part is to count how many terms will *not* vanish.

We need to count those double multi-indices (I_{2k}, J_{2k}) such that neither conditions in Lemma 4 are satisfied. Consequently, I_{2k} must have no singleton, *i.e.* $I_{2k} \in \mathcal{E}_{N, 2k}$. And for fixed $I_{2k} \in \mathcal{E}_{N, 2k}$, J_{2k} shall be chosen according to I_{2k} . In particular, J_{2k} belongs to the set defined as

$$\mathcal{P}_{N, 2k}^{I_{2k}} := \left\{ J_{2k} \in \mathcal{T}_{N, 2k} \mid \begin{array}{l} \text{either for all } 1 \leq \nu \leq 2k, j_\nu \in \{i_1, \dots, i_{2k}\}; \\ \text{or for any } \nu \text{ such that } j_\nu \notin \{i_1, \dots, i_{2k}\}, \\ \exists \nu' \neq \nu, \text{ such that } j_\nu = j_{\nu'}. \end{array} \right\}.$$

By Lemma 18 in Chapter 6, we have the right order for the cardinality for $\mathcal{P}_{N, 2k}^{I_{2k}}$, that is

$$|\mathcal{P}_{N, 2k}^{I_{2k}}| \leq C^k k^k N^k,$$

where C is a universal constant.

Therefore one reaches the following lemma

Lemma 5 *In the expanding (1.29), the number of double multi-indices (I_{2k}, J_{2k}) for non-vanishing terms can be bounded by $C^k k^{2k} N^{2k}$, where C is a universal constant.*

Proof. The proof is simple. Since the number of all possible choices of I_{2k} is bounded by $C^k k^k N^k$ by Lemma 5.5, and for fixed I_{2k} , the choices of J_{2k} is also bounded by $C^k k^k N^k$, the multiplication principle of counting will complete the proof. \square

The total number of all the double multi-indices (I_{2k}, J_{2k}) is N^{4k} . The total number of the non-vanishing multi-indices for (1.29) is in the order $C^k k^{2k} N^{2k}$, which is roughly $\sqrt{N^{4k}}$. This agrees with the case in classical Laws of Large Numbers, in particular Prop. 4.

Now we can prove Prop. 6 by assuming Lemma 5.

Proof of Prop. 6 First, by Lemma 4 and Lemma 5, the expanding (1.29) can be reduced to

$$\frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{I_{2k} \in \mathcal{E}_{N,2k}, J_{2k} \in \mathcal{P}_{N,2k}^{I_{2k}}} \int_{E^N} (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N \, dZ.$$

Combining with the trivial bound (1.31), it can be further bounded by

$$\frac{1}{(2k)!} \frac{1}{N^{2k}} C^k k^{2k} N^{2k} (2\|K\|_{L^\infty} \|\nabla_v \log f\|_{L^\infty})^{2k}.$$

Applying $p^p \leq p!e^p$ to $p = 2k$ will give the final form in Prop. 6. \square

More complete treatments of the combinatorics argument would appear in Chapter 5 –Chapter 7. The stronger assumptions (1.26) for the Law of Large Numbers in L^{2k} space agree with similar strong moment assumptions for the limit law f in our problem: here we assume a stronger assumption $\nabla_v \log f \in L^\infty$, which can be relaxed to the assumption (1.10) or an equivalent moment assumption $\sup_{p \geq 1} \frac{M_p}{p} = \sup_{p \geq 1} \frac{\|\nabla_v \log f\|_{L^p(df)}}{p} < \infty$. See the discussion in Section 3.2.

Chapter 2: Main results for the 1st order system and the comparison with the literature

From this chapter, we deal with the technical issues of the mean field limit in our framework in a complete way. Our main result for the 1st order system will be presented and compared to the existing literature. In the last part, some related problems will also be discussed.

2.1 Existence of weak solutions of the Liouville equations

As illustrated in the previous chapter, we are working at the level of the Liouville equations. Recall that for the 2nd order system (1.1), the evolution of the joint law $f_N(t) = \text{Law}(X_1(t), V_1(t), \dots, X_N(t), V_N(t))$ is given by the Liouville equation

$$\partial_t f_N + \sum_{i=1}^N v_i \cdot \nabla_{x_i} f_N + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i}^N K(x_i - x_j) \cdot \nabla_{v_i} f_N = \sigma_N \sum_{i=1}^N \Delta_{v_i} f_N, \quad (2.1)$$

as first stated in (1.6).

Similarly, for the 1st order system (1.2), the joint law/distribution

$$\rho_N(t, X) = \text{Law}(X_1(t), \dots, X_N(t)) \in \mathcal{P}(\mathbb{T}^d)$$

solves the corresponding Liouville equation

$$\partial_t \rho_N + \sum_{i=1}^N \operatorname{div}_{x_i} (\rho_N F(x_i)) + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \operatorname{div}_{x_i} (\rho_N K(x_i - x_j)) = \sigma_N \sum_{i=1}^N \Delta_{x_i} \rho_N. \quad (2.2)$$

Again the Liouville equation (2.2) can be derived from Itô's formula (see [88]), applied to $\phi(X_1(t), \dots, X_N(t))$ directly.

The transition from the original particle systems (1.1) and (1.2) to their corresponding Liouville equations (2.1) and (2.2) respectively enjoys an obvious advantage: We can now consider more general kernels K . Indeed, if we work at the level of SDEs, the Itô's theory on the well-posedness for the Cauchy problem of (1.1) or (1.2) requires that K be at least locally Lipschitz. However, very weak assumptions on K (for instance $K \in L^\infty$ in the 2nd order case) can guarantee the existence of weak solutions to the Liouville equations (2.1) and (2.2).

2.1.1 The 2nd order case

For the completeness we first present the existence of the weak solutions to the Liouville equation (2.1) in the 2nd order case.

Proposition 7 (Existence of weak solutions to the Liouville equation (2.1))

Assume that $K \in L^\infty$ and that the initial data $f_N^0 \geq 0$ satisfies the following assumptions

$$\begin{aligned} i) \quad & f_N^0 \in L^1((\mathbb{D} \times \mathbb{R}^d)^N) \text{ with } \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N^0 \, dZ = 1, \\ ii) \quad & \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N^0 \log f_N^0 \, dZ < \infty, \end{aligned} \quad (2.3)$$

together with the moment assumption

$$iii) \quad \int_{(\mathbb{D} \times \mathbb{R}^d)^N} \sum_{i=1}^N (1 + |x_i|^{2k} + |v_i|^{2k}) f_N^0 \, dZ < \infty, \quad (2.4)$$

for some $k > 0$. Then there exists $f_N \geq 0$ in $L^\infty(\mathbb{R}_+, L^1(\mathbb{D} \times \mathbb{R}^d)^N)$, which is a solution to (2.1) in the sense of distribution and satisfies

$$\begin{aligned}
i) \quad & \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N(t, Z) \, dZ = 1, \quad \text{for a.e. } t, \\
ii) \quad & \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N(t, Z) \log f_N(t, Z) \, dZ + \sigma_N \int_0^t \int_{(\mathbb{D} \times \mathbb{R}^d)^N} \frac{|\nabla_V f_N(s, Z)|^2}{f_N(s, Z)} \, dZ \, ds \\
& \leq \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N^0 \log f_N^0 \, dZ, \quad \text{for a.e. } t, \\
iii) \quad & \sup_{t \in [0, T]} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} \sum_{i=1}^N (1 + |x_i|^{2k} + |v_i|^{2k}) f_N(t, Z) \, dZ < \infty, \quad \text{for any } T < \infty.
\end{aligned} \tag{2.5}$$

For fixed N , we can find weak solutions to (2.1) which dissipate the entropy and propagate the moment provided that the initial entropy and moment are finite. More detailed discussions will appear in Section 2.1.3.

2.1.2 The 1st order case

Before stating the existence result for the law $\rho_N(t)$ solving the Liouville equation (2.2), let us fix some notations first.

The notation $K \in W^{-1, \infty}$ means that there exists a $d \times d$ matrix-valued function $V = (V_{hl})_{1 \leq h, l \leq d}$ defined on $\mathbb{D} = \mathbb{T}^d$ with

$$\|K\|_{W^{-1, \infty}} := \|V\|_{L^\infty} = \sup_{1 \leq h, l \leq d} \|V_{hl}\|_{L^\infty} < +\infty$$

such that

$$K = (K^1, \dots, K^d) \quad \text{and} \quad K^h = \sum_{l=1}^d \partial_{x^l} V_{hl}, \quad h = 1, \dots, d. \tag{2.6}$$

Sometimes we also write (2.6) as $K = \operatorname{div} V$ for simplicity.

Proposition 8 (Existence of weak solutions of the Liouville equation (2.2))

Assume that the underlying domain is $\mathbb{D} = \mathbb{T}^d$. Assume that $\operatorname{div}F \in L^\infty$ and that K permits a decomposition $K = K_1 + K_2$ where $K_2 \in L^\infty$, $K_1 = \operatorname{div}V$ with $V = (V_{hl})_{1 \leq h, l \leq d}$ is a $d \times d$ matrix valued function such that

$$\sup_{1 \leq h, l \leq d} |V_{hl}(x)| \leq C \sqrt{|\log |x||}, \quad \text{for any } x \in \mathbb{D}.$$

We further assume that the initial data $\rho_N^0 \geq 0$ satisfies the following assumptions

$$\begin{aligned} i) \quad & \rho_N^0 \in L^1(\mathbb{D}^N) \text{ with } \int_{\mathbb{D}^N} \rho_N^0 \, dX = 1, \\ ii) \quad & \int_{\mathbb{D}^N} \rho_N^0 \log \rho_N^0 \, dX < \infty. \end{aligned} \tag{2.7}$$

Then there exists $\rho_N \geq 0$ in $L^\infty(\mathbb{R}_+, L^1(\mathbb{D}^N))$, which is a solution to (2.2) in

the sense of distribution and satisfies

$$\begin{aligned} i) \quad & \int_{\mathbb{D}^N} \rho_N(t, X) \, dX = 1, \quad \text{for a.e. } t, \\ ii) \quad & \int_{\mathbb{D}^N} \rho_N(t, X) \log \rho_N(t, X) \, dX + \sigma_N \int_0^t \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N(s, X)|^2}{\rho_N(s, X)} \, dX \, ds \\ & \leq \int_{\mathbb{D}^N} \rho_N^0 \log \rho_N^0 \, dX - \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \int_0^t \int_{\mathbb{D}^N} \rho_N(s, X) (\operatorname{div}_x K_1)(x_i - x_j) \, dX \, ds \\ & + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \int_0^t \int_{\mathbb{D}^N} \nabla_{x_i} \rho_N K_2(x_i - x_j) \, dX \, ds \\ & + \sum_{i=1}^N \int_0^t \int_{\mathbb{D}^N} \rho_N (\operatorname{div}_x F)(x_i) \, dX \, ds, \quad \text{for a.e. } t. \end{aligned} \tag{2.8}$$

2.1.3 Remarks on Proposition 7 and Proposition 8

We omit the proofs of Proposition 7 and Proposition 8. It is straightforward by approximating K by a sequence of smooth kernels K_ϵ and then passing to limit. The weak solutions we used here are those dissipating the entropy. And the dissipation of the entropy, *i.e.* the Fisher information (see Appendix A), will be helpful in the 1st order case to obtain the mean field limit for $K \in W^{-1, \infty}$ for instance.

We do not have uniqueness in Proposition 7 and Proposition 8: There could be several such solutions. Even though we do not have uniqueness here, any weak solution $f_N(t, Z)$ or $\rho_N(t, X)$ prescribed in the above propositions will be close to the limit f_t or ρ_t in the scaled relative entropy sense as illustrated in Chapter 1 when N is large: Main results on propagation of chaos do not rely on the specific choice of weak solutions f_N or ρ_N .

Uniqueness and in general the well-posedness of the Cauchy problem for transport equations like (2.1) or (2.2) with $\sigma_N = 0$ are usually handled through the theory of renormalized solutions as introduced in [47] and improved in [4] (see also [6] and [44] for a good introduction).

Renormalized solutions not only give well-posedness to transport equations like (2.1) with $\sigma_N = 0$ but also provide the existence of a flow to the corresponding ODE system thus giving a meaning to the ODE system (1.1) with $\sigma_N = 0$ for instance.

In the case $\sigma_N = 0$, the general setting of [4] would require $K \in BV$. That may sometimes be improved for 2nd order systems like (1.1), see [22, 23, 38, 92]. However for a system in large dimension like (1.1), it seems out of reach to obtain renormalized solutions or a well posed flow with only $K \in L^\infty$. Therefore in that case, it is actually *critical* to be able to work with only weak solutions to (2.2).

If one had a full diffusion, that is $\Delta_x f_N + \Delta_v f_N$ in the Liouville equation (2.1) or $\Delta_x \rho_N$ in (2.2), it would in general be possible to obtain uniqueness together with a flow for the system (1.1) or (1.2) respectively in some sense, see for instance

[38,39,55,106]. Note though that even for $\sigma_N > 0$, the diffusion in (2.1) is degenerate (diffusion only in the v_i variables) so that even for $\sigma_N > 0$, well posedness for the equation (2.1) does not seem easy with only $K \in L^\infty$.

Of course our analysis also applies to more regular interactions K for which it may be possible to have solutions to the SDE systems (1.1) and (1.2) even if only for short times (for instance K is continuous).

2.2 Main results: Mean field limit for the 1st order system

Now we present our main results for the 1st order system (1.2). Note that in the 2nd order system (1.1) or (2.1) the space domain \mathbb{D} can be whole space \mathbb{R}^d or the flat torus \mathbb{T}^d while in the 1st order case (1.2) or (2.2) it can only be $\mathbb{D} = \mathbb{T}^d$ due to the regularity restrictions for the limit law ρ_t .

We have the following propagation of chaos result for the general first order system (1.2).

Theorem 3 (Propagation of chaos for the 1st order system) *Assume that $\operatorname{div} F \in L^\infty$ and that K permits a decomposition $K = K_1 + K_2$, where $K_2 \in L^\infty$ and $K_1 = \operatorname{div} V$, $V = (V_{hl})_{1 \leq h, l \leq d}$ is an anti-symmetric matrix valued function with*

$$\sup_{1 \leq h, l \leq d} |V_{hl}(x)| \leq C \sqrt{|\log |x||}, \quad \text{for any } x \in \mathbb{D} = \mathbb{T}^d. \quad (2.9)$$

We further assume that $\rho(t, x) \in L^\infty([0, T], L_+^\infty(\mathbb{D}) \cap W^{2,p})$ for every $1 \leq p \leq \infty$ solves the macroscopic equation (1.4) with

$$\sup_{t \in [0, T]} \|\nabla_x \log \rho\|_{L^\infty} < \infty, \quad \sup_{t \in [0, T]} \left(\sup_{p \geq 1} \frac{\|R\|_{L^p(\rho dx)}}{p} \right) < \infty, \quad (2.10)$$

where we define

$$R_{hl}(x) = \frac{1}{\rho(x)} \partial_l \partial_h \rho(x), \quad R(x) = \sum_{h,l=1}^d |R_{hl}(x)|. \quad (2.11)$$

Assume that the initial data ρ_N^0 of the Liouville equation satisfies assumptions (2.7)

and

$$\sup_{N \geq 2} \frac{1}{N} \int_{\mathbb{D}^N} \rho_N^0 \log \rho_N^0 \, dZ < \infty, \quad (2.12)$$

as well as

$$H_N(\rho_N^0 | \rho_0^{\otimes N}) = \frac{1}{N} \int_{\mathbb{D}^N} \rho_N^0 \log \left(\frac{\rho_N^0}{\rho_0^{\otimes N}} \right) \, dX \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Then for any corresponding weak solution ρ_N to the Liouville equation as given by Proposition 8, one has for any $t \leq T$,

$$H_N(\rho_N(t) | \rho_t^{\otimes N}) \leq (H_N(\rho_N^0 | \rho_0^{\otimes N}) + \frac{C}{N} + \Lambda_0(\sigma - \sigma_N)^2)^{\exp(-C \int_0^t [1+h(s)] \, ds)}, \quad (2.13)$$

where C is a universal constant and $0 \leq h \in L^1[0, T]$ with $\int_0^t h(s) \, ds = C_t < \infty$, C_t depending on t . Consequently, for any fixed k , the k -marginal $\rho_{N,k}$ of ρ_N converges to the k -tensor product of ρ in L^1 as $N \rightarrow \infty$, i.e.

$$\|\rho_{N,k} - \rho^{\otimes k}\|_{L^1} \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Similar consequences of Theorem 3 would follow the discussion in Section 1.5.3.

For instance we would have the mean field limit etc. We remark in particular that in Theorem 3, the assumption (2.10) is satisfied if the law ρ is comparable to the Lebesgue measure on the torus \mathbb{T}^d and its 1st and 2nd derivatives are all bounded.

The main estimate (2.13) has the very special double exponential rate, which shall be compared to Theorem 1. Theorem 3 indicates the possibility to propagate

the relative entropy (then chaos) for the very general systems (1.2) or (2.2) even with singular kernels. Indeed, in Theorem 3, the time T can be arbitrarily large a priori as long as the assumptions (2.10) on ρ are satisfied.

The assumption that V is anti-symmetric can be replaced by $\operatorname{div}K_1 \in L^\infty$. Our proof in Chapter 4 applies to this case in an identical way. We still keep the original form in Theorem 3 simply because it is more natural physically, considering the Helmholtz decomposition for instance.

2.3 The difference between the 2nd order case and the 1st order case.

For the 2nd order system (1.1) and its corresponding limit (1.3), our framework applies in an identical manner in the following three cases

- No randomness $\sigma_N = 0$ where (1.1) reduces to the deterministic Newton dynamics.
- Fixed randomness $\sigma_N \rightarrow \sigma > 0$ as $N \rightarrow +\infty$.
- Vanishing randomness $\sigma_N \rightarrow \sigma = 0$ as $N \rightarrow +\infty$.

In the general 1st order system (1.2) with $K \in L^\infty$, the presence of the noise $\sigma > 0$ is essential. However, for the 1st order system with the Hamiltonian structure, *i.e.* $\operatorname{div}K = 0$, the following corollary should be a direct consequence of Theorem 1 and a more advanced combinatorics result (for instance Prop. 9)

Corollary 1 *For the 1st order system (1.2) with $\operatorname{div}F \in L^\infty$, $K \in L^\infty$ and $\operatorname{div}K = 0$. One has the propagation of chaos and hence mean field limit result for (1.2)*

towards (1.4) in the vanishing viscosity cases $\sigma_N \rightarrow 0$ or in the purely deterministic case $\sigma_N = \sigma = 0$ under proper assumptions as in Theorem 1.

The generalization is straightforward: the mean field PDE (1.4) with $\sigma = 0$ and $F = 0$ can be written as

$$\partial_t \rho + K \star \rho \cdot \nabla_x \rho = 0, \quad (2.14)$$

since $\operatorname{div} K = 0$. Now the velocity field $K \star \rho$ is also divergence free as $(v, K \star \rho)$ in the 2nd order case. The same type cancellation rules as in Lemma 4 shall be expected for the 1st order system (2.14) with the Hamiltonian structure.

We write the cancellation rules as the following lemma

Lemma 6 *Assume that $4k \leq N$. Consider a function $\phi : E \times E \rightarrow \mathbb{R}$, with the following cancellation rules*

$$\int_E \phi(x, \cdot) \rho(x) dx = 0, \quad \int_E \phi(\cdot, x) \rho(x) dx = 0. \quad (2.15)$$

Let $\bar{\rho}_N = \rho^{\otimes N}$. Then for a double multi-index (I_{2k}, J_{2k}) with $I_{2k} = (i_1, \dots, i_{2k})$, $J_{2k} = (j_1, \dots, j_{2k})$ and $1 \leq i_\alpha, j_\beta \leq N$, the following integral vanishes, i.e.

$$\int_{E^N} \phi(x_{i_1}, x_{j_1}) \cdots \phi(x_{i_{2k}}, x_{j_{2k}}) \bar{\rho}_N dx_1 \cdots dx_N = 0$$

provided that there exists i_α such that

$$i_\alpha \notin \{i_1, \dots, i_{\alpha-1}, i_{\alpha+1}, \dots, i_{2k}, j_1, \dots, j_{2k}\},$$

or there exists j_β such that

$$j_\beta \notin \{i_1, \dots, i_{2k}, j_1, \dots, j_{\beta-1}, j_{\beta+1}, \dots, j_{2k}\}.$$

Lemma 6 can be easily proved by Fubini's Theorem and the two type cancellation rules (2.15). Notice that Lemma 6 applies both to the 2nd order case (Lemma 4) and the 1st order case (Lemma 19). In particular, for the 1st order system (1.2) with the Hamiltonian structure ($\operatorname{div}K = 0$), one can set $E = \mathbb{D}$ and

$$\phi(x_i, x_j) = \nabla_{x_i} \log \rho(x_i) \cdot \{K(x_i - x_j) - K \star \rho(x_i)\}.$$

As we have already seen in Chapter 1, the crucial part is to count the effective/non-vanishing double multi-indices (I_{2k}, J_{2k}) . This is answered by the following combinatorics result.

Proposition 9 *Assume that $4k \leq N$. Consider double multi-indices (I_{2k}, J_{2k}) with $I_{2k} = (i_1, \dots, i_{2k})$ and $J_{2k} = (j_1, \dots, j_{2k})$, where each component is chosen from $\{1, 2, \dots, N\}$. Define the singleton of I_{2k} as*

$$\operatorname{Sing}(I_{2k}) = \{i_\alpha \mid i_\alpha \neq i_\beta, \text{ for any } \beta \neq \alpha\}.$$

Similarly we can define the singleton of J_{2k} . We further define the general diagonal multi-indices as

$$\mathcal{D}_{N,2k} := \{(I_{2k}, J_{2k}) \mid \operatorname{Sing}(I_{2k}) \subset J_{2k}, \operatorname{Sing}(J_{2k}) \subset I_{2k}\},$$

where in $\operatorname{Sing}(I_{2k}) \subset J_{2k}$ we treat J_{2k} as a set of its components and the same for $\operatorname{Sing}(J_{2k}) \subset I_{2k}$. Then one has

$$|\mathcal{D}_{N,2k}| \leq C^k k^{2k} N^{2k},$$

where C is a universal constant.

Prop. 9 plays a fundamental role in obtaining Theorem 3 and also the above Corollary 1. A slightly stronger version of this proposition will be proved in Chapter 7, which is much more difficult than the counterpart for the 2nd order case, Lemma 4 for instance.

The total number of all the double multi-indices (I_{2k}, J_{2k}) is N^{4k} , while the total number of the general diagonal multi-indices (or non-vanishing multi-indices) is in the order of $C^k k^{2k} N^{2k}$, which is roughly $\sqrt{N^{4k}}$. This agrees with the order in classical Laws of Large Numbers Prop. 4. Furthermore, if we simply set $J_{2k} = \emptyset$, then $\mathcal{D}_{N,2k}$ will reduce to the effective set $\mathcal{E}_{N,2k}$ as in classical Laws of Large Numbers Prop. 4.

2.4 Comparison with the literature

The mean field limit and propagation of chaos is more well-investigated for the 1st order deterministic systems (the system (1.2) with $\sigma_N = 0$ for instance). Systems like (1.2) with kernels K non smooth only at the origin $x = 0$ enjoy additional symmetries with respect to the 2nd order case which make the derivation easier. We refer to [91] for a more thorough comparison.

The main example of the deterministic 1st order system is the point vortex method for the 2D Euler equations. The mean field limit has been obtained for well distributed initial conditions, see for example [41, 72] while the proof of propagation of chaos can be found in [129, 130]. We refer to [79] for the best results so far for general multi-dimensional 1st order systems.

In comparison with the deterministic case, the stochastic case, $\sigma_N > 0$ in (1.1) or (1.2), seems harder as many of the techniques developed in the deterministic settings are not applicable. The Lipschitz case, $K \in W_{loc}^{1,\infty}$ can still be handled through Gronwall like inequalities, see for instance [17, 29].

In the non degenerate case, $\sigma_N \rightarrow \sigma > 0$ in (1.2) for instance, the regularizing properties of the stochastic part can actually be exploited to handle some singularity in K (up to order $1/|x|$). For 1st order systems, propagation of chaos can hence be proved for the 2D viscous or stochastic vortex systems for the Euler equations, leading to the 2D incompressible Navier-Stokes system; see [58, 63, 124]. See also [64, 68, 108] for Keller-Segel systems with similar techniques.

Compared to the results in [63, 108, 124], Theorem 3 is weak in the singularity of kernels: The stream or potential function can at most have the “square root of logarithmic” singularity. For example we can deal with a variant stochastic vortex model for 2D Navier-Stokes by setting

$$K(x) = \gamma \nabla^\perp \left(\sqrt{\log |x|} \right),$$

which is less singular than the Biot-Savart kernel or the Poisson kernel in 2D.

But we only make assumptions on the order of the singularity, not on the specific structure or symmetries of the kernels. In particular, K can be *anisotropic* in our results. Also we can even treat measure-valued kernels. For instance we can choose that

$$K(x_1, x_2) = (\phi(x_2), \delta_0(x_1)), \quad K(x_1, x_2) = \nabla^\perp \mathbf{1}_{B_R(0)}(x_1, x_2)$$

where ϕ is a smooth function and δ_0 is the Dirac mass at the origin while $B_R(0)$ is

the open ball centered in 0 with the radius $R > 0$. Theorem 3 applies to both cases since $K \in W^{-1,\infty}$, $\operatorname{div}K = 0$ in both examples. The systems (1.2) with those K can model certain 2D collision models.

Furthermore, we have obtained explicit convergence rate for the relative entropy and for the L^1 distance between the marginals and the limit, which is usually absent from the literature. There are certain results [63, 108, 124] on the weak convergence up to any positive time t , but without any rate since they all rely on compactness arguments.

We also want to mention several recent results of the mean field limit for the 1st order systems under various assumptions of K . Firstly, in [87] a new coupling strategy and a Glivenko-Cantelli theorem are used to show the mean field limits for systems (1.1) or (1.2) with global Hölder continuous interaction kernels $K \in C^{0,\alpha}$. For 1st order system, $\alpha > 0$ is enough. But it requires $\alpha > \frac{2}{3}$ for 2nd systems in order to ensure the existence of a differentiable stochastic flow (see [134] for instance). The results are given essentially in the sense of large deviation. It is not directly comparable to our results. For instance for a kernel $K \in L^\infty$ satisfying our assumption might not be Hölder continuous (even discontinuous) at all.

Recently, inspired by the work in [131], a mean field limit result is obtained for the 1st order systems with an s -Riesz interaction gradient, *i.e.* $K = -\nabla W$ where W is defined in (1.5). We also refer to the recent preprints [10, 11] for the mean field limit for the 1st order system (1.2) in 1D with kernels like

$$K(x) = \frac{\phi(x)}{|x|^\alpha} \cdot \frac{x}{|x|}, \quad \alpha \in [0, 2)$$

where $\phi \in C_b^2(\mathbb{R}^d)$. The proof relies on the convexity property of the interaction potential and the corresponding functional. Gradient flow techniques [5] in metric spaces or Γ -convergence for certain functionals play a crucial role there.

The technical tools developed in this thesis could be applied to more complicated systems. We expect that the relative method will be a standard tool and a useful norm to study the mean field limit and related problems.

2.5 Related problems

There are many other interesting questions that are related to mean field limit for stochastic systems but that are out of the scope of this thesis. For instance

- The derivation of collisional models and Kac's Program in kinetic theory. After the seminal in [101] and later [37], the rigorous derivation of the Boltzmann equation was finally achieved in [65] but only for a short time (of the order of the average time between collisions). The derivation for longer time is still widely open in spite of some critical progress when close to equilibrium in [14]. Many tools and concepts that are used for mean field limits were initially introduced for collisional models, such as the ideas in the now famous Kac's program. Kac first introduced a probabilistic approach to simulate the spatially homogeneous Boltzmann equation in [97] and formulated several related conjecture. For most recent progress in Kac's program, we refer in particular to [83, 116, 117].

The Boltzmann type kinetic equations have been derived formally in certain

flocking models with topological instead of metric interactions [12, 78].

- Stochastic vortex dynamics with multiplicative (instead of additive) noise leading to Stochastic 2D Euler equation. In [58], the authors showed that the point vortex dynamics becomes fully well-posed for every initial configuration when a generic stochastic perturbation (in the form of multiplicative noises) compatible with the Euler description is introduced. The SDE systems in [58] will converge to the stochastic Euler equation, rather than Navier-Stokes equation as the number N of point vortices goes to infinity. However, the rigorous proof of the convergence is difficult and still open.
- Scaling limit (hydrodynamic limit) of random walks on discrete spaces, for instance on lattice \mathbb{Z}^d for which we refer to [98]. In this setting, one also tries to obtain a continuum model, usually a deterministic PDE, from a discrete particle model on a lattice, as $N \rightarrow \infty$ and of course the mesh size h converges to 0. An interesting observation is that we can use a stochastic PDE as a correction to the limit deterministic PDE, see [48].
- Quantum many particle systems and the derivation of non linear Schrödinger equations (nonlinear Hartree or Gross-Pitaevskii for instance) from N -particle linear Schrödinger equation. See for instance [40, 53, 54, 73] and the references therein. Now a very common strategy is to follow the BBGKY hierarchy and in particular show the uniqueness of the infinite hierarchy. It is natural and tempting to cook up a controllable norm (as the relative entropy in our case) to compare the quantum many particle systems and the corresponding limits.

We refer to [70,131] in particular for recent successes in this spirit.

Chapter 3: Proof of the main result: The 2nd order case

In this chapter, we give the proof of Theorem 1 by assuming the Main Estimate (5). The Main Estimate (5) will be proved in Chapter 6 by combinatorics argument. The main techniques is the entropy analysis which can also be applied to get a weak-strong uniqueness result Theorem (4) at the PDE level.

3.1 The Vlasov equation (1.3): Weak-strong uniqueness

Our framework, the relative entropy method at the level of the Liouville, is initially inspired by a classical weak-strong uniqueness argument for the Vlasov equation, based on the relative entropy of two solutions. Consider two non-negative solutions f and \tilde{f} with total mass 1 to Eq. (1.3). If f is smooth enough then it is possible to control the distance between them through the relative entropy of \tilde{f} with respect to f or

$$H(t) := H(\tilde{f}|f)(t) = \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} \log\left(\frac{\tilde{f}}{f}\right) dx dv.$$

More precisely, one has the following result

Theorem 4 (Weak-strong Uniqueness) *Assume that $K \in L^\infty$, that $f(t, x, v) \in L^\infty([0, T], L^1(\mathbb{D} \times \mathbb{R}^d) \cap W^{1,p})$ for any $1 \leq p \leq \infty$ is a strong solution to (1.3) with*

(1.10) for some $\lambda_f > 0$. Then for any $\tilde{f} \in L^\infty([0, T], L^1(\mathbb{D} \times \mathbb{R}^d))$, weak solution to (1.3) with mass 1, initial value \tilde{f}^0 and satisfying

$$\int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} \log \tilde{f} \, dx \, dv + \sigma \int_0^t \int_{\mathbb{D} \times \mathbb{R}^d} \frac{|\nabla_v \tilde{f}|^2}{\tilde{f}} \, dx \, dv \, ds \leq \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f}^0 \log \tilde{f}^0 \, dx \, dv,$$

one has for some constant $C > 0$ and any $t \in [0, T]$ that as long as $H(\tilde{f} | f)(s) \leq 1$ for any $s \in [0, t]$,

$$H(\tilde{f} | f)(t) \leq \exp(Ct \|K\|_{L^\infty} (1 + \log \theta_f)) H(\tilde{f} | f)(t=0).$$

In particular if initially $f(t=0) = \tilde{f}^0$ then $f = \tilde{f}$ at any later time.

The short proof of Theorem 4 is given in subsection B.1. It relies at the key step on a weighted Csiszár-Kullback-Pinsker inequality (see [21]).

Theorem 4 requires enough smoothness on f . Fortunately such solutions are guaranteed to exist, at least on some bounded time interval as per

Proposition 10 *Assume that $K \in L^\infty$, $f^0 \in L^1(\mathbb{D} \times \mathbb{R}^d) \cap W^{1,p}$ for every $1 \leq p \leq \infty$ and s.t. for some $\lambda_0 > 0$*

$$\int_{\mathbb{D} \times \mathbb{R}^d} e^{\lambda_0 |\nabla_{(x,v)} \log f^0|} f^0 \, dx \, dv < \infty.$$

Then there exists T depending on f^0 and $f \in L^\infty([0, T], L^1(\mathbb{D} \times \mathbb{R}^d) \cap W^{1,p})$ solution to (1.3) s.t. (1.10) holds for some $\lambda_f > 0$. Furthermore, if $\sigma = 0$ and we assume that

$$|\nabla_{(x,v)} \log f^0| \leq C(1 + |x|^k + |v|^k)$$

for some $k > 0$, then

$$\sup_{t \in [0, T]} |\nabla_{(x,v)} \log f(t, x, v)| \leq C e^{CT} (1 + |x|^k + |v|^k).$$

The proof of Prop. 10 is straightforward and also given in the appendix.

It is tempting to try to use directly a result like Theorem 4 to prove the Mean Field limit. In the case of the purely deterministic system (1.1) with $\sigma_N = 0$, one may associate to each solution the so-called empirical measure μ_N defined in (1.7). If $(X_i, V_i)_{i=1\dots N}$ solves (1.1) in an appropriate sense (for instance it comes from a flow), then μ_N defined through (1.7) is a solution to Eq. (1.3) in the sense of distribution. If one could then use a weak-strong uniqueness principle to compare μ_N to the expected smooth limit f then the Mean Field limit and propagation of chaos would follow.

This general idea plays an important role in the recent [104] for instance (see also [16, 103]), leading to an improved truncation parameter (see the discussion after the main result in Chapter 1). However Theorem 4 relies on a very different weak-strong uniqueness principle than the one used in [104] and cannot be used directly as it is. There are several reasons for that: In particular Theorem 4 requires the weak solution \tilde{f} to have a bounded entropy, which cannot be the case of the empirical measure μ_N .

Instead the main result in this article consists in extending Theorem 4 to the Liouville Eq. (1.6).

The study of well-posedness for Vlasov-type systems is now classical and mostly focused on the Vlasov-Poisson case ($K = C x/|x|^d$). The existence of weak solutions was obtained in [7] but global existence of strong solutions in dimen-

sion 3 had long been difficult (see [8] for small initial data) before being obtained in [126, 128] and concurrently in [107] through the propagation of moments (see also [125] for more recent estimates). The most general uniqueness result for the Vlasov-Poisson system was obtained in [109].

3.2 Main Estimate: The need of combinatorics

Instead of trying to use directly Theorem 4, our approach is to try to mimic its relative entropy estimate but at the level of the Liouville equation (1.6).

First define the tensor product of the expected limit f by

$$\bar{f}_N(t, X, V) = \prod_{i=1}^N f(t, x_i, v_i),$$

We can now directly compare f_N to \bar{f}_N through the N dimensional relative entropy

$$H_N(f_N|\bar{f}_N)(t) = \frac{1}{N} \int_{\mathbb{D}^N \times (\mathbb{R}^d)^N} f_N \log\left(\frac{f_N}{\bar{f}_N}\right) dZ.$$

We will also write $H_N(t) := H_N(f_N|\bar{f}_N)(t)$ in short. The key difficulty is to find a suitable replacement for the weighted Csiszár-Kullback-Pinsker inequality used in the proof of Theorem 4. This turns out to be very delicate and it is the main technical contribution of the article.

Define for any $p \geq 1$

$$M_p := \left(\int_{\mathbb{D} \times \mathbb{R}^d} |\nabla_v \log f|^p f \, dx \, dv \right)^{\frac{1}{p}},$$

then one has

Theorem 5 *Assume that $f \in L^\infty \cap L^1(\mathbb{D} \times \mathbb{R}^d)$ with $f \geq 0$ and $\int f = 1$, that $\nabla_v f \in W_{loc}^{1,p}$ for every $1 \leq p \leq \infty$ with $\sup_{1 \leq p < \infty} \frac{M_p}{p} < \infty$ and that $\|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) < \frac{1}{8e^2}$,*

then

$$\int_{(\mathbb{D} \times \mathbb{R}^d)^N} \bar{f}_N \exp(|R_N|) dZ \leq 5 + 6 \left(\frac{8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right)}{1 - \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^2} \right)^2 < \infty,$$

where $\bar{f}_N = \prod_{i=1}^N f(t, x_i, v_i)$ and R_N is defined by

$$R_N = \frac{1}{N} \sum_{i,j=1}^N \nabla_{v_i} \log f(x_i, v_i) \cdot \{K(x_i - x_j) - K \star \rho(x_i)\}. \quad (3.1)$$

It is straightforward to see why R_N as defined in Eq. (3.1) is the key quantity.

Defining the Liouville operator as

$$L_N = \sum_{i=1}^N v_i \cdot \nabla_{x_i} + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} K(x_i - x_j) \cdot \nabla_{v_i} - \sigma_N \sum_{i=1}^N \Delta_{v_i},$$

the Liouville Eq. (1.6) can be written as

$$\partial_t f_N + L_N f_N = 0.$$

Indeed since f solves the limit Eq. (1.3) then \bar{f}_N solves the Liouville Eq. with a right-hand side given by R_N

$$\partial_t \bar{f}_N + L_N \bar{f}_N = R_N \bar{f}_N + (\sigma - \sigma_N) \sum_{i=1}^N \Delta_{v_i} \bar{f}_N.$$

Theorem 5 is a sort of modified law of large numbers, written at an exponential or large deviation scale. Contrary to usual laws of large numbers, that have been used for Mean Field limits recently in [70] with $K \in W_{loc}^{1,\infty}$, R_N here exhibits a double sum so that *a priori* $R_N = O(N)$ and the challenge in Theorem 5 is to prove that in fact $R_N = O(1)$.

Finally we observe that the assumption $\sup_p \frac{M_p}{p} < \infty$ is essentially equivalent to the assumption (1.10) in Theorem 1. Indeed,

i) $\sup_p \frac{M_p}{p} < \Lambda$ implies $\int f e^{\lambda |\nabla_v \log f|} dz$ is finite for any $\lambda < \frac{1}{e\Lambda}$: By Taylor expansion for e^x ,

$$\begin{aligned} \int f e^{\lambda |\nabla_v \log f|} dz &\leq 1 + \sum_{p=1}^{\infty} \frac{1}{p!} \lambda^p \int f |\nabla_v \log f|^p dz \\ &\leq 1 + \sum_{p=1}^{\infty} \frac{1}{p!} \lambda^p (\Lambda p)^p \leq 1 + \sum_{p=1}^{\infty} (e\Lambda\lambda)^p. \end{aligned}$$

ii) Assumption (1.10) implies $\sup_p \frac{M_p}{p} \leq \frac{\theta_f}{\lambda_f}$. Indeed, for any $p = 1, 2, \dots$,

$$\int f |\nabla_v \log f|^p dz \leq p! \lambda_f^{-p} \int f e^{\lambda_f |\nabla_v \log f|} dz.$$

Since $p! \leq p^p$,

$$\sup_p \frac{M_p}{p} \leq \frac{1}{\lambda_f} \sup_p \left(\int f e^{\lambda_f |\nabla_v \log f|} dz \right)^{\frac{1}{p}} < \infty.$$

3.3 From combinatorics and Theorem 5, to Theorem 1

Recall that f is a strong solution to the Vlasov Eq. (1.3). Therefore \bar{f}_N solves

$$\partial_t \bar{f}_N + L_N \bar{f}_N = \bar{f}_N R_N + (\sigma - \sigma_N) \sum_{i=1}^N \Delta_{v_i} \bar{f}_N, \quad (3.2)$$

where R_N is defined as (3.1) with the convention that $K(0) = 0$.

From this point the initial calculations exactly follow the proof of Theorem 4.

Since f_N is a weak solution to the Liouville Eq. according to Prop. 7

$$\begin{aligned} H_N(t) &= \frac{1}{N} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N \log \left(\frac{f_N}{\bar{f}_N} \right) dZ = \frac{1}{N} \int f_N \log f_N - \frac{1}{N} \int f_N \log \bar{f}_N \\ &\leq \frac{1}{N} \int f_N^0 \log f_N^0 - \frac{\sigma_N}{N} \int_0^t \int \frac{|\nabla_V f_N|^2}{f_N} - \frac{1}{N} \int f_N \log \bar{f}_N, \end{aligned}$$

per the assumption of dissipation of entropy for f_N in Prop. 8.

Since \bar{f}_N is smooth, $\log \bar{f}_N$ can be used as a test function against f_N which is a weak solution to the Liouville Eq. (1.6) so that

$$\int f_N \log \bar{f}_N = \int f_N^0 \log \bar{f}_N^0 + \int_0^t \int f_N(s, X, V) (\partial_t \log \bar{f}_N + L_N^* \log \bar{f}_N) dZ ds,$$

where

$$L_N^* = \sum_{i=1}^N v_i \cdot \nabla_{x_i} + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i}^N K(x_i - x_j) \cdot \nabla_{v_i} + \sigma_N \sum_{i=1}^N \Delta_{v_i}.$$

Since \bar{f}_N is a strong solution to (3.2), this leads to

$$\begin{aligned} \int f_N \log \bar{f}_N &= \int f_N^0 \log \bar{f}_N^0 + \int_0^t \int f_N R_N dZ ds \\ &+ \sigma_N \int_0^t \int f_N \left(\frac{\Delta_V \bar{f}_N}{\bar{f}_N} + \Delta_V \log \bar{f}_N \right) dZ ds + (\sigma - \sigma_N) \int_0^t \int f_N \frac{\Delta_V \bar{f}_N}{\bar{f}_N} dZ ds. \end{aligned}$$

Hence,

$$\begin{aligned} H_N(t) &\leq H_N(0) - \frac{1}{N} \int_0^t \int f_N R_N dZ ds \\ &- \frac{\sigma_N}{N} \int_0^t \int \left[\frac{|\nabla_V f_N|^2}{f_N} + f_N \left(\frac{\Delta_V \bar{f}_N}{\bar{f}_N} + \Delta_V \log \bar{f}_N \right) \right] dZ ds \\ &- \frac{\sigma - \sigma_N}{N} \int_0^t \int f_N \frac{\Delta_V \bar{f}_N}{\bar{f}_N} dZ ds. \end{aligned} \tag{3.3}$$

Entropy analysis gives us the following estimate for (3.3).

Lemma 7 *One has the estimate for the diffusion terms in (3.3)*

$$\begin{aligned} &- \frac{\sigma_N}{N} \int \left[\frac{|\nabla_V f_N|^2}{f_N} + f_N \left(\frac{\Delta_V \bar{f}_N}{\bar{f}_N} + \Delta_V \log \bar{f}_N \right) \right] dZ - \frac{\sigma - \sigma_N}{N} \int f_N \frac{\Delta_V \bar{f}_N}{\bar{f}_N} dZ \\ &\leq \alpha_N \rightarrow 0, \end{aligned}$$

as $N \rightarrow \infty$. In particular, we can choose $\alpha_N = 0$ in the case $\sigma_N \equiv \sigma \geq 0$, $\alpha_N = C \frac{(\sigma - \sigma_N)^2}{4\sigma\sigma_N}$ and $\alpha_N = C\sigma_N$ with a universal constant C in the case $\sigma_N \rightarrow \sigma > 0$ and $\sigma_N \rightarrow 0$ respectively.

Proof We now treat the three types of the choices of σ_N separately.

Case I: $\sigma_N = \sigma \geq 0$. In this case, the last term in the right-hand side of (3.3)

vanishes. Classical entropy estimates show that

$$\int \frac{|\nabla_V f_N|^2}{f_N} + \int f_N \left(\frac{\Delta_V \bar{f}_N}{\bar{f}_N} + \Delta_V \log \bar{f}_N \right) dZ = \int f_N |\nabla_V \log \frac{f_N}{\bar{f}_N}|^2 dZ \geq 0,$$

see the proof of Theorem 4 for detailed calculations.

Therefore we finally obtain that

$$H_N(t) \leq H_N(0) - \frac{1}{N} \int_0^t \int f_N R_N \, dZ \, ds. \quad (3.4)$$

Case II: $\sigma_N \rightarrow \sigma > 0$. The terms in (3.3) induced by randomness can be bounded

by the entropy of f_N^0 ,

$$\begin{aligned} & -\frac{1}{N} \int_0^t \int \left[\sigma_N \frac{|\nabla_V f_N|^2}{f_N} + \sigma f_N \frac{\Delta_V \bar{f}_N}{\bar{f}_N} + \sigma_N f_N \Delta_V \log \bar{f}_N \right] \, dZ \, ds \\ &= -\frac{1}{N} \int_0^t \int f_N \sigma |\nabla_V \log \bar{f}_N - \frac{\sigma + \sigma_N}{2\sigma} \nabla_V \log f_N|^2 \, dZ \, ds \\ & \quad + \frac{(\sigma - \sigma_N)^2}{4\sigma} \int_0^t \int \frac{|\nabla_V f_N|^2}{f_N} \, dZ \, ds \\ & \leq \frac{(\sigma - \sigma_N)^2}{4\sigma} \frac{1}{N} \int_0^t \int \frac{|\nabla_V f_N|^2}{f_N} \, dZ \, ds \\ & \leq \frac{(\sigma - \sigma_N)^2}{4\sigma\sigma_N} \left[\frac{1}{N} \int f_N^0 \log f_N^0 - \frac{1}{N} \int f_N(t) \log f_N(t) \right]. \end{aligned}$$

Recalling the assumption (1.12) and Prop. 8, one has for any $t \in [0, T]$

$$\sup_{N \geq 2} \frac{1}{N} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} f_N \log f_N \, dZ \geq C_d - \sup_{t \in [0, T]} \frac{1}{N} \int \sum_{i=1}^N (1 + |z_i|^2) f_N(t, Z) \, dZ \geq -C,$$

where C_d is a universal constant only depending on the dimension d and $C > 0$ is a universal constant only depending on the uniform bound in (1.12), the time interval T and the dimension d . Therefore, we obtain that

$$H_N(t) \leq H_N(0) - \frac{1}{N} \int_0^t \int f_N R_N \, dZ \, ds + \alpha_N, \quad (3.5)$$

where

$$\alpha_N := \frac{(\sigma - \sigma_N)^2}{4\sigma\sigma_N} \left[\frac{1}{N} \int f_N^0 \log f_N^0 - \frac{1}{N} \int f_N(t) \log f_N(t) \right] \leq C \frac{(\sigma - \sigma_N)^2}{4\sigma\sigma_N}$$

goes to 0 as $N \rightarrow \infty$ and again C only depends on the uniform bounds in (1.12)

and the time T and the dimension d .

Case III: $\sigma_N \rightarrow \sigma = 0$. This is the vanishing randomness case, that is there is no diffusion in the limit Vlasov equation. The terms in (3.3) induced by randomness in N -particle system can also be bounded but by some moment bounds for f_N^0 ,

$$\begin{aligned} S(\sigma_N) &:= -\frac{\sigma_N}{N} \int_0^t \int \left[\frac{|\nabla_V f_N|^2}{f_N} + f_N \Delta_V \log \bar{f}_N \right] dZ ds \\ &= -\frac{\sigma_N}{N} \int_0^t \int \frac{|\nabla_V f_N|^2}{f_N} + \frac{\sigma_N}{N} \int_0^t \int \nabla_V f_N \cdot \nabla_V \log \bar{f}_N \\ &\leq \frac{\sigma_N}{4N} \int_0^t \int f_N |\nabla_V \log \bar{f}_N|^2 dZ ds. \end{aligned}$$

This is the reason why we add here extra moment restrictions. Recall that (1.11) and the second part of Proposition 10, *i.e.*

$$|\nabla_v \log f| \leq |\nabla_{(x,v)} \log f| \leq C(1 + |x|^k + |v|^k).$$

Therefore,

$$S(\sigma_N) \leq C \frac{\sigma_N}{4} \left(\frac{1}{N} \int_0^t \int \sum_{i=1}^N (1 + |x_i|^{2k} + |v_i|^{2k}) f_N dZ ds \right) \rightarrow 0,$$

as $N \rightarrow \infty$. Hence, we also obtain (3.5) in this case with $\alpha_N \leq C\sigma_N \rightarrow 0$ as $N \rightarrow \infty$. This completes the proof. \square

Now we can proceed to prove the estimate for $H_N(t)$. Recall the Frenchel's inequality for the function $u(x) = x \log x$: For all $x, y \geq 0$

$$xy \leq x \log x + \exp(y - 1).$$

Hence for $\nu > 0$

$$-f_N R_N \leq \frac{\bar{f}_N}{\nu} \left(\frac{f_N}{\bar{f}_N} \nu |R_N| \right) \leq \frac{\bar{f}_N}{\nu} \left(\frac{f_N}{\bar{f}_N} \log \left(\frac{f_N}{\bar{f}_N} \right) + \exp(\nu |R_N|) \right).$$

Therefore Eq. (3.5) becomes

$$H_N(t) \leq H_N(0) + \alpha_N + \frac{1}{\nu} \int_0^t H_N(s) ds + \frac{1}{\nu} \frac{1}{N} \int_0^t \int \bar{f}_N \exp(\nu |R_N|) dZ ds. \quad (3.6)$$

Now define $\tilde{K} = \nu K$ and take ν s.t.

$$\|\tilde{K}\|_{L^\infty} \sup_p \frac{M_p}{p} = \nu \|K\|_{L^\infty} \sup_p \frac{M_p}{p} \leq \nu \|K\|_{L^\infty} \theta_f / \lambda_f \leq \frac{1}{16 e^2}.$$

We may apply Theorem 5 to \tilde{K} and $\tilde{R}_N = \nu R_N$. This implies that

$$L = \sup_N \sup_{t \in [0, T]} \int \bar{f}_N \exp(\nu |R_N|) dZ \leq 10.$$

Inserting this in (3.6) gives

$$H_N(t) \leq H_N(0) + \alpha_N + \frac{1}{\nu} \int_0^t H_N(s) ds + \frac{10t}{\nu N},$$

and up to time $T > 0$, by Gronwall's inequality

$$H_N(f_N | \bar{f}_N)(t) \leq \left(H_N(f_N | \bar{f}_N)(0) + \alpha_N + \frac{10}{N} \right) \exp(t/\nu), \quad (3.7)$$

which gives the first part of Theorem 1 taking $\nu^{-1} = 16 e^2 \|K\|_{L^\infty} \theta_f / \lambda_f$.

Now we apply Prop. 1, for any fixed $k \geq 1$,

$$H_k(f_{N,k} | f^{\otimes k}) = \frac{1}{k} \int_{(\mathbb{D} \times \mathbb{R}^d)^k} f_{N,k} \log \left(\frac{f_{N,k}}{f^{\otimes k}} \right) dz_1 \cdots dz_k \leq H_N(f_N | \bar{f}_N) \rightarrow 0,$$

as $N \rightarrow \infty$.

The classical Csiszár-Kullback-Pinsker inequality (see chapter 22 in [138]) then implies that

$$\|f_{N,k} - f^{\otimes k}\|_{L^1} \leq \sqrt{2k H_k(f_{N,k} | f^{\otimes k})} \rightarrow 0$$

as $N \rightarrow \infty$. This completes the proof of Theorem 1.

Chapter 4: Proof of the main results: The 1st order case

In this chapter, we prove the main result Theorem 3 for the 1st order system by assuming two main estimates Theorem 6 and Theorem 7 whose proof are the main technical difficulties in this thesis and will be given in Chapter 7. The difference to the 2nd order case is that presence of the noise $\sigma > 0$ is essential since we need the diffusion of the relative entropy (with minus sign) to cancel some bad terms splitting from integration by parts. Moreover, we apply the entropy estimate in Prop. 8 to control the average minimal distance (Prop. 12) of particles, which in turn controls the contribution of the singular part.

4.1 Main Estimates: Combinatorics results

The idea to prove the main result Theorem 3, is rather straightforward: we study the evolution of the relative entropy and try to control the growth of it.

In this chapter, for simplicity we write that

$$\bar{\rho}_N(t, X) = \prod_{i=1}^N \rho_t(x_i) = \rho_t^{\otimes N}, \quad H_N(t) := H_N(\rho_N(t) | \rho_t^{\otimes N}).$$

In Theorem 3, we assume that the kernel K permits a decomposition $K = K_1 + K_2$, where $K_1 = \operatorname{div} V$, V is a matrix valued function. The components of K_1 can be

written as

$$K_1^h = \sum_{l=1}^d \partial_{x^l} V_{hl}, \quad h = 1, \dots, d. \quad (4.1)$$

We further define

$$\delta V_{hl}^{ij} = V_{hl}(x_i - x_j) - V_{hl} \star \rho(x_i), \quad (4.2)$$

and adopt the convention that $V_{hl}(0) = 0$ for each $1 \leq h, l \leq d$. Finally we define

$$\Delta^{ij} = (\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i). \quad (4.3)$$

The main estimates in the 1st order case can be formulated as the following

Theorem 6 (Main Estimate I) *Suppose that $\rho \in L^\infty \cap L^1(\mathbb{D})$ with $\rho \geq 0$ and $\int_{\mathbb{D}} \rho(x) dx = 1$. Assume that $g, \phi \in L^\infty(\mathbb{D})$ with $\|\phi\|_{L^\infty} \|g\|_{L^\infty} < \frac{1}{2e}$. Then*

$$\int_{\mathbb{D}^N} \bar{\rho}_N \exp(R_{N,i}^{\phi,g}) dX \leq 3 \left(1 + \frac{5\alpha}{(1-\alpha)^3} + \frac{\beta}{1-\beta} \right) < \infty, \quad (4.4)$$

where $\bar{\rho}_N = \prod_{i=1}^N \rho(t, x_i)$ and $R_{N,i}^{\phi,g}$ is defined by

$$R_{N,i}^{\phi,g} = \frac{1}{N} \sum_{j_1, j_2=1}^N g^2(x_i) \delta\phi^{ij_1} \delta\phi^{ij_2},$$

with $\delta\phi^{ij} = \phi(x_i - x_j) - \phi \star \rho(x_i)$ and

$$\alpha = (2e \|V\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty})^4 < 1, \quad \beta = \left(2\sqrt{2e} \|V\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty} \right)^4 < 1.$$

In Theorem 6, the bounded functions ϕ, g can represent V_{hl} and $|\partial_{x^h} \log \rho|$ respectively for instance. Therefore one has the following corollary

Corollary 2 *Suppose that $\rho \in L^\infty \cap L^1(\mathbb{D})$ with $\rho \geq 0$ and $\int_{\mathbb{D}} \rho(x) dx = 1$. Assume that $V, \nabla_x \log \rho \in L^\infty(\mathbb{D})$ with $\|V\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty} < \frac{1}{2e}$. Then*

$$\sup_{1 \leq i \leq N} \sup_{1 \leq h, l \leq d} \int_{\mathbb{D}^N} \bar{\rho}_N \exp(\Upsilon_{hl}^{N,i}) dX \leq 3 \left(1 + \frac{5\alpha}{(1-\alpha)^3} + \frac{\beta}{1-\beta} \right) < \infty, \quad (4.5)$$

where $\bar{\rho}_N = \Pi_{i=1}^N \rho(t, x_i)$ and $\Upsilon_{hl}^{N,i}$ is defined by

$$\Upsilon_{hl}^{N,i} = \frac{1}{N} \sum_{j_1, j_2=1}^N (\partial_{x_i^h} \log \rho(x_i)) \delta V_{hl}^{ij_1} \delta V_{hl}^{ij_2}, \quad (4.6)$$

and α, β are defined similarly as in Theorem 6.

Theorem 7 (Main Estimate II) *Suppose that $\rho \in L^\infty \cap L^1(\mathbb{D})$ with $\rho \geq 0$ and $\int_{\mathbb{D}} \rho dx = 1$, the vector field $K = \operatorname{div} V$, $V = (V_{hl})_{1 \leq h, l \leq d}$ is a matrix valued function and that as defined in (2.11) in Theorem 3,*

$$\sup_{p \geq 1} \frac{\|R\|_{L^p(\rho dx)}}{p} < \infty$$

and

$$\gamma := \left(C [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{\|R\|_{L^p(\rho dx)}}{p} + 1 \right) \right)^2 < 1,$$

where C is a universal constant. Then

$$\int_{\mathbb{D}^N} \bar{\rho}_N \exp(\Theta_N) dX \leq \frac{3}{1 - \gamma} < \infty, \quad (4.7)$$

where $\bar{\rho}_N(t, X) = \Pi_{i=1}^N \rho(t, x_i)$ and Θ_N is defined by

$$\Theta_N \equiv \frac{1}{N} \sum_{i,j=1}^N \left[\left(\sum_{h,l=1}^d R_{hl}(x_i) \delta V_{hl}^{ij} \right) - \Delta^{ij} \right],$$

where R, R_{hl} are defined in (2.11) while δV_{hl} and Δ^{ij} are defined in (4.2) and (4.3) respectively.

The proof of the previous main estimates is the main technical difficulty of the article and will be given in Chapter 7.

4.2 The evolution of the relative entropy

The starting point of the proof is the evolution of the relative entropy as per

Lemma 8 (Evolution of Relative Entropy) *For $\rho_N(t)$ solving the Liouville Eq. (2.2) and the ρ_t a strong solution to (1.4) with initial data $\rho_N(0)$ and ρ_0 respectively, the relative entropy can be estimated as*

$$\begin{aligned} H_N(t) &\leq H_N(0) - \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N G_N \, dX \, ds - \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N Q_N \, dX \, ds - \frac{1}{N} \int_0^t \mathcal{D}_N \, ds \\ &= H_N(0) + I + II + III \end{aligned} \quad (4.8)$$

where

$$G_N \equiv \frac{1}{N} \sum_{i,j=1}^N \nabla_{x_i} \log \rho(x_i) \{K(x_i - x_j) - K \star \rho(x_i)\}, \quad (4.9)$$

$$Q_N \equiv \frac{1}{N} \sum_{i,j=1}^N \{(\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i)\} \quad (4.10)$$

with the convention that $K(0) = 0$ and $(\operatorname{div}_x K)(0) = 0$ and the diffusion term (depends on σ , σ_N obviously) is defined as

$$\mathcal{D}_N \equiv \sigma_N \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N|^2}{\rho_N} \, dX + \sigma_N \int_{\mathbb{D}^N} \rho_N \Delta_X \log \bar{\rho}_N \, dX + \sigma \int_{\mathbb{D}^N} \rho_N \frac{\Delta_X \bar{\rho}_N}{\bar{\rho}_N} \, dX.$$

Proof Since ρ_N is a weak solution to the Liouville equation (2.2), the relative entropy $H_N(t)$ thus can be estimated as follows

$$\begin{aligned} H_N(t) &\leq \frac{1}{N} \int_{\mathbb{D}^N} \rho_N \log \frac{\rho_N}{\bar{\rho}_N} \, dX = \frac{1}{N} \int_{\mathbb{D}^N} \rho_N \log \rho_N \, dX - \frac{1}{N} \int_{\mathbb{D}^N} \rho_N \log \bar{\rho}_N \, dX \\ &\leq \frac{1}{N} \int_{\mathbb{D}^N} \rho_N^0 \log \rho_N^0 \, dX - \frac{\sigma_N}{N} \int_0^t \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N(s, X)|^2}{\rho_N(s, X)} \, dX \, ds \\ &\quad - \frac{1}{N} \sum_{i=1}^N \int_0^t \int_{\mathbb{D}^N} \rho_N(s, X) [(\operatorname{div}_x F)(x_i) + \frac{1}{N} \sum_{j \neq i} (\operatorname{div}_x K)(x_i - x_j)] \, dX \, ds \\ &\quad - \frac{1}{N} \int_{\mathbb{D}^N} \rho_N \log \bar{\rho}_N \, dX \end{aligned} \quad (4.11)$$

according to the assumption of dissipation of entropy for ρ_N as in Prop. 8.

Recall that ρ is a strong solution to the macroscopic PDE (1.4), then $\log \bar{\rho}_N(X) = \sum_{i=1}^N \log \rho(x_i)$ can be used as a test function against ρ_N which is a weak solution to the Liouville equation (2.2) such a way that

$$\begin{aligned} \frac{1}{N} \int_{\mathbb{D}^N} \rho_N \log \bar{\rho}_N \, dX &= \frac{1}{N} \int_{\mathbb{D}^N} \rho_N^0 \log \bar{\rho}_N^0 \, dX \\ &+ \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N(s, X) \{ \partial_t \log \bar{\rho}_N + L_N^* \log \bar{\rho}_N \} \, dX \, ds, \end{aligned} \quad (4.12)$$

where the dual of the differential operator L_N is given by

$$L_N^* = \sum_{i=1}^N F(x_i) \cdot \nabla_{x_i} + \frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} K(x_i - x_j) \cdot \nabla_{x_i} + \sigma_N \sum_{i=1}^N \Delta_{x_i}.$$

A simple computation shows that

$$\begin{aligned} \partial_t \log \bar{\rho}_N + L_N^* \log \bar{\rho}_N &= \sum_{i=1}^N \nabla_{x_i} \log \rho(x_i) \cdot \left\{ \frac{1}{N} \sum_{j \neq i} K(x_i - x_j) - K \star \rho(x_i) \right\} \\ &- \sum_{i=1}^N (\operatorname{div}_x F)(x_i) - \sum_{i=1}^N (\operatorname{div}_x K) \star \rho(x_i) + \sigma_N \Delta_X \log \bar{\rho}_N + \sigma \frac{\Delta_X \bar{\rho}_N}{\bar{\rho}_N}. \end{aligned} \quad (4.13)$$

Combing (4.11), (4.12) and (4.13), we prove this lemma. \square

In the following, we treat three terms I , II and III in (4.8) one by one. A priori trivial bounds for the first two term read

$$|I| \leq \|\nabla_x \log \rho\|_{L^\infty} \|K\|_{L^\infty}, \quad |II| \leq \|\operatorname{div} K\|_{L^\infty},$$

which are both in the order 1 and will make it impossible to obtain the expected smallness of $H_N(t)$, *i.e.* $H_N(t) \rightarrow 0$ when $N \rightarrow \infty$. More precise combinatorics results, considering the subtle cancellation rules in the integrals I and II , will be *critical* to get this proof done.

The last term, due to the randomness in the particle system (1.2) and the corresponding diffusion in the limit (1.4), will help to cancel some bad terms splitting

from I for instance by integration by parts. That is the *reason* we need the viscosity σ to be strictly positive even though it can be arbitrarily small.

We deal with the term III first. The essential property of K is that K permits the decomposition $K = K_1 + K_2$ with $\operatorname{div}K_1 \in L^\infty, K_2 \in L^\infty$. We write the estimate for III as the following lemma

Lemma 9 *Assume that $\operatorname{div}F \in L^\infty$ and that the kernel K permits a decomposition $K = K_1 + K_2$ with $\operatorname{div}K_1 \in L^\infty$ and $K_2 \in L^\infty$. Then the term III in (4.8) can be estimated as*

$$III = -\frac{1}{N} \int_0^t \mathcal{D}_N \, ds \leq -\frac{\sigma}{2N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 \, dX \, ds + \Lambda_0 (\sigma - \sigma_N)^2,$$

where the constant Λ_0 has the explicit form

$$\Lambda_0 \equiv \frac{2}{\sigma^2} \left(\sup_{N \geq 2} \frac{1}{N} \int \rho_N^0 \log \rho_N^0 \, dX + T \|\operatorname{div}_x K_1\|_{L^\infty} + T \|\operatorname{div}_x F\|_{L^\infty} + \frac{T}{\sigma} \|K_2\|_{L^\infty}^2 \right). \quad (4.14)$$

Proof of Lemma 9 We discuss two types of the choices of σ_N separately.

Case I: $\sigma_N \equiv \sigma$ for any $N \geq 2$, *i.e.* the strength of the noise does not depend on the number of interacting particles. Then \mathcal{D}_N in III coincides with the diffusion of the relative entropy

$$\sigma \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 \, dX.$$

In this case

$$III = -\frac{\sigma}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 \, dX \, ds,$$

which gives the thesis.

Case II: $\sigma_N \rightarrow \sigma > 0$. Without loss of generality, assume that $\sigma_N \geq \sigma/2$ for any $N \geq 2$ since we are interested in the asymptotic behavior as $N \rightarrow \infty$. Then now \mathcal{D}_N in (4.8) can be rewritten as

$$\begin{aligned} \mathcal{D}_N &= \frac{\sigma}{2} \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX + \frac{\sigma}{2} \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \bar{\rho}_N - \frac{\sigma_N}{\sigma} \nabla_X \log \rho_N \right|^2 dX \\ &\quad - \frac{(\sigma - \sigma_N)^2}{2\sigma} \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N|^2}{\rho_N} dX \end{aligned}$$

which is thus trivially bounded from below by

$$\frac{\sigma}{2} \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX - \frac{(\sigma - \sigma_N)^2}{2\sigma} \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N|^2}{\rho_N} dX.$$

Inserting this back to the term *III*, we get

$$\begin{aligned} III &\leq -\frac{\sigma}{2N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX ds \\ &\quad + \frac{(\sigma - \sigma_N)^2}{2\sigma\sigma_N} \left[\sigma_N \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N|^2}{\rho_N} dX ds \right]. \end{aligned} \tag{4.15}$$

Thanks to the estimate ii) in Proposition 8, we can then bound the term inside the bracket $[\cdot]$. Indeed, by Cauchy-Schwarz inequality, one has

$$\begin{aligned} &\frac{1}{N} \sum_{i=1} \sum_{j \neq i} \int \nabla_{x_i} \rho_N K_2(x_i - x_j) dX \\ &\leq \frac{\sigma}{4} \int \frac{|\nabla_X \rho_N|^2}{\rho_N} dX + \frac{1}{\sigma} N \int \rho_N |K_2|^2(x_1 - x_2) dX. \end{aligned}$$

Consequently, combining with the fact that $\sigma_N \geq \sigma/2$, the inequality ii) in Proposition 8 becomes

$$\begin{aligned} &\int_{\mathbb{D}^N} \rho_N(t, X) \log \rho_N(t, X) dX + \frac{\sigma_N}{2} \int_0^t \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N(s, X)|^2}{\rho_N(s, X)} dX ds \\ &\leq \int_{\mathbb{D}^N} \rho_N^0 \log \rho_N^0 dX + NT \left(\frac{1}{\sigma} \|K_2\|_{L^\infty}^2 + \|\operatorname{div}_x K_1\|_{L^\infty} + \|\operatorname{div}_x F\|_{L^\infty} \right). \end{aligned} \tag{4.16}$$

Since the entropy (if well-defined) of a probability measure on torus $\mathbb{D} = \mathbb{T}^d$ is always non-negative, we can estimate the quantity inside the bracket $[\cdot]$ with

$$2 \left(\frac{1}{N} \int \rho_N^0 \log \rho_N^0 dX + \frac{T}{\sigma} \|K_2\|_{L^\infty}^2 + T \|\operatorname{div}_x K_1\|_{L^\infty} + T \|\operatorname{div}_x F\|_{L^\infty} \right).$$

Combining with (4.15), one reaches the thesis and the constant Λ_0 is given by (4.14).

□

Up to now, we have not considered the specific structure of the kernel K . Recall that K permits a decomposition $K = K_1 + K_2$ with $K_1 = \operatorname{div}V$ where $K_2 \in L^\infty$, V is an anti-symmetric matrix valued function with a square root of logarithmic singularity at the origin as in (2.9). We use the usual *divide and conquer* strategy as per the following lemma

Lemma 10 *For $\rho_N(t)$ solving the Liouville Eq. (2.2) and the ρ_t a strong solution to (1.4) with initial data $\rho_N(0)$ and ρ_0 respectively, the relative entropy can be estimated as*

$$H_N(t) \leq H_N(0) + \Lambda_0(\sigma - \sigma_N)^2 + \mathcal{J}_1 + \mathcal{J}_2, \quad (4.17)$$

where the constant Λ_0 is given in (4.14) and for $\nu = 1, 2$

$$\begin{aligned} \mathcal{J}_\nu &= -\frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N G_N^\nu \, dX \, ds - \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N Q_N^\nu \, dX \, ds \\ &\quad - \frac{\sigma}{4N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\rho_N} \right|^2 \, dX \, ds, \end{aligned} \quad (4.18)$$

with

$$\begin{aligned} G_N^\nu &\equiv \frac{1}{N} \sum_{i,j=1}^N \nabla_{x_i} \log \rho(x_i) \{K_\nu(x_i - x_j) - K_\nu \star \rho(x_i)\}, \\ Q_N^\nu &\equiv \frac{1}{N} \sum_{i,j=1}^N \{(\operatorname{div}_x K_\nu)(x_i - x_j) - (\operatorname{div}_x K_\nu) \star \rho(x_i)\}. \end{aligned}$$

Lemma 10 is a direct consequence of Lemma 8 and Lemma 9. We note that Q_N^1 vanishes since

$$\operatorname{div}K_1 = \sum_h \partial_{x^h} \left(\sum_l \partial_{x^l} V_{hl} \right) = 0$$

since V is anti-symmetric. In the following, we actually treat a more general case when $\operatorname{div}K_1 \in L^\infty$. We now proceed to bound \mathcal{J}_1 and \mathcal{J}_2 respectively.

4.3 Control of the K_1 part with $K_1 = \operatorname{div}V$

In this section, we assume that $K_1 = \operatorname{div}V$, where V is matrix valued function with a singularity as (2.9). We use a more general assumption $\operatorname{div}K_1 \in L^\infty$, while $\operatorname{div}K_1 = 0$ for $K = \operatorname{div}V$ when V is anti-symmetric as in Theorem 3.

We decompose K_1 and correspondingly V_{hl} into bounded parts and singular parts. For each fixed time t , we choose a small parameter $\varepsilon_N(t) < 1$ (to be determined later) and define

$$K_1 = K_b + K_s, \quad K_b(x) = K_1(x) \mathbf{1}_{|x| \geq \varepsilon_N}(x).$$

Correspondingly we write $V_{hl} = V_{hl}^b + V_{hl}^s$ with $V_{hl}^b(x) = V_{hl}(x) \mathbf{1}_{|x| \geq \varepsilon_N}(x)$. Therefore

$$K_b^h = \sum_{l=1}^d \partial_{x^l} V_{hl}^b, \quad h = 1, \dots, d.$$

and

$$\sup_{1 \leq h, l \leq d} \|V_{hl}^b\|_{L^\infty} \leq C \sqrt{|\log |\varepsilon_N||}.$$

We can then decompose \mathcal{J}_1 defined in (4.18) as the following

$$\mathcal{J}_1 = \mathcal{J}_1^b + \mathcal{J}_1^s,$$

where

$$\begin{aligned} \mathcal{J}_1^b &= -\frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N G_N^{1,b} dX ds - \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N Q_N^{1,b} dX ds \\ &\quad - \frac{\sigma}{8N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX ds, \end{aligned} \tag{4.19}$$

and

$$\begin{aligned} \mathcal{J}_1^s &= -\frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N G_N^{1,s} dX ds - \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N Q_N^{1,s} dX ds \\ &\quad - \frac{\sigma}{8N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX ds, \end{aligned} \tag{4.20}$$

with

$$G_N^{1,b} = \frac{1}{N} \sum_{i,j=1}^N \nabla_{x_i} \log \rho(x_i) \{K_b(x_i - x_j) - K_b \star \rho(x_i)\}$$

and $G_N^{1,s}$, $Q_N^{1,b}$ and $Q_N^{1,s}$ can be defined similarly.

4.3.1 Control of \mathcal{J}_1^b : The bounded part of K_1

Recall that (4.18) with $\nu = 1$ and in this subsection we define further

$$I_1 := -\frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N G_N^{1,b} dX ds, \quad II_1 := -\frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N Q_N^{1,b} dX ds.$$

Now we proceed to bound the terms I_1 and II_1 above. We recall the definition of δV_{hl}^{ij} in (4.2) and $G_N^{1,b}$ etc. as above.

We firstly split the term I by integration by parts, that is

$$\begin{aligned} I_1 &= -\frac{1}{N} \int_0^t \int \rho_N G_N^{1,b} dX ds \\ &= -\frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \int_0^t \int \bar{\rho}_N \frac{\rho_N}{\rho_N} \partial_{x_i^h} \log \rho(x_i) \{(\partial_{x_i^l} V_{hl}^b)(x_i - x_j) - (\partial_{x_i^l} V_{hl}^b) \star \rho(x_i)\} dX ds \\ &= D_N^{1,b} + D_N^{2,b}, \end{aligned}$$

where

$$D_N^{1,b} = \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \int_0^t \int \rho_N \partial_{x_i^l} \left(\log \frac{\rho_N}{\rho_N} \right) \partial_{x_i^h} \log \rho(x_i) \delta(V_{hl}^b)^{ij} dX ds,$$

and

$$D_N^{2,b} = \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \int_0^t \int \rho_N \frac{1}{\rho(x_i)} \partial_{x_i^l} \partial_{x_i^h} \rho(x_i) \delta(V_{hl}^b)^{ij} dX ds.$$

Then applying Cauchy's inequality with $\varepsilon = \frac{\sigma}{8d}$, *i.e.* $ab \leq \varepsilon a^2 + \frac{1}{4\varepsilon} b^2$ for $a, b \in \mathbb{R}$,

we extract the diffusion of the relative entropy term out of D_N^1

$$\begin{aligned} |D_N^{1,b}| &\leq \frac{1}{N} \sum_{i=1}^N \sum_{h=1}^d \int_0^t \int \rho_N \sum_{l=1}^d \left| \partial_{x_i^l} \log \frac{\rho_N}{\rho_N} \right| \left| \frac{1}{N} \sum_{j=1}^N \partial_{x_i^h} \log \rho(x_i) \delta V_{hl}^{ij} \right| dX ds \\ &\leq \frac{\sigma}{8N} \int_0^t \int \rho_N |\nabla_X \log \frac{\rho_N}{\rho_N}|^2 + \frac{2d}{\sigma} \frac{1}{N} \sum_{i=1}^N \sum_{h,l=1}^d \int_0^t \int \rho_N \left(\frac{1}{N} \sum_{j=1}^N \partial_{x_i^h} \log \rho(x_i) \delta V_{hl}^{ij} \right)^2. \end{aligned}$$

Combining Lemma 9, the decomposition of I_1 and the estimate of $D_N^{1,b}$ above, one obtains

$$\mathcal{J}_1^b \leq M_1 + M_2, \quad (4.21)$$

where

$$M_1 \equiv \frac{2d}{\sigma} \frac{1}{N} \sum_{i=1}^N \sum_{h,l=1}^d \int_0^t \int_{\mathbb{D}^N} \rho_N \left(\frac{1}{N} \sum_{j=1}^N \partial_{x_i^h} \log \rho(x_i) \delta(V_{hl}^b)^{ij} \right)^2 dX ds,$$

and

$$M_2 \equiv \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N \Theta_N dX ds,$$

where

$$\Theta_N \equiv \frac{1}{N} \sum_{i,j=1}^N \left[\left(\sum_{h,l=1}^d R_{hl}(x_i) \delta(V_{hl}^b)^{ij} \right) - \Delta_b^{ij} \right] \quad (4.22)$$

recalling the definitions of functions $R_{hl}, R : \mathbb{D} \rightarrow \mathbb{R}$ in (2.11) in Theorem 3 and δV_{hl}^{ij} in (4.2) and Δ^{ij} in (4.3) but we use the bounded parts of V_{hl} and $\text{div}K_1$.

It suffices to control M_1 and M_2 now, which will be given by our main estimates Theorem 6 and Theorem 7.

We now proceed to bound M_1 and M_2 in (4.21). assuming Theorem 6 and Theorem 7.

Estimate on M_1 . Applying the Frenchel's inequality for the function $u(x) = x \log x$, that is for $x, y \geq 0$, $xy \leq x \log x + e^{y-1}$, we obtain that for any $\eta > 0$ (to be determined later)

$$\begin{aligned} M_1 &= \int_0^t \frac{2d}{\sigma} \frac{1}{N} \sum_{i=1}^N \sum_{h,l=1}^d \left[\frac{1}{\eta} \frac{1}{N} \int \bar{\rho}_N \frac{\rho_N}{\bar{\rho}_N} \eta \Upsilon_{hl}^{N,i} dX \right] ds \\ &\leq \int_0^t \frac{2d}{\sigma} \frac{1}{N} \sum_{i=1}^N \sum_{h,l=1}^d \left[\frac{1}{\eta} \frac{1}{N} \int \bar{\rho}_N \left(\frac{\rho_N}{\bar{\rho}_N} \log \frac{\rho_N}{\bar{\rho}_N} + \exp(\eta \Upsilon_{hl}^{N,i}) \right) dX \right] ds \\ &\leq \int_0^t \frac{2d^3}{\eta \sigma} \left(H_N(s) + \frac{1}{N} \sup_{1 \leq i \leq N} \sup_{1 \leq h,l \leq d} \int_{\mathbb{D}^N} \bar{\rho}_N \exp(\eta \Upsilon_{hl}^{N,i}) dX \right) ds, \end{aligned}$$

where we recall that as in Theorem 6 and its following corollary

$$\Upsilon_{hl}^{N,i} = N \left(\frac{1}{N} \sum_{j=1}^N \partial_{x_i^h} \log \rho(x_i) \delta(V_{hl}^b)^{ij} \right)^2 = \frac{1}{N} \sum_{j_1, j_2=1}^N (\partial_{x_i^h} \log \rho(x_i))^2 \delta(V_{hl}^b)^{ij_1} \delta(V_{hl}^b)^{ij_2}.$$

Consequently, as long as $\sqrt{\eta} \|V^b\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty} < \frac{1}{2e}$, applying Theorem 6 to

$\tilde{V}^b = \sqrt{\eta} V^b$ and thus to $\tilde{\Upsilon}_{hl}^{N,i} = \eta \Upsilon_{hl}^{N,i}$, one has for any $t \in [0, T]$,

$$\sup_{1 \leq i \leq N} \sup_{1 \leq h, l \leq d} \int_{\mathbb{D}^N} \bar{\rho}_N \exp(\tilde{\Upsilon}_{hl}^{N,i}) dX \leq 3 \left(1 + \frac{5\tilde{\alpha}}{(1-\tilde{\alpha})^3} + \frac{\tilde{\beta}}{1-\tilde{\beta}} \right),$$

where

$$\tilde{\alpha} = (2e\sqrt{\eta} \|V^b\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty})^4 < 1, \quad \tilde{\beta} = (2\sqrt{2e\eta} \|V^b\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty})^4 < 1. \quad (4.23)$$

Consequently, M_1 in (4.21) can be estimated with

$$M_1 \leq \int_0^t \frac{d^3}{2\eta\sigma} \left[H_N(s) ds + \frac{3\Lambda_1}{N} \right] ds, \quad (4.24)$$

where $0 < \eta < \left(\frac{1}{2e\|V^b\|_{L^\infty} \|\nabla_x \log \rho\|_{L^\infty}} \right)^2$ and

$$\Lambda_1 \equiv 1 + \frac{5\tilde{\alpha}}{(1-\tilde{\alpha})^3} + \frac{\tilde{\beta}}{1-\tilde{\beta}}. \quad (4.25)$$

The definition of $\tilde{\alpha}$ and $\tilde{\beta}$ is given above in (4.23). Here $\eta^{-1} \sim \|V^b\|_{L^\infty}^2 \sim |\log \varepsilon_N|$,

where ε_N is the cut-off parameter which can be time dependent.

Estimate on M_2 . Finally we estimate M_2 in (4.21). By the same trick used in bounding M_1 , for a new $\eta > 0$ which might be smaller than the previous one, we have

$$\begin{aligned} M_2 &= \int_0^t \frac{1}{\eta N} \int_{\mathbb{D}^N} \bar{\rho}_N \left(\frac{\rho_N}{\bar{\rho}_N}(\eta\Theta_N) \right) dX ds \\ &\leq \int_0^t \frac{1}{\eta N} \int_{\mathbb{D}^N} \bar{\rho}_N \left(\frac{\rho_N}{\bar{\rho}_N} \log \frac{\rho_N}{\bar{\rho}_N} + \exp(\eta\Theta_N) \right) dX ds \\ &\leq \int_0^t \frac{1}{\eta} H_N(s) ds + \frac{1}{\eta N} \int_0^t \int_{\mathbb{D}^N} \bar{\rho}_N \exp(\eta\Theta_N) dX ds. \end{aligned}$$

Choose η small enough such that

$$\tilde{\gamma} := \left(C [\eta \|V^b\|_{L^\infty} + \eta \|\operatorname{div}_x K_b\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{\|R\|_{L^p(\rho dx)}}{p} + 1 \right) \right)^2 < 1, \quad (4.26)$$

where C is a universal constant as in Theorem 7. Then applying Theorem 7 to $\tilde{V}^b = \eta V^b$ and therefore $\tilde{K}_b = \eta K_b$, $\tilde{\Theta}_N = \eta \Theta_N$, one has for $t \in [0, T]$

$$\int_{\mathbb{D}^N} \bar{\rho}_N \exp(\tilde{\Theta}_N) dX \leq \frac{3}{1 - \tilde{\gamma}}.$$

Therefore, M_2 in (4.21) can be estimated as follows

$$M_2 \leq \int_0^t \left[\frac{1}{\eta} H_N(s) + \frac{3}{\eta N(1 - \tilde{\gamma})} \right] ds, \quad (4.27)$$

where η should satisfy (4.26).

Combing (4.21), (4.24) and (4.27), we have

$$\mathcal{J}_1^b \leq \int_0^t \frac{1}{\eta} \left(\frac{d^3}{2\sigma} + 1 \right) \left[H_N(s) + \frac{3}{N} \left(\frac{d^3 \Lambda_1}{2\sigma} + \frac{1}{1 - \tilde{\gamma}} \right) \right] ds$$

where η is a small fixed constant satisfies both (4.23) and (4.26) and Λ_0, Λ_1 and $\tilde{\gamma}$ are all fixed constants given in (4.14), (4.25) and (4.26) respectively. In particular, we remark that η can be choose according to

$$\frac{1}{\eta} \sim 1 + \|V^b\|_{L^\infty}^2 + \|\operatorname{div} K_b\|_{L^\infty} = 1 + \|K_b\|_{W^{-1,\infty}}^2 + \|\operatorname{div} K\|_{L^\infty}. \quad (4.28)$$

For very small $\varepsilon_N = \varepsilon_N(t) \ll 1$, one can choose $\eta^{-1} = C |\log \varepsilon_N|$. Therefore we can obtain

Proposition 11 *The contribution \mathcal{J}_1^b of the bounded part of K_1 can be bounded as*

$$\mathcal{J}_1^b \leq C \int_0^t |\log \varepsilon_N| \left[H_N(s) + \frac{C}{N} \right] ds, \quad (4.29)$$

where C is a universal constant depending on the uniform bounds in the assumptions (2.10), $\|\nabla_x \log \rho\|_{L^\infty}$ in particular, and the dimension d , the viscosity σ and the assumptions on K_1 or V_{hl} .

4.3.2 Control of \mathcal{J}_1^s : The singular part of K_1

The following part is to control \mathcal{J}_1^s . As what we did in the previous subsection, we split the first term in \mathcal{J}_1^s (4.20) by integration by parts

$$-\frac{1}{N} \int_0^t \int \rho_N G_N^{1,s} dX ds = D_N^{1,s} + D_N^{2,s},$$

where

$$D_N^{1,s} = \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \int_0^t \int \rho_N \partial_{x_i^l} \left(\log \frac{\rho_N}{\bar{\rho}_N} \right) \partial_{x_i^h} \log \rho(x_i) [V_{hl}^s(x_i - x_j) - V_{hl}^s \star \rho(x_i)],$$

and

$$D_N^{2,s} = \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \int_0^t \int \rho_N \frac{1}{\rho(x_i)} \partial_{x_i^l} \partial_{x_i^h} \rho(x_i) [V_{hl}^s(x_i - x_j) - V_{hl}^s \star \rho(x_i)].$$

Since V is anti-symmetric and ρ is smooth ($\rho \in W_{loc}^{2,p}$),

$$\sum_{h,l=1}^d \int \rho_N \frac{1}{\rho(x_i)} \partial_{x_i^l} \partial_{x_i^h} \rho(x_i) \delta V_{hl}^{ij} dX ds = 0,$$

and therefore $D_N^{2,s} = 0$. Again for the case that V is anti-symmetric, $Q_N^{1,s} = 0$. Even under a more general assumption, for instance $\operatorname{div} K_1 \in L^\infty$ or even with a small singularity, we can still get similar results for \mathcal{J}_1^s .

Now we only need to consider $D_N^{1,s}$. Applying the Cauchy-Schwarz inequality to $D_N^{1,s}$, one has

$$\begin{aligned} D_N^{1,s} &\leq \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \frac{\sigma}{8d} \int_0^t \int \rho_N |\partial_{x_i^l} \log \frac{\rho_N}{\bar{\rho}_N}|^2 dX ds \\ &\quad + \frac{1}{N^2} \sum_{i,j=1}^N \sum_{h,l=1}^d \frac{2d}{\sigma} \|\nabla_x \log \rho\|_{L^\infty}^2 \int_0^t \int \rho_N |V_{hl}^s(x_i - x_j) - V_{hl}^s \star \rho(x_i)|^2 dX ds. \end{aligned}$$

Then the \mathcal{J}_1^s defined in (4.19) can be estimated as

$$\mathcal{J}_1^s \leq \frac{2d^3}{\sigma} \|\nabla_x \log \rho\|_{L^\infty}^2 \sup_{1 \leq i, j \leq N} \sup_{1 \leq h, l \leq d} \int_0^t \int \rho_N |V_{hl}^s(x_i - x_j) - V_{hl}^s \star \rho(x_i)|^2 dX ds. \quad (4.30)$$

To complete the estimate of \mathcal{J}_1^s (4.30), we only need to bound for $i \neq j$

$$\int_{\mathbb{D}^N} \rho_N |V_{hl}^s(x_i - x_j)|^2 dX = \int_{\mathbb{D}^2} \rho_{N,2} |V_{hl}^s(x_1 - x_2)|^2 dx_1 dx_2$$

and similarly

$$\int \rho_{N,1} |V_{hl}^s \star \rho(x)|^2 dx.$$

Indeed, we need to show the smallness of the following quantity

$$\sup_{N \geq 2} \int_0^T \int_{\mathbb{D}^2} \rho_{N,2} \frac{\mathbf{1}_{|x_1 - x_2| \leq \varepsilon_N}}{|x_1 - x_2|^\alpha} dx_1 dx_2 dt, \quad (4.31)$$

or equivalently the functional

$$\int_0^T \mathbb{E} \frac{1}{N^2} \sum_{i \neq j} \frac{\mathbf{1}_{|X_i - X_j| \leq \varepsilon_N}}{|X_i - X_j|^\alpha} dt,$$

where $\alpha > 0$ is a index to be determined. The estimates for these quantities rely on the bound given by the dissipation of the entropy or the Fisher information.

Proposition 12 (Control of the average distance) *For any fixed time $T > 0$, one has*

$$\sup_{N \geq 2} \int_0^T \mathbb{E} \frac{1}{N^2} \sum_{i \neq j} \frac{1}{|X_i - X_j|^\beta} dt < C_T < \infty, \quad (4.32)$$

where $0 < \beta \leq 2$ for $d \geq 3$ and $\beta < 2$ for $d = 2$ and C_T is a constant depending on T .

Assuming the above proposition, we can easily bound the contribution from the singular part. Indeed, one has the following proposition

Corollary 3 For a small parameter $\varepsilon_N = \varepsilon_N(t)$,

$$\sup_{N \geq 2} \int_0^T \mathbb{E} \frac{1}{N^2} \sum_{i \neq j} \frac{\mathbf{1}_{|X_i - X_j| \leq \varepsilon_N(t)}}{|X_i - X_j|^\alpha} dt \leq \int_0^t \varepsilon_N(t)^{\beta - \alpha} h(t) dt < \infty, \quad (4.33)$$

where $0 \leq h \in L^1_{loc}$ with $\int_0^T h(t) dt < C_T$ and C_t is defined in the previous Proposition 12.

Assuming Corollary 3, one can thus estimate \mathcal{J}_1^s as per

Proposition 13 The contribution \mathcal{J}_1^s of the singular part of K_1 can be estimated as

$$\mathcal{J}_1^s \leq C \int_0^t \varepsilon_N(s) (h(s) + 1) ds, \quad (4.34)$$

where $h \in L^1_{loc}$ and C is a universal constant depending on the dimension d , the viscosity σ and the assumptions (2.10) on ρ in particular $\|\nabla_x \rho\|_{L^\infty}$ and $\|\rho\|_{L^\infty}$.

Proof of Prop.13. Apply Corollary 3 to $|V_{hl}^s(x)| \leq C \sqrt{|\log|x||} \leq C/|x|^{1/4}$ for $\varepsilon_N \ll 1$. Indeed,

$$\int_0^t \int_{\mathbb{D}^2} \rho_{N,2} |V_{hl}^s(x_1 - x_2)|^2 dx_1 dx_2 ds \leq \int_0^t \varepsilon_N(s) h(s) ds,$$

if we take β very close to 2 and $\alpha = 1/2$ and where $h \in L^1_{loc}$. By the definition of the convolution,

$$|V_{hl}^s \star \rho(x)| \leq \|\rho\|_{L^\infty} \int_{|x| \leq \varepsilon_N} \frac{C}{|x|^{1/4}} dx \leq C \|\rho\|_{L^\infty} \varepsilon_N^{d-1/4} \leq C \varepsilon_N$$

and

$$\int \rho_{N,1} |V_{hl}^s \star \rho(x)|^2 dx \leq C \varepsilon_N$$

for $d = 2, 3, \dots$. \square

The proof of the above Corollary is straightforward. We only give a short proof of the Proposition 12.

Proof of Proposition 12. Since the joint law ρ_N is symmetric, we only need to bound the quantity

$$\mathcal{I}_N^\beta = \int_0^T \int_{\mathbb{D}^2} \frac{\rho_{N,2}}{|x_1 - x_2|^\beta} dx_1 dx_2 dt.$$

Recalling the inequality (4.16) in the proof of Lemma 9, one has

$$\sigma_N \frac{1}{N} \int_0^T \int_{\mathbb{D}^N} \frac{|\nabla_X \rho_N|^2}{\rho_N} dX ds \leq C_T^1,$$

where

$$C_T^1 = 2 \frac{1}{N} \int \rho_N^0 \log \rho_N^0 dX + 2T \left(\frac{\|K_2\|_{L^\infty}^2}{\sigma} + \|\operatorname{div} K_1\|_{L^\infty} + \|\operatorname{div}_x F\|_{L^\infty} \right).$$

Again by symmetry, one has

$$\int_0^T \int_{\mathbb{D}^N} \frac{|\nabla_{x_1} \rho_N|^2}{\rho_N} dX ds \leq C_T^1 / \sigma_N.$$

Combing with the trivial inequality

$$\int_0^T \int_{\mathbb{D}^N} |\nabla_{x_1} \log \rho_{N,2}(t, x_1, x_2) - \nabla_{x_1} \log \rho_N(t, X)|^2 \rho_N dX dt \geq 0,$$

one then obtains

$$\int_0^T \int_{\mathbb{D}^2} \frac{|\nabla_{x_1} \rho_{N,2}|^2}{\rho_{N,2}} dx_1 dx_2 dt \leq \int_0^T \int_{\mathbb{D}^N} \frac{|\nabla_{x_1} \rho_N|^2}{\rho_N} dX ds \leq C_T^1 / \sigma_N.$$

Choose $L(z) = 1/|z|^\beta - C$ where $C = \int_{\mathbb{D}} 1/|z|^\beta dz$ for $\beta \leq 2$ for $d \geq 3$ and in particular $\beta < d$ for $d = 1, 2$. Solve the Poisson equation

$$-\Delta \varphi = L,$$

then one has

$$\nabla\varphi = \phi, \quad -\operatorname{div}\phi = L.$$

One can choose ϕ such that $|\phi(z)| \leq C/|z|^{\beta-1}$. Now one has

$$\begin{aligned} \mathcal{I}_N^\beta &= CT + \int_0^T \int_{\mathbb{D}^2} \rho_{N,2}(t, x_1, x_2) L(x_1 - x_2) dx_1 dx_2 dt \\ &= CT + \int_0^T \int_{\mathbb{D}^2} \nabla_{x_1} \rho_{N,2} \cdot \phi(x_1 - x_2) dx_1 dx_2 dt \\ &\leq CT + \left(\int_0^T \int_{\mathbb{D}^2} \frac{|\nabla_{x_1} \rho_{N,2}|^2}{\rho_{N,2}} \right)^{1/2} \left(\int_0^T \int_{\mathbb{D}^2} \rho_{N,2} |\phi|^2 \right)^{1/2} \\ &\leq CT + C \sqrt{C_T^1 / \sigma_N} \left(\int_0^T \int_{\mathbb{D}^2} \frac{\rho_{N,2}}{|x_1 - x_2|^{2\beta-2}} \right)^{1/2}. \end{aligned}$$

As long as $\beta \leq 2$, one has

$$\frac{1}{|x_1 - x_2|^{2\beta-2}} \leq d^{(2-\beta)/2} \frac{1}{|x_1 - x_2|^\beta}.$$

Consequently we can estimate as

$$\mathcal{I}_N^\beta \leq CT + C \sqrt{C_T^1 / \sigma_N} d^{(2-\beta)/4} \sqrt{\mathcal{I}_N^\beta},$$

which implies

$$\mathcal{I}_N^\beta \leq C_T := 2CT + 2d^{(2-\beta)/2} C^2 C_T^1 / \sigma_N,$$

which completes the proof. \square

4.4 Control of the K_2 part with $K_2 \in L^\infty$

We recall that for a function ϕ ,

$$\delta\phi^{ij} = \phi(x_i - x_j) - \phi \star \rho(x_i),$$

according to the definition (4.2). In this section, for a vector field $K = (K^1, \dots, K^d)$,

we also write

$$\delta K^{ij} = K(x_i - x_j) - K \star \rho(x_i), \quad \delta(K^h)_{ij} = K^h(x_i - x_j) - K^h \star \rho(x_i) \quad (4.35)$$

where $h = 1, \dots, d$.

By the similar method in the previous two sections, one can easily obtain the following

Proposition 14 *The contribution \mathcal{J}_2 from the K_2 part with $K_2 \in L^\infty$ can be estimated as*

$$\mathcal{J}_2 \leq \int_0^t \frac{d}{\eta\sigma} \left[H_N(s) + \frac{C}{N} \right] ds, \quad (4.36)$$

where η^{-1} is in the order of $\|K_2\|_{L^\infty}^2$ and C is a universal constant depending on $\|K_2\|_{L^\infty}$.

Proof of Prop. 14. Now we proceed to control \mathcal{J}_2 as in (4.18) with $\nu = 2$, that is

$$\begin{aligned} \mathcal{J}_2 &= -\frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N G_N^2 dX ds - \frac{1}{N} \int_0^t \int_{\mathbb{D}^N} \rho_N Q_N^2 dX ds \\ &\quad - \frac{\sigma}{4N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX ds. \end{aligned}$$

Firstly, we rewrite the 2nd term above as

$$\begin{aligned} &-\frac{1}{N} \int \rho_N Q_N^2 dX \\ &= -\frac{1}{N^2} \sum_{i,j=1}^N \int \bar{\rho}_N \frac{\rho_N}{\bar{\rho}_N} [(\operatorname{div}_x K_2)(x_i - x_j) - (\operatorname{div}_x K_2) \star \rho(x_i)] dX \\ &= \frac{1}{N^2} \sum_{i,j=1}^N \left\{ \int \rho_N \nabla_{x_i} \log \rho(x_i) \delta K_2^{ij} dX + \int \rho_N \nabla_{x_i} \log \frac{\rho_N}{\bar{\rho}_N} \delta K_2^{ij} dX \right\} \end{aligned}$$

by integration by parts, where δK_2^{ij} is defined in (4.35).

Consequently, one has

$$\begin{aligned} \mathcal{J}_2 &= \frac{1}{N^2} \sum_{i,j=1}^N \int_0^t \int \rho_N \nabla_{x_i} \log \frac{\rho_N}{\bar{\rho}_N} \delta K_2^{ij} dX ds - \frac{\sigma}{4N} \int_0^t \int_{\mathbb{D}^N} \rho_N \left| \nabla_X \log \frac{\rho_N}{\bar{\rho}_N} \right|^2 dX ds \\ &\leq \frac{1}{\sigma N} \sum_{i=1}^N \int_0^t \int \rho_N \left(\frac{1}{N} \sum_{j=1}^N \delta K_2^{ij} \right)^2 dX ds. \end{aligned}$$

The last inequality is ensured by the Cauchy-Schwarz inequality. For fixed i , one has

$$\int \rho_N \left(\frac{1}{N} \sum_{j=1}^N \delta K_2^{ij} \right)^2 dX = \sum_{h=1}^d \int \rho_N \left(\frac{1}{N} \sum_{j=1}^N \delta(K_2^h)_{ij} \right)^2 dX.$$

By the same trick in bounding \mathcal{J}_1 and applying the main estimate Theorem 6 with $g \equiv 1$, one has for η small such that $\sqrt{\eta}\|K_2\|_{L^\infty} < \frac{1}{2e}$,

$$\int \rho_N \left(\frac{1}{N} \sum_{j=1}^N \delta(K_2^h)_{ij} \right)^2 dX \leq \frac{1}{\eta} H_N(t) + \frac{3\Lambda_2}{\eta N},$$

where

$$\Lambda_2 \equiv 1 + \frac{5\bar{\alpha}}{(1-\bar{\alpha})^3} + \frac{\bar{\beta}}{1-\bar{\beta}} < \infty \quad (4.37)$$

since

$$\bar{\alpha} = (2e\sqrt{\eta}\|K_2\|_{L^\infty})^4 < 1, \quad \bar{\beta} = \left(2\sqrt{2e\eta}\|K_2\|_{L^\infty}\right)^4 < 1.$$

Integrating over time completes the proof. \square

4.5 Final step of the Proof of Theorem 3

Now we can prove Theorem 3. Indeed, combining Lemma 10, Prop. 11, Prop. 13 and Prop. 14, one finally reaches

$$\begin{aligned} H_N(t) &\leq H_N(0) + \Lambda_0(\sigma - \sigma_N)^2 \\ &+ C \int_0^t |\log \varepsilon_N(s)| \left(H_N(s) + \frac{C}{N}\right) ds + C \int_0^t \varepsilon_N(s)(h(s) + 1) ds. \end{aligned} \quad (4.38)$$

If one chooses the parameter $\varepsilon_N(s) = H_N(s) + \frac{C}{N}$ at time s , then the above inequality reduces to

$$H_N(t) \leq H_N(0) + \Lambda_0(\sigma - \sigma_N)^2 + C \int_0^t |\log(H_N(s) + \frac{C}{N})| \left(H_N(s) + \frac{C}{N}\right) (1+h(s)) ds.$$

By Gronwall's inequality, one finally reaches

$$H_N(t) \leq \left(H_N(0) + \frac{C}{N} + \Lambda_0(\sigma - \sigma_N)^2 \right)^{\exp(-C \int_0^t (1+h(s)) ds)},$$

where C is a universal constant depending on the time length t , the dimension d , the viscosity σ and some uniform bounds in (2.10) and assumptions on K while $0 \leq h \in L^1_{loc}$ with $\int_0^t h \, ds = C_t$ as in Corollary 3. Other parts follows exactly as the 2nd order cases.

This completes the proof of Theorem 3.

Chapter 5: Preliminary of combinatorics

In this chapter, we list several classical combinatorics results that will be used in the proof of the main estimates. We first recall Stirling's formula for $n = 1, 2, \dots$,

$$n! = \lambda_n \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad (5.1)$$

where $1 < \lambda_n < \frac{11}{10}$ and $\lambda_n \rightarrow 1$ as $n \rightarrow \infty$. A straightforward consequence of the Stirling's formula is

$$p^p \leq p! e^p,$$

which will be used frequently for simplicity.

Lemma 11 *For any $1 \leq p \leq q$, one has*

$$\binom{q}{p} \leq e^p q^p p^{-p}.$$

Proof of Lemma 11 The proof is straightforward by the Stirling's formula. \square

Lemma 12 *For any $1 \leq p \leq q$, one has*

$$|\{(b_1, \dots, b_p) \in \mathbb{N}^p \mid \forall l \ 1 \leq b_l \leq q, \quad b_1 + b_2 + \dots + b_p = q\}| = \binom{q-1}{p-1}.$$

Proof of Lemma 12 When $p = 1$, the lemma trivially holds true with the convention $\binom{0}{0} = 1$ if $p = q = 1$. We thus assume $p \geq 2$ in the following. Since each p -tuple

(b_1, b_2, \dots, b_p) uniquely determines a $(p-1)$ -tuple $(c_1, c_2, \dots, c_{p-1})$ and reciprocally via

$$c_1 = b_1, c_2 = b_1 + b_2, \dots, c_{p-1} = b_1 + b_2 + \dots + b_{p-1},$$

it suffices to verify that

$$|\{(c_1, c_2, \dots, c_{p-1}) \mid 1 \leq c_1 < c_2 < \dots < c_{p-1} \leq q-1\}| = \binom{q-1}{p-1}.$$

This is simply obtained by choosing $p-1$ distinct integers from the set $\{1, 2, \dots, q-1\}$ and assigning the smallest one to c_1 , the second smallest to c_2 , and so on. \square

Much of the combinatorics that we handle is based only on the multiplicity in the multi-indices. It is therefore convenient to know how many multi-indices can have the same multiplicity signature

Lemma 13 *For any $a_1, \dots, a_q \in \mathbb{N}$ s.t. $a_1 + \dots + a_q = p$, then the set of multi-indices $I_p = (i_1, \dots, i_p)$ with $1 \leq i_k \leq q$ and corresponding multiplicities has cardinal*

$$\left| \left\{ (i_1, \dots, i_p) \in \{1, \dots, q\}^p \mid \forall l \ a_l = |\{k \mid i_k = l\}| \right\} \right| = \frac{p!}{a_1! \dots a_q!}.$$

Proof This is the basic multinomial relation: We have to choose 1 a_1 times among p positions, 2 a_2 times among the remaining positions and so on... \square

Similarly as for the binomial coefficients, $\frac{p!}{a_1! \dots a_q!}$ is the coefficient of $x_1^{a_1} \dots x_q^{a_q}$ in the expansion of $(x_1 + \dots + x_q)^p$ leading to the obvious estimate

$$\sum_{a_1, \dots, a_q \geq 0, a_1 + \dots + a_q = p} \frac{p!}{a_1! \dots a_q!} = q^p. \tag{5.2}$$

Definitions To let the presentation simple, let us introduce some notations here.

We write the integer valued p -tuple as $I_p = (i_1, \dots, i_p)$. The overall set $\mathcal{T}_{q,p}$ of those indices is defined as

$$\mathcal{T}_{q,p} = \{I_p = (i_1, \dots, i_p) | 1 \leq i_\nu \leq q, \text{ for all } 1 \leq \nu \leq p\}. \quad (5.3)$$

We thus define the *multiplicity function* $\Phi_{q,p} : \mathcal{T}_{q,p} \rightarrow \{0, 1, \dots, p\}^q$ with $\Phi_{q,p}(I_p) = A_q = (a_1, a_2, \dots, a_q)$, where

$$a_l = |\{1 \leq \nu \leq p | i_\nu = l\}|.$$

With the multiplicity function $\Phi_{q,p}$, we can proceed to define the “*effective set*” $\mathcal{E}_{q,p}$ of index I_p as

$$\mathcal{E}_{q,p} = \{I_p \in \mathcal{T}_{q,p} | \Phi_{q,p}(I_p) = A_q = (a_1, \dots, a_q) \text{ with } a_\nu \neq 1 \text{ for any } 1 \leq \nu \leq q\}. \quad (5.4)$$

The cardinality of the effective set $\mathcal{E}_{N,2k}$ or $\mathcal{E}_{N,4k}$ play a crucial role in the combinatorics argument later, which is given by the following lemma

Lemma 14 *Assume that $1 \leq p \leq q$. Then*

$$|\mathcal{E}_{q,p}| \leq \sum_{l=1}^{\lfloor \frac{p}{2} \rfloor} \binom{q}{l} l^p \leq \lfloor \frac{p}{2} \rfloor \binom{q}{\lfloor \frac{p}{2} \rfloor} \left(\lfloor \frac{p}{2} \rfloor \right)^p \leq \frac{p}{2} e^{\frac{p}{2}} q^{\frac{p}{2}} \left(\frac{p}{2} \right)^{\frac{p}{2}}. \quad (5.5)$$

Proof If $p = 1$, then $\mathcal{E}_{q,p} = \emptyset$. The estimate (5.5) holds trivially. Hence we assume that $p \geq 2$. Assume that there are l distinct integers in I_p , then $1 \leq l \leq \lfloor \frac{p}{2} \rfloor$ by the definition of $\mathcal{E}_{q,p}$.

We count the number $|\{I_p \in \mathcal{E}_{q,p} | |\{i_1, \dots, i_p\}| = l\}|$ by the multiple principle. Indeed, at the first step, we choose l distinct integers from $\{1, 2, \dots, q\}$. There are

$\binom{q}{l}$ choices in this step. Then we order those integers at p positions with possible repetition. The total number of all orderings can be bounded by l^p , the total number without any restriction. Therefore one reaches

$$|\mathcal{E}_{q,p}| = \sum_{l=1}^{\lfloor \frac{p}{2} \rfloor} |\{I_p \in \mathcal{E}_{q,p} \mid |\{i_1, \dots, i_p\}| = l\}| = \sum_{l=1}^{\lfloor \frac{p}{2} \rfloor} \binom{q}{l} l^p.$$

Simply applying the Lemma 11, we complete the proof. \square

Chapter 6: Main estimates: The 2nd order case

In this chapter, we establish the main estimates Theorem 5 for the 2nd order case. This chapter is the most technical part in the article [93]. We simplify the proof and emphasize the key properties to make use of the *Laws of Large Numbers* corresponding to certain cancellation rules.

6.1 Intuitive calculations: the scaling of R_N

We present some of the basic scaling properties of R_N . Recall that the definition (3.1) of R_N in the 2nd order case, that is the double summation

$$R_N = \frac{1}{N} \sum_{i,j=1}^N \nabla_{v_i} \log f(x_i, v_i) \cdot \{K(x_i - x_j) - K \star \rho(x_i)\}.$$

Then a trivial bound for $|R_N|$ is simply

$$|R_N| \leq (2\|K\|_{L^\infty} \|\nabla_v \log f\|_{L^\infty}) N. \quad (6.1)$$

However inserting this bound in the evolution of the relative entropy

$$H_N(t) \leq H_N(0) - \frac{1}{N} \int_0^t \int f_N R_N \, dZ \, ds$$

for the purely deterministic case for instance would only give that $H_N(t) = O(1)$ without any chance of converging. Instead Theorem 5 essentially proves that R_N is of order 1 and not of order N , which indicates the need of certain cancellation rules.

To get

$$\int_{(\mathbb{D} \times \mathbb{R}^d)^N} \bar{f}_N \exp(|R_N|) dZ \leq C < \infty,$$

where C doesn't depend on N , we expand $\exp(|R_N|)$ by Taylor expansion. Note though that

$$\begin{aligned} \frac{1}{(2k+1)!} |R_N|^{2k+1} &\leq \frac{1}{(2k+1)!} |R_N|^{2k} \left(\frac{2k+1}{2} + \frac{1}{2(2k+1)} |R_N|^2 \right) \\ &\leq \frac{1}{2} \frac{1}{(2k)!} |R_N|^{2k} + \frac{1}{(2k+2)!} |R_N|^{2k+2}, \end{aligned}$$

so that we only have to bound the even terms and have

$$\exp(|R_N|) = \sum_{k=0}^{\infty} \frac{1}{k!} |R_N|^k \leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} |R_N|^{2k}.$$

Consequently, we have

$$\int \bar{f}_N \exp(|R_N|) dZ \leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ. \quad (6.2)$$

The basic idea of the proof for Theorem 5 is to expand the sum defining R_N in $|R_N|^{2k}$ and show that a large number of terms vanish under integral with respect to \bar{f}_N . Indeed, the k -th term in (6.2) can be expanded as

$$\begin{aligned} &\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ \\ &= \frac{1}{(2k)!} \frac{1}{N^{2k}} \int \sum_{1 \leq i_1, j_1 \leq N} \cdots \sum_{1 \leq i_{2k}, j_{2k} \leq N} (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N dZ \\ &= \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{I_{2k} \in \mathcal{T}_{N, 2k}} \sum_{J_{2k} \in \mathcal{T}_{N, 2k}} \int (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N dZ. \end{aligned} \quad (6.3)$$

where we recall the definition (5.3) of $\mathcal{T}_{q,p}$ for $q = N$ and $p = 2k$ and we define in this chapter

$$F_i = \nabla_{v_i} \log f(x_i, v_i), \quad \delta K^{i,j} = K(x_i - x_j) - K \star \rho(x_i).$$

In the expansion (6.3), one typical term with fixed index (I_{2k}, J_{2k}) can be bounded simply by

$$\begin{aligned} & \left| \int (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N \, dZ \right| \\ & \leq (2\|K\|_{L^\infty})^{2k} \int |\nabla_{v_1} \log f(z_1)|^{a_1} \cdots |\nabla_{v_N} \log f(z_N)|^{a_N} \bar{f}_N \, dZ \end{aligned} \quad (6.4)$$

with (a_1, \dots, a_N) is just the multiplicity of I_{2k} . Even under stronger assumption $\|\nabla_v \log f\|_{L^\infty} < \infty$ compared to the one $\sup_p \frac{M_p}{p} < \infty$ in Theorem 5, another trivial bound for each term in the series (6.2) based on (6.4) will give us

$$\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ \leq \frac{1}{(2k)!} N^{2k} (2\|K\|_{L^\infty} \|\nabla_v \log f\|_{L^\infty})^{2k}, \quad (6.5)$$

which will blow up if we fix k but let $N \rightarrow \infty$. However, the above simple bound tells us we only need to focus on the case k is small compared to N : the trivial bound discussed here will be enough for the case $k \gg N$.

In the following, we only present two basic calculations, indicative of the type of cancellations that we use

Lemma 15 *Assume that $f \in L^\infty \cap L^1(\mathbb{D} \times \mathbb{R}^d)$ with $f \geq 0$ and $\int f = 1$. Assume that $K \in L^\infty$ and that $\nabla_v f \in L^1_{loc}$, then*

$$\int_{\mathbb{D}^N \times (\mathbb{R}^d)^N} R_N \bar{f}_N \, dZ = 0.$$

Proof Simply expanding R_N , we get

$$\begin{aligned} \int R_N \bar{f}_N \, dZ &= \frac{1}{N} \sum_{i,j=1}^N \int \nabla_{v_i} \log f(x_i, v_i) \cdot \{K(x_i - x_j) - K \star \rho(x_i)\} \\ & f(x_1, v_1) \cdots f(x_N, v_N) \, dZ. \end{aligned}$$

For fixed (i, j) , notice that $f(x_i, v_i) \nabla_{v_i} \log f(x_i, v_i) = \nabla_{v_i} f(x_i, v_i)$, and no other terms depend on v_i . Integration by parts thus implies that the integral vanishes.

Indeed, by Fubini's Theorem, without loss of generality, we only need to check

$$\int \nabla_v f(x, v) K \star \rho(x) dx dv = 0 \quad (6.6)$$

and

$$\int \nabla_{v_1} f(x_1, v_1) \{K(x_1 - x_2) - K \star \rho(x_1)\} \rho(x_2) dx_1 dv_1 dx_2 = 0. \quad (6.7)$$

□

Lemma 15 only illustrates the simplest cancellation in R_N . It is also straightforward to show some orthogonality property between the terms in the sum defining R_N . This leads to the first indication that indeed R_N is of order 1 and not N .

Lemma 16 *Assume that $f \in L^\infty \cap L^1(\mathbb{D} \times \mathbb{R}^d)$ with $f \geq 0$ and $\int f = 1$. Assume that $K \in L^\infty$ and that $\nabla_v f \in L^2_{loc}$, then*

$$\int_{\mathbb{D}^N \times (\mathbb{R}^d)^N} |R_N|^2 \bar{f}_N dZ \leq 4 \|K\|_{L^\infty}^2 \int_{\mathbb{D} \times \mathbb{R}^d} |\nabla_v \log f|^2 f dx dv.$$

Proof For convenience we recall

$$F_i = \nabla_{v_i} \log f(x_i, v_i), \quad \delta K^{i,j} = K(x_i - x_j) - K \star \rho(x_i).$$

Simply expand the left-hand side

$$\int |R_N|^2 \bar{f}_N dZ = \frac{1}{N^2} \sum_{i_1, i_2=1}^N \sum_{j_1, j_2=1}^N \int F_{i_1} \cdot \delta K^{i_1, j_1} F_{i_2} \cdot \delta K^{i_2, j_2} \bar{f}_N dZ.$$

If $i_1 \neq i_2$, then by integration by parts,

$$\int F_{i_1} \cdot \delta K^{i_1, j_1} F_{i_2} \cdot \delta K^{i_2, j_2} \bar{f}_N dZ = 0.$$

Indeed, without loss of generality, let $i_1 = 1$ and $i_2 = 2$, then

$$\begin{aligned} & \int F_{i_1} \cdot \delta K^{i_1, j_1} F_{i_2} \cdot \delta K^{i_2, j_2} \bar{f}_N dZ \\ &= \int_{(\mathbb{D} \times \mathbb{R}^d)^2} \nabla_{v_1} f(x_1, v_1) \cdot \delta K^{1, j_1} \nabla_{v_2} f(x_2, v_2) \cdot \delta K^{2, j_2} dz_1 dz_2 = 0, \end{aligned}$$

by integration by parts since $\delta K^{1,j_1}$ and $\delta K^{2,j_2}$ do not depend any v variables.

If $i_1 = i_2$ while $j_1 \neq j_2$, then at least one of $\{j_1, j_2\}$ is not equal to i_1 , then this type of integral vanishes by the definition of convolution. Indeed, without lost of generality, let assume that $i_1 = i_2 = 1$ and $j_1 = 2$ while $j_2 \neq 2$, then

$$\begin{aligned}
& \int_{(\mathbb{D} \times \mathbb{R}^d)^N} F_{i_1} \cdot \delta K^{i_1, j_1} F_{i_2} \cdot \delta K^{i_2, j_2} \bar{f}_N \, dZ \\
&= \int_{(\mathbb{D} \times \mathbb{R}^d)^N} [\nabla_{v_1} \log f(x_1, v_1) \cdot \{K(x_1 - x_2) - K \star \rho(x_1)\}] \\
&\quad \cdot [\nabla_{v_1} \log f(x_1, v_1) \cdot \{K(x_1 - x_{j_2}) - K \star \rho(x_1)\}] \bar{f}_N \, dZ \\
&= \int_{(\mathbb{D} \times \mathbb{R}^d)^{N-1}} [\nabla_{v_1} \log f(x_1, v_1) \cdot \{K(x_1 - x_{j_2}) - K \star \rho(x_1)\}] \Pi_{i \neq 2} f(x_i, v_i) \, dz_i \\
&\quad \cdot (\nabla_{v_1} \log f(x_1, v_1) \cdot \int_{\mathbb{D}} \{K(x_1 - x_2) - K \star \rho(x_1)\} \rho(x_2) \, dx_2) \\
&= 0,
\end{aligned}$$

where we used that

$$\int_{\mathbb{D}} \{K(x_1 - x_2) - K \star \rho(x_1)\} \rho(x_2) \, dx_2 = 0,$$

by the definition of convolution, and since ρ has integral 1.

Hence after integration only those terms with indices $i_1 = i_2$ and $j_1 = j_2$ contribute to the summation. That is

$$\begin{aligned}
& \frac{1}{N^2} \sum_{i_1, i_2=1}^N \sum_{j_1, j_2=1}^N \int F_{i_1} \cdot \delta K^{i_1, j_1} F_{i_2} \cdot \delta K^{i_2, j_2} \bar{f}_N \, dZ \\
&= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \int (F_i \cdot \delta K^{i, j})^2 \bar{f}_N \, dZ \leq 4 \|K\|_{L^\infty}^2 \int_{\mathbb{D} \times \mathbb{R}^d} |\nabla_v \log f|^2 f \, dx \, dv,
\end{aligned}$$

which completes the proof. \square

6.2 Main Estimates: Proof of Theorem 5

Now we are ready to give the complete proof of Theorem 5. From the discussion above, it only show the convergence of the series

$$\sum_{k=0}^{\infty} \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ$$

which we divide in two different cases: k is small compared to N or k is comparable or larger than N . The first part, $3k \leq N$, is more delicate and requires some preparatory combinatorics work. The second part, $3k > N$, is almost trivial since now the coefficients $\frac{1}{(2k)!}$ dominates and we can simply use the trivial bound similar to (6.5). We remark the general strategy is the same for the 1st order systems but it is more difficult there and the cancellation rules there are more tricky.

Accordingly Theorem 5 is a consequence of the following two propositions

Proposition 15 *For $3k \leq N$, we have*

$$\sum_{k=0}^{\lfloor \frac{N}{3} \rfloor} \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ \leq 1 + 2 \sum_{k=1}^{\lfloor \frac{N}{3} \rfloor} k \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k}.$$

Proposition 16 *For $3k > N$, we have*

$$\sum_{k=\lfloor \frac{N}{3} \rfloor + 1}^{\infty} \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ \leq \sum_{k=\lfloor \frac{N}{3} \rfloor + 1}^{\infty} \left(5e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k}.$$

Let us briefly explain how we can prove Theorem 5 from Proposition 15 and Proposition 16.

Proof of Theorem 5 Recall that

$$\sum_{k=1}^{\infty} k r^k = r \frac{d}{dr} \sum_{k=0}^{\infty} r^k = \frac{r}{(1-r)^2}.$$

Under the assumption $\|K\|_{L^\infty} \sup_p \frac{M_p}{p} < \frac{1}{8e^2}$, we have that

$$\begin{aligned} \sum_{k=1}^{\lfloor \frac{N}{3} \rfloor} k \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} &\leq \sum_{k=1}^{\infty} k \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} \\ &= \frac{\left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^2}{\left(1 - \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^2 \right)^2} < \infty, \end{aligned}$$

and

$$\sum_{k=\lfloor \frac{N}{3} \rfloor + 1}^{\infty} \left(5e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} \leq \sum_{k=1}^{\infty} \left(\frac{5}{8} \right)^{2k} \leq \frac{\left(\frac{5}{8} \right)^2}{1 - \left(\frac{5}{8} \right)^2} < \infty.$$

Hence, by (6.2), Proposition 15 and Proposition 16 we have that

$$\begin{aligned} \int \bar{f}_N \exp(|R_N|) dZ &\leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ \\ &\leq 3 \left(1 + 2 \sum_{k=1}^{\infty} k \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} + \sum_{k=1}^{\infty} \left(\frac{5}{8} \right)^{2k} \right) \\ &\leq 5 + 6 \left(\frac{8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right)}{1 - \left(8e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^2} \right)^2. \end{aligned}$$

This completes the proof. \square

6.2.1 The case $3k \leq N$: Proof of Proposition 15

We start with the general rule for cancellation in the expansion (6.3).

Lemma 17 (General Cancellation Rule) *Fix an integer $p \geq 1$. Take any pair of multi-indices (I_p, J_p) , where $I_p = (i_1, i_2, \dots, i_p)$ and $J_p = (j_1, j_2, \dots, j_p)$. All components of I_p and J_p are taken from the set $\{1, 2, \dots, N\}$. Then*

$$\begin{aligned} \int_{(\mathbb{D} \times \mathbb{R}^d)^N} \left(\nabla_{v_{i_1}} \log f(x_{i_1}, v_{i_1}) \cdot \{K(x_{i_1} - x_{j_1}) - K * \rho(x_{i_1})\} \right) \\ \cdots \left(\nabla_{v_{i_p}} \log f(x_{i_p}, v_{i_p}) \cdot \{K(x_{i_p} - x_{j_p}) - K * \rho(x_{i_p})\} \right) \bar{f}_N dZ = 0 \end{aligned} \quad (6.8)$$

provided that one of the following statements is satisfied:

- 1) there exists one i_ν , such that $i_\nu \notin \{i_1, \dots, i_{\nu-1}, i_{\nu+1}, \dots, i_p\}$;
- 2) there exists one j_ν , such that $j_\nu \notin \{i_1, i_2, \dots, i_p\} \cup \{j_1, \dots, j_{\nu-1}, j_{\nu+1}, \dots, j_p\}$.

Proof The proof of the this lemma is essentially the same as Lemma 15 and Lemma 16. For completeness, we give a short proof here. Let us first check the case 1) above. Without loss of generality, we can assume $i_\nu = i_1 = 1$ while $i_2 \neq 1, \dots, i_p \neq 1$. Now use the conventions F_i and $\delta K^{i,j}$ to simplify notations. Hence the integral becomes

$$\begin{aligned} & \int_{(\mathbb{D} \times \mathbb{R}^d)^N} (F_{i_1} \cdot \delta K^{1,j_1}) \cdot (F_{i_2} \cdot \delta K^{i_2,j_2}) \cdots (F_{i_p} \cdot \delta K^{i_p,j_p}) \bar{f}_N \, dZ \\ &= \int \nabla_{v_1} f(x_1, v_1) \cdot \delta K^{1,j_1} (F_{i_2} \cdot \delta K^{i_2,j_2}) \cdots (F_{i_p} \cdot \delta K^{i_p,j_p}) \prod_{i=2}^N f(x_i, v_i) \, dZ, \end{aligned}$$

where the only term depending on v_1 is $f(x_1, v_1)$. Integration by parts shows that (6.8) holds.

In the second case, without loss of generality, we can assume that $j_1 = 1$, while $j_2 \neq 1, \dots, j_p \neq 1$ and $i_1 \neq 1, \dots, i_p \neq 1$. Hence the integral becomes

$$\begin{aligned} & \int F_{i_1} \cdot \{K(x_{i_1} - x_1) - K * \rho(x_{i_1})\} (F_{i_2} \cdot \delta K^{i_2,j_2}) \cdots (F_{i_p} \cdot \delta K^{i_p,j_p}) \bar{f}_N \, dZ \\ &= \int_{(\mathbb{D} \times \mathbb{R}^d)^{N-1}} (F_{i_2} \cdot \delta K^{i_2,j_2}) \cdots (F_{i_p} \cdot \delta K^{i_p,j_p}) \prod_{i=2}^N f(x_i, v_i) \, dz_2 \cdots dz_N \\ & \quad \cdot \int_{\mathbb{D} \times \mathbb{R}^d} \nabla_{v_1} \log f(x_{i_1}, v_{i_1}) \cdot \{K(x_{i_1} - x_1) - K * \rho(x_{i_1})\} f(x_1, v_1) \, dx_1 \, dv_1 \\ &= \int_{(\mathbb{D} \times \mathbb{R}^d)^{N-1}} (F_{i_2} \cdot \delta K^{i_2,j_2}) \cdots (F_{i_p} \cdot \delta K^{i_p,j_p}) \prod_{i=2}^N f(x_i, v_i) \, dz_2 \cdots dz_N \\ & \quad \cdot \left(\nabla_{v_1} \log f(x_{i_1}, v_{i_1}) \cdot \int_{\mathbb{D} \times \mathbb{R}^d} \{K(x_{i_1} - x_1) - K * \rho(x_{i_1})\} f(x_1, v_1) \, dx_1 \, dv_1 \right), \end{aligned}$$

where only $K(x_{i_1} - x_1)$ and $f(x_1, v_1)$ are (x_1, v_1) -dependent. As in Lemma 16

$$\int_{\mathbb{D} \times \mathbb{R}^d} \{K(x_{i_1} - x_1) - K * \rho(x_{i_1})\} f(x_1, v_1) \, dx_1 \, dv_1 = 0,$$

and hence again (6.8) holds, completing the proof.

□

Using the notation $\mathcal{E}_{N,2k}$ introduced in (5.4), the first cancellation rule above means that once the index $I_{2k} \notin \mathcal{E}_{N,2k}$, then

$$\int (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N \, dZ = 0.$$

Therefore, we only need to count those couples (I_{2k}, J_{2k}) with $I_{2k} \in \mathcal{E}_{N,2k}$ and J_{2k} belongs to the set defined as

$$\mathcal{P}_{N,2k}^{I_{2k}} := \left\{ J_{2k} \in \mathcal{T}_{N,2k} \mid \begin{array}{l} \text{either for all } 1 \leq \nu \leq 2k, j_\nu \in \{i_1, \dots, i_{2k}\}; \\ \text{or for any } \nu \text{ such that } j_\nu \notin \{i_1, \dots, i_{2k}\}, \\ \exists \nu' \neq \nu, \text{ such that } j_\nu = j_{\nu'}. \end{array} \right\}.$$

We now turn to bounding the number of choices of J_{2k} in $\mathcal{P}_{N,2k}^{I_{2k}}$ with $I_{2k} \in \mathcal{E}_{N,2k}$.

Lemma 18 (*Choices of the multi-indices J_{2k}*) *Assume that $3k \leq N$ and $I_{2k} \in \mathcal{E}_{N,2k}$ with $|\{i_1, \dots, i_{2k}\}| = l$. Recall that $1 \leq l \leq k$. Then we have*

$$|\mathcal{P}_{N,2k}^{I_{2k}}| = l^{2k} + \sum_{h=2}^{2k} l^{2k-h} \binom{2k}{h} |\mathcal{E}_{N-l,h}|. \quad (6.9)$$

Furthermore,

$$|\mathcal{P}_{N,2k}^{I_{2k}}| \leq P_{N,2k} := 2k e^k 2^{2k} k^k N^k. \quad (6.10)$$

Proof Without loss of generality, we assume that as a set $\{i_1, \dots, i_{2k}\} = \{1, 2, \dots, l\}$. By the definition of the set $\mathcal{P}_{N,2k}^{I_{2k}}$, we have two cases. The first case is that all j_ν are chosen from $\{1, 2, \dots, l\}$. The total number of such J_{2k} is l^{2k} since each j_ν can be any integer from 1 to l .

In the second case, there exists some j_ν in $\{l+1, \dots, N\}$ and for each such $j_\nu \geq l+1$, there exists $\nu' \neq \nu$ such that $j_\nu = j_{\nu'}$. That is to say, each component

$j_\nu \geq l + 1$ is repeated. Denote by

$$h = |\{1 \leq \nu \leq 2k | j_\nu \geq l + 1\}|$$

the number of components of J_{2k} which are larger than l . We thus have $2 \leq h \leq 2k$.

For a fixed h , we need to choose h positions in J_{2k} to put integers bigger than l for $\binom{2k}{h}$ choices.

The remaining $(2k - h)$ positions of J_{2k} can be filled with any integer in $\{1, 2, \dots, l\}$, for l^{2k-h} choices.

Finally, we choose h integers from the set $\{l + 1, \dots, N\}$ for each of the h positions in J_{2k} that we chose initially. Again, the multiplicity for each integer chosen is at least two and the order is taken into account. This coincides with the definition of $\mathcal{E}_{N-l,h}$. Hence, in this step, the total number is just $|\mathcal{E}_{N-l,h}|$.

Therefore for a fixed h , one has that

$$|\{J_{2k} \in \mathcal{P}_{N,2k}^{I_{2k}} | h \text{ components of } J_{2k} \text{ are larger than } l\}| = \binom{2k}{h} l^{2k-h} |\mathcal{E}_{N-l,h}|.$$

Adding all the cases together, we obtain

$$\begin{aligned} |\mathcal{P}_{N,2k}^{I_{2k}}| &= l^{2k} + \sum_{h=2}^{2k} |\{J_{2k} \in \mathcal{P}_{N,2k}^{I_{2k}} | h \text{ components of } J_{2k} \text{ are larger than } l\}| \\ &= l^{2k} + \sum_{h=2}^{2k} \binom{2k}{h} l^{2k-h} |\mathcal{E}_{N-l,h}|, \end{aligned}$$

which is exactly (6.9).

Now we simplify the bound for $|\mathcal{P}_{N,2k}^{I_{2k}}|$. Applying Lemma 14, we have

$$|\mathcal{E}_{N-l,h}| \leq \frac{h}{2} e^{\frac{h}{2}} (N-l)^{\frac{h}{2}} \left(\frac{h}{2}\right)^{\frac{h}{2}}.$$

Therefore

$$\begin{aligned}
|\mathcal{P}_{N,2k}^{I_{2k}}| &\leq l^{2k} + \sum_{h=2}^{2k} l^{2k-h} \binom{2k}{h} h e^{\frac{h}{2}} (N-l)^{\frac{h}{2}} \left(\frac{h}{2}\right)^{\frac{h}{2}} \\
&\leq l^{2k} + 2k e^k \sum_{h=2}^{2k} l^{2k-h} \binom{2k}{h} (N-l)^{\frac{h}{2}} k^{\frac{h}{2}} \\
&\leq 2k e^k \left\{ \sum_{h=0}^{2k} \binom{2k}{h} l^{2k-h} (N-l)^{\frac{h}{2}} k^{\frac{h}{2}} \right\} = 2k e^k \left(l + \sqrt{k(N-l)} \right)^{2k} \\
&\leq 2k e^k 2^{2k} k^k N^k.
\end{aligned}$$

This completes the proof.

□

We are now ready to prove Prop. 15 by combining Lemma 14 and Lemma 18.

Proof of Prop. 15 Applying Lemma 17, the expansion in 6.3 becomes

$$\begin{aligned}
&\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ \\
&= \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{I_{2k} \in \mathcal{E}_{N,2k}} \sum_{J_{2k} \in \mathcal{P}_{N,2k}^{I_{2k}}} \int (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N \, dZ.
\end{aligned} \tag{6.11}$$

Combining the typical bound (6.4) and the definition $M_p = \|\nabla_v \log f\|_{L^p(df)}$, *i.e.*

$$\begin{aligned}
&\int |\nabla_{v_1} \log f(x_1, v_1)|^{a_1} \cdots |\nabla_{v_N} \log f(x_N, v_N)|^{a_N} \bar{f}_N \, dZ \\
&= M_{a_1}^{a_1} M_{a_2}^{a_2} \cdots M_{a_N}^{a_N} \leq \left(\sup_p \frac{M_p}{p} \right)^{2k} a_1^{a_1} \cdots a_N^{a_N},
\end{aligned}$$

with the convention that $0^0 = 1$, we obtain

$$\begin{aligned}
&\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ \\
&= \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{l=1}^k \sum_{I_{2k} \in \mathcal{E}_{N,2k}, |\{i_1, \dots, i_{2k}\}|=l} \sum_{J_{2k} \in \mathcal{P}_{N,2k}^{I_{2k}}} \int (F_{i_1} \cdot \delta K^{i_1, j_1}) \cdots (F_{i_{2k}} \cdot \delta K^{i_{2k}, j_{2k}}) \bar{f}_N \, dZ \\
&\leq \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{l=1}^k \sum_{I_{2k} \in \mathcal{E}_{N,2k}, |\{i_1, \dots, i_{2k}\}|=l} P_{N,2k} \left(2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} a_1^{a_1} \cdots a_N^{a_N}
\end{aligned} \tag{6.12}$$

where we recall that $P_{N,2k} = 2k e^k 2^{2k} k^k N^k$ which is the bound obtained on $|\mathcal{P}_{N,2k}^{I_{2k}}|$

in Lemma 18.

Observe that for a given l and given multiplicities a_1, \dots, a_l , the number of $I_{2k} \in \mathcal{E}_{N,2k}$ with such multiplicities is bounded by

$$\frac{(2k)!}{(a_1)! \cdots (a_l)!},$$

by Lemma 13.

Thus

$$\begin{aligned} & \sum_{l=1}^k \sum_{I_{2k} \in \mathcal{E}_{N,2k}, |\{i_1, \dots, i_{2k}\}|=l} a_1^{a_1} \cdots a_N^{a_N} \\ &= \sum_{l=1}^k \binom{N}{l} \sum_{a_1 + \dots + a_l = 2k, a_1 \geq 2, \dots, a_l \geq 2} \frac{(2k)!}{(a_1)! \cdots (a_l)!} a_1^{a_1} \cdots a_l^{a_l} \\ &\leq (2k)! e^{2k} \sum_{l=1}^N \binom{N}{l} \binom{2k-l-1}{l-1}, \end{aligned}$$

where the last inequality is ensured by $a_i^{a_i} \leq a_i! e^{a_i}$ and a direct consequence of Lemma (12)

$$\begin{aligned} & |\{(a_1, \dots, a_l) | a_1 + \dots + a_l = 2k, a_1 \geq 2, \dots, a_l \geq 2\}| \\ &= |\{(b_1, \dots, b_l) | b_1 + \dots + b_l = 2k, b_1 \geq 1, \dots, b_l \geq 1\}| = \binom{2k-l-1}{l-1}. \end{aligned}$$

Since $1 \leq l \leq k$, Stirling's formula gives

$$\binom{2k-l-1}{l-1} \leq \binom{2k}{k} \leq \frac{1}{\sqrt{k}} 2^{2k}.$$

Combining all the estimates, one obtains

$$\begin{aligned} & \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N dZ \\ &\leq \frac{(2\|K\|_{L^\infty})^{2k}}{(2k)!} \frac{1}{N^{2k}} \left(\sup_p \frac{M_p}{p} \right)^{2k} (2k e^k 2^{2k} k^k N^k) \left(\frac{1}{\sqrt{k}} (2e)^{2k} (2k)! \right) \sum_{l=1}^k \binom{N}{l} \\ &\leq 2\sqrt{k} \left(8\|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} e^{3k} \frac{k^k}{N^k} k \binom{N}{k}. \end{aligned} \tag{6.13}$$

Now we use Stirling's formula again to simplify the binomial coefficient above,

$$\frac{k^k}{N^k} \binom{N}{k} = \frac{k^k}{N^k} \frac{N!}{(N-k)!k!} \leq \frac{1}{\sqrt{\pi k}} \sqrt{\frac{N}{N-k}} \left(\frac{N}{N-k} \right)^{N-k}.$$

The assumption $3k \leq N$ gives that $\frac{N}{N-k} \leq \frac{3}{2}$ and further implies

$$\frac{k^k}{N^k} \binom{N}{k} \leq \sqrt{\frac{3}{2\pi k}} e^k.$$

Using this bound in (6.13) we complete the proof of Prop 15. \square

6.2.2 The case $3k > N$: Proof of Proposition 16

Now we establish the estimate for large k .

Proof of Proposition 16 We only need the trivial bound for R_N , that is

$$|R_N| \leq 2\|K\|_{L^\infty} \sum_{i=1}^N |\nabla_{v_i} \log f|.$$

As what we did in the previous section, one proceeds as

$$\begin{aligned} \frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ &\leq \frac{(2\|K\|_{L^\infty})^{2k}}{(2k)!} \int \left(\sum_{i=1}^N |\nabla_{v_i} \log f(x_i, v_i)| \right)^{2k} \bar{f}_N \, dZ \\ &= \frac{(2\|K\|_{L^\infty})^{2k}}{(2k)!} \sum_{a_1+\dots+a_N=2k, a_1 \geq 0, \dots, a_N \geq 0} \frac{(2k)!}{(a_1)! \dots (a_N)!} M_{a_1}^{a_1} \dots M_{a_N}^{a_N} \\ &\leq \left(2e\|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k} \left[\sum_{a_1+\dots+a_N=2k, a_1 \geq 0, \dots, a_N \geq 0} 1 \right] \end{aligned} \quad (6.14)$$

where the summation inside the bracket $[\cdot]$ equals to

$$|\{(a_1, \dots, a_N) | a_1 + \dots + a_N = 2k, a_i \geq 0 \text{ for } 1 \leq i \leq N\}| = \binom{2k + N - 1}{N - 1}$$

by applying Lemma 12 with $b_i = a_i + 1$. To simplify this binomial coefficient, we

write $N - 1 = 2ks$, where $s < \frac{3}{2}$, yielding

$$\binom{2k + N - 1}{N - 1} = \frac{(2k(1+s))!}{(2ks)!(2k)!}.$$

Apply Stirling's formula to the factorials above and using the fact that $(1 + \frac{1}{s})^s < e$

for $s > 0$, we get $N \geq 2$ and $3k > N$,

$$\binom{2k + N - 1}{N - 1} \leq \left(\frac{5}{2} \right)^{2k} e^{2k}.$$

Inserting this control back into (6.14), one finally reaches

$$\frac{1}{(2k)!} \int |R_N|^{2k} \bar{f}_N \, dZ \leq \left(5e^2 \|K\|_{L^\infty} \left(\sup_p \frac{M_p}{p} \right) \right)^{2k}.$$

Summation over all $k > \frac{N}{3}$ completes the proof. \square

Chapter 7: Main Estimates: The 1st order case

In this chapter, we prove the main estimates Theorem 6 and Theorem 7, the most technical part in the article [95]. The main idea is essentially the same as the one in the second order case. However, the combinatorics arguments here are more complicated than the 2nd order case. Indeed, in the Newton dynamics or in the limit Vlasov system, the velocity field $(v, K \star \rho(x))$ in the limit is divergence free and in particular the velocity in the x -direction only depends on v while the velocity in the v -direction only depends on x , more directly leading to cancellation rules for instance in the expansion (6.3) induced by integration by parts.

7.1 Main estimate I : Proof of Theorem 6

Quite different to the main estimate in the 2nd order cases, here essentially we only have one index $J_{4k} \in \mathcal{T}_{N,4k}$ rather than the couple (I_{2k}, J_{2k}) . And the cancellation rule is essentially only due to the definition of convolution for instance for a function ϕ and integers $i \neq j$

$$\int_{\mathbb{D}} \{\phi(x_i - x_j) - \phi \star \rho(x_i)\} \rho(x_j) dx_j = 0.$$

Proof of Theorem 6 As we did before, we write

$$\delta\phi^{ij} = \phi(x_i - x_j) - \phi \star \rho(x_i).$$

By symmetry we assume that $i = 1$ in the following and call $R_N = R_{N,1}^{\phi,g}$ in this proof.

By Taylor's expansion and the discussion before (6.2), one has

$$\exp(R_N) = \sum_{k=0}^{\infty} \frac{1}{k!} (R_N)^k \leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} (R_N)^{2k}.$$

Hence it suffices only to bound the series with even terms

$$\int_{\mathbb{D}^N} \bar{\rho}_N \exp(R_N) \, dX \leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} \int \bar{\rho}_N (R_N)^{2k} \, dX, \quad (7.1)$$

where in general the k -th even term can be expanded as

$$\begin{aligned} & \frac{1}{(2k)!} \int \bar{\rho}_N (R_N)^{2k} \, dX \\ &= \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{j_1, \dots, j_{4k}=1}^N \int_{\mathbb{D}^N} \bar{\rho}_N g^{4k}(x_1) \delta\phi^{1j_1} \dots \delta\phi^{1j_{4k}} \, dX. \end{aligned} \quad (7.2)$$

We divide the proof in two different cases: k is small compared to N or k is comparable to or larger than N . In the first case, $4k \leq N$, we apply Lemma 14 to get the combinatorics work done. The second part, $4k > N$, is almost trivial since now the coefficients $\frac{1}{(2k)!}$ dominates.

Case: $4 \leq 4k \leq N$ Recall the definitions of the overall set (see (5.3)) and the multiplicity function in chapter 5. Write

$$J_{4k} = (j_1, \dots, j_{4k}) \in \mathcal{T}_{N,4k}, \quad \Phi_{N,4k}(J_{4k}) = B_N = (b_1, b_2, \dots, b_N).$$

In (7.2), the integral with index $J_{4k} = (j_1, \dots, j_{4k}) \in \mathcal{T}_{N,4k}$ vanishes provided that for some $2 \leq l \leq N$, the multiplicity $b_l = 1$. That means number l is only be taken

once in J_{4k} , say only $j_\nu = l$. Indeed,

$$\begin{aligned}
& \int_{\mathbb{D}^N} \bar{\rho}_N(g(x_1))^{4k} \delta\phi^{1j_1} \dots \delta\phi^{1j_{4k}} dX \\
&= \int_{\mathbb{D}^{N-1}} \frac{\bar{\rho}_N}{\rho(x_l)} (g(x_1))^{4k} \prod_{\nu' \neq \nu} \delta\phi^{1j_{\nu'}} \prod_{l' \neq l} dx_{l'} \\
&\quad \cdot \left(\int_{\mathbb{D}} (\phi(x_1 - x_l) - \phi \star \rho(x_1)) \rho(x_l) dx_l \right) \\
&= 0,
\end{aligned}$$

by the definition of the convolution

$$\phi \star \rho(x_1) = \int_{\mathbb{D}} \phi(x_1 - x_2) \rho(x_2) dx_2.$$

Consequently, the indices $J_{4k} \in \mathcal{T}_{N,4k}$ for the non-vanishing terms in (7.2) only have two types: either 1) $b_l \neq 1$ for all $1 \leq l \leq N$ or 2) $b_1 = 1$ but $b_l \neq 1$ for all $2 \leq l \leq N$. Recall the definition of the “**effective set**” $\mathcal{E}_{q,p}$ before lemma 14. The total number of the first type indices J_{4k} with all $b_l \neq 1$ is just $|\mathcal{E}_{N,4k}|$.

Let us count the other type indices, i.e. those J_{4k} with $b_1 = 1$ but $b_l \neq 1$ for any $2 \leq l \leq N$. In the first step, we choose one position from $4k$ ones for number 1 since $a_1 = 1$. We have $\binom{4k}{1}$ choices in this step. All other components of J_{4k} are chosen from the set $\{2, 3, \dots, N\}$ but no number is only chosen once. Hence, in this step the total choice should be $|\mathcal{E}_{N-1,4k-1}|$. By the multiplication rule, the total number of this type J_{4k} is $4k|\mathcal{E}_{N-1,4k-1}|$.

Consequently, the number of the non-vanishing terms in (7.2) after integral is no larger than

$$|\mathcal{E}_{N,4k}| + 4k|\mathcal{E}_{N-1,4k-1}| \leq (1 + 4k)|\mathcal{E}_{N,4k}| \leq 10k^2 e^{2k} N^{2k} (2k)^{2k}. \quad (7.3)$$

The last inequality above is ensured by Lemma 14.

For each term in the summation of (7.2), one trivially has

$$\int_{\mathbb{D}^N} \bar{\rho}_N (g(x_1))^{4k} \delta\phi^{1j_1} \dots \delta\phi^{1j_{4k}} dX \leq (2\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k}. \quad (7.4)$$

Consequently, combining (7.2), (7.3) and (7.4), for $1 \leq k \leq \lfloor \frac{N}{4} \rfloor$, we obtain

$$\begin{aligned} \frac{1}{(2k)!} \int \bar{\rho}_N (R_N)^{2k} dX &\leq \frac{1}{(2k)!} \frac{1}{N^{2k}} (10k^2 e^{2k} N^{2k} (2k)^{2k}) (2\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k}. \\ &\leq 5k^{\frac{3}{2}} (2e\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k}. \end{aligned} \quad (7.5)$$

The last inequality in (7.5) is ensured by the Stirling's formula for $n = 2k$.

Case: $4k > N$ In this case, we don't need count how many terms in (7.2) will remain after integral. We can simply use the total number $|\mathcal{T}_{N,4k}| = N^{4k}$ in the calculation. Combing (7.2) and (7.4), we have for $k > \lfloor \frac{N}{4} \rfloor$

$$\begin{aligned} \frac{1}{(2k)!} \int \bar{\rho}_N (R_N)^{2k} dX &\leq \frac{1}{(2k)!} \frac{1}{N^{2k}} N^{4k} (2\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k} \\ &\leq k^{-\frac{1}{2}} (2\sqrt{2}e\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k}. \end{aligned} \quad (7.6)$$

The last inequality in (7.6) is also obtained by the Stirling's formula.

Combing (7.5), (7.6) and (7.1), we establish

$$\begin{aligned} \int_{\mathbb{D}^N} \bar{\rho}_N \exp(R_N) dX &\leq 3 \left(1 + \sum_{k=1}^{\lfloor \frac{N}{4} \rfloor} 5k^{\frac{3}{2}} (2e\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k} \right. \\ &\quad \left. + \sum_{k=\lfloor \frac{N}{4} \rfloor + 1}^{\infty} k^{-\frac{1}{2}} (2\sqrt{2}e\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k} \right). \end{aligned}$$

The proof of (4.5) is completed by

$$\begin{aligned} \sum_{k=1}^{\lfloor \frac{N}{4} \rfloor} 5k^{\frac{3}{2}} (2e\|\phi\|_{L^\infty} \|g\|_{L^\infty})^{4k} &\leq \frac{5}{2} \alpha \sum_{k=1}^{\infty} k(k+1) \alpha^{k-1} \\ &= \frac{5}{2} \alpha \frac{d^2}{d\alpha^2} \left(\sum_{k=0}^{\infty} \alpha^k \right) = \frac{5}{2} \alpha \left(\frac{1}{1-\alpha} \right)'' = \frac{5\alpha}{(1-\alpha)^3} < \infty \end{aligned}$$

and

$$\sum_{k=\lfloor \frac{N}{4} \rfloor + 1}^{\infty} k^{-\frac{1}{2}} \left(2\sqrt{2}e\|\phi\|_{L^\infty} \|g\|_{L^\infty} \right)^{4k} \leq \sum_{k=1}^{\infty} \beta^k = \frac{1}{1-\beta} - 1 = \frac{\beta}{1-\beta} < \infty.$$

□

7.2 Main estimate II: Proof of Theorem 7

The combinatorics here is the most difficult one in this thesis. Let us recall some important notations

$$G^{ij} = \left[\sum_{h,l=1}^d R_{hl}(x_i) \delta V_{hl}^{ij} \right] - \Delta^{ij}, \quad (7.7)$$

where

$$R_{hl}(x) = \frac{1}{\rho(x)} \partial_l \partial_h \rho(x), \quad R(x) = \sum_{h,l=1}^d |R_{hl}(x)|,$$

and

$$\delta V_{hl}^{ij} = V_{hl}(x_i - x_j) - V_{hl} \star \rho(x_i), \quad \Delta^{ij} = (\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i).$$

Using the new notion of G^{ij} , we can write $\Theta_N = \Theta_N = \frac{1}{N} \sum_{i,j=1}^N G^{ij}$.

For any $p \geq 1$, we further denote

$$M_p \equiv \left(\int_{\mathbb{D}} (|R(x)| + 1)^p \rho(x) \, dx \right)^{\frac{1}{p}}. \quad (7.8)$$

Consequently, a simple calculation can show that

$$\sup_{p \geq 1} \frac{M_p}{p} \leq \sup_{p \geq 1} \frac{\|R\|_{L^p(\rho \, dx)}}{p} + 1 < \infty. \quad (7.9)$$

The quantity $\sup_{p \geq 1} \frac{M_p}{p}$ will enter the estimates below.

As in the proof of Theorem 6, since

$$\int_{\mathbb{D}^N} \bar{\rho}_N \exp(\Theta_N) \, dX \leq 3 \sum_{k=0}^{\infty} \frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} \, dX, \quad (7.10)$$

it suffices to show the convergence of the series of even terms

$$\begin{aligned} & 1 + \sum_{k=1}^{\infty} \frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} \, dX \\ &= 1 + \sum_{k=1}^{\infty} \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{i_1, \dots, i_{2k}=1}^N \sum_{j_1, \dots, j_{2k}=1}^N \int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} \, dX. \end{aligned} \quad (7.11)$$

As the previous section, we divide the proof into two cases: the technical case when k is relatively small compared to N and the trivial case when k is comparable to or larger than N .

Accordingly Theorem 7 is a consequence of the following two propositions

Proposition 17 *If $4k > N$, one has*

$$\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} dX \leq \left(6e^2 [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{M_p}{p} \right) \right)^{2k}.$$

Proposition 18 *For $4 \leq 4k \leq N$, one has*

$$\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} dX \leq \left(1600 [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{M_p}{p} \right) \right)^{2k}.$$

Let us give a quick proof of Theorem 7 assuming Proposition 17 and Proposition 18. *Proof of Theorem 7* By (7.10) and Proposition 17 and Proposition 18, one has

$$\begin{aligned} \int_{\mathbb{D}^N} \bar{\rho}_N \exp(\Theta_N) dX &\leq 3 \left(1 + \sum_{k=1}^{\lfloor \frac{N}{4} \rfloor} \left(1600 [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{M_p}{p} \right) \right)^{2k} \right. \\ &\quad \left. + \sum_{k=\lfloor \frac{N}{4} \rfloor + 1}^{\infty} \left(6e^2 [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{M_p}{p} \right) \right)^{2k} \right). \end{aligned}$$

By the definition of γ and (7.9), one obtains

$$\int_{\mathbb{D}^N} \bar{\rho}_N \exp(\Theta_N) dX \leq 3 \sum_{k=0}^{\infty} \gamma^k = \frac{3}{1-\gamma} < \infty.$$

This completes the proof of Theorem 7. \square

We now proceed to establish the above propositions. We start with the easier one.

7.2.1 The case $4k > N$: Proof of Proposition 17

In this case, it is sufficient to apply the trivial bound for G^{ij} without considering any possible cancellation due to the integration against $\bar{\rho}_N$, that is

$$|G^{ij}| \leq 2\|V\|_{L^\infty}|R(x_i)| + 2\|\operatorname{div}_x K\|_{L^\infty} \leq 2[\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] (|R(x_i)| + 1). \quad (7.12)$$

Consequently, for $k > \frac{N}{4}$ the k -th even term in (7.11) can be estimated with

$$\begin{aligned} & \frac{1}{(2k)!} 2^{2k} (\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty})^{2k} \sum_{i_1, \dots, i_{2k}=1}^N \int_{\mathbb{D}^N} \bar{\rho}_N (|R(x_{i_1})| + 1) \cdots (|R(x_{i_{2k}})| + 1) \, dX \\ &= \frac{2^{2k} (\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty})^{2k}}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N \left(\sum_{i=1}^N (|R(x_i)| + 1) \right)^{2k} \, dX \\ &= \frac{2^{2k} (\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty})^{2k}}{(2k)!} \sum_{a_1 + \dots + a_N = 2k, a_1 \geq 0, \dots, a_N \geq 0} \frac{(2k)!}{(a_1)! \cdots (a_N)!} M_{a_1}^{a_1} \cdots M_{a_N}^{a_N}, \end{aligned} \quad (7.13)$$

where we recall that (7.8), *i.e.*

$$M_{a_i}^{a_i} = \int_{\mathbb{D}} (|R(x)| + 1)^{a_i} \rho(x) \, dx$$

and we use the convention that $M_0^0 = 1$ as well as $0! = 0^0 = 1$. By the fact (7.9), one has for any $1 \leq i \leq N$

$$M_{a_i}^{a_i} \leq a_i^{a_i} \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{a_i} \leq e^{a_i} (a_i)! \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{a_i},$$

where the last inequality uses the fact $n^n \leq e^n n!$. Inserting it into (7.13), one obtains

$$\begin{aligned} & \frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} \, dX \\ & \leq \left(2e [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{M_p}{p} \right) \right)^{2k} \left[\sum_{a_1 + \dots + a_N = 2k, a_1 \geq 0, \dots, a_N \geq 0} 1 \right], \end{aligned} \quad (7.14)$$

where the summation inside the bracket $[\cdot]$ is nothing but $\binom{2k+N-1}{N-1}$. Again it is a consequence of Lemma 12 in Chapter 5. See the argument in the proof of Prop. 16.

We set that $N - 1 = 2ks$ where $s < 2$ since $4k > N$, leading to

$$\binom{2k + N - 1}{N - 1} = \frac{(2k(1 + s))!}{(2ks)!(2k)!}$$

By the Stirling's formula, the fact $(1 + \frac{1}{s})^s < e$ for any $s > 0$ and here $0 < s < 2$,

one has the estimate

$$\binom{2k + N - 1}{N - 1} \leq (3e)^{2k}.$$

Inserting it into (7.14), one obtains that for $4k > N$

$$\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} dX \leq \left(6e^2 [\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty}] \left(\sup_{p \geq 1} \frac{M_p}{p} \right) \right)^{2k}. \quad (7.15)$$

This gives the proof of Proposition 17.

Now we proceed to prove the case when $4k \leq N$. It is the most technical part of this article. We need several new combinatorics lemmas.

7.2.2 The case $4 \leq 4k \leq N$: Proof of Proposition 18

In this case, the purely trivial estimate as the previous case for

$$\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k}$$

is far from enough to show the convergence of (7.11). Indeed, even under a strong

assumption $\|R\|_{L^\infty} < \infty$, for fixed k , the following estimate

$$\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N |\Theta_N|^{2k} \leq N^{2k} \frac{2^{2k}}{(2k)!} (\|V\|_{L^\infty} + \|\operatorname{div}_x K\|_{L^\infty})^{2k} (\|R\|_{L^\infty} + 1)^{2k}$$

will blow up when we send N to infinity. Fortunately, most of the terms in the

following expansion

$$\begin{aligned} & \frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N \left| \frac{1}{N} \sum_{i,j=1}^N G^{ij} \right|^{2k} dX \\ &= \frac{1}{(2k)!} \frac{1}{N^{2k}} \sum_{i_1, \dots, i_{2k}=1}^N \sum_{j_1, \dots, j_{2k}=1}^N \int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} dX. \end{aligned} \quad (7.16)$$

will *vanish* after the integration against $\bar{\rho}_N$.

Now we need to count the pairs (I_{2k}, J_{2k}) again. Recall the definitions of $\mathcal{T}_{N,2k}$ in (5.3) and of the multiplicity function $\Phi_{N,2k}$. As a convention, we denote that $\Phi_{N,2k}(I_{2k}) = A_N = (a_1, a_2, \dots, a_N)$ and $\Phi_{N,2k}(J_{2k}) = (b_1, \dots, b_N)$. Finally we denote

$$m_I = |\{l \mid a_l = 1\}|, \quad n_I = |\{l \mid a_l > 1\}|,$$

s.t. $m_I + n_I$ is exactly the number of integers present in I_{2k} : $m_I + n_I = |\{l \mid a_l \geq 1\}|$.

Observe that for a particular choice of I_{2k} and J_{2k}

$$\begin{aligned} & \int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} \, dX \\ & \leq \int_{\mathbb{D}^N} \bar{\rho}_N \prod_{\nu=1}^{2k} 2(\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty}) (|R(x_{i_\nu})| + 1) \, dX \\ & \leq 2^{2k} (\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty})^{2k} \int_{\mathbb{D}^N} \bar{\rho}_N (R(x_1) + 1)^{a_1} \dots (R(x_N) + 1)^{a_N} \, dX. \end{aligned} \tag{7.17}$$

As one readily sees this bound only depends on the multiplicity in I_{2k} and therefore the main difficulty is to identify and count for which I_{2k} and J_{2k} the above integral does not vanish.

We start by the following lemma which, for every I_{2k} , identifies the only possible J_{2k} s.t. the integral does not vanish.

First we simplify the possible expression of I_{2k} which makes the counting easier by using the natural symmetry by permutation of the problem. For any $\tau \in \mathcal{S}_N$, we simply define $\tau(I_{2k}) = (\tau(i_1), \dots, \tau(i_{2k}))$. Thus τ is a one-to-one application on the I_{2k} and moreover

$$\int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} \, dX = \int_{\mathbb{D}^N} \bar{\rho}_N G^{\tau(i_1), \tau(j_1)} \dots G^{\tau(i_{2k}), \tau(j_{2k})} \, dX.$$

Therefore we only need to consider one I_{2k} in each of the equivalence classes $\{\tau(I_{2k}), \forall \tau \in \mathcal{S}_N\}$, leading to

Definition 4 *A multi-index I_{2k} belongs to the reduced form set $\mathcal{R}_{N,2k}$ iff $0 < a_1 \leq a_2 \dots \leq a_n$ and $a_{n+1} = \dots = a_N = 0$.*

Note that for any I_{2k} there exists only one $\tilde{I}_{2k} \in \mathcal{R}_{N,2k}$ that belongs to the same class, even though there can be several τ s.t. $\tau(I_{2k}) = \tilde{I}_{2k}$.

By the definition of m_I and n_I , if $I_{2k} \in \mathcal{R}_{N,2k}$, one hence has $a_l = 1$ for $l = 1, \dots, m_I$, $a_l > 1$ for $l = m_I + 1, \dots, m_I + n_I$ and $a_l = 0$ for $l > m_I + n_I$.

Lemma 19 (Cancellation Rules) *For any m, n , define as $\mathcal{J}_{m,n}$ the set of indices J_{2k} with multiplicity (b_1, \dots, b_N) satisfying*

- $b_l \geq 1$ for any $l = 1 \dots m$;
- $b_l \neq 1$ for any $l > m + n$.

Then for any $I_{2k} \in \mathcal{R}_{N,2k}$ and any $J_{2k} \notin \mathcal{J}_{m_I, n_I}$, one has that

$$\int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} dX = 0.$$

This lemma implies that we only need to count for each $I_{2k} \in \mathcal{R}_{N,2k}$, the indices $J_{2k} \in \mathcal{J}_{m_I, n_I}$ as the others lead to vanishing integrals. Lemma 19 is not an equivalence: There are still indices $J_{2k} \in \mathcal{J}_{m_I, n_I}$ giving a vanishing integral. But the formulation above allows for simpler combinatorics and in particular \mathcal{J}_{m_I, n_I} only depends in a basic manner on I_{2k} through the two integers m_I and n_I .

Proof of Lemma 19 Choose any $I_{2k} \in \mathcal{R}_{N,2k}$, without loss of generality, assume that I_{2k} has the following form

$$I_{2k} = \left(1, 2, \dots, m_I, \underbrace{m_I + 1, \dots, m_I + 1}_{a_{m_I+1}}, \dots, \underbrace{m_I + n_I, \dots, m_I + n_I}_{a_{m_I+n_I}} \right).$$

Choose any $J_{2k} \notin \mathcal{J}_{m_I, n_I}$. That means that there exists $l \leq m_I$ s.t. $b_l = 0$ or that there exists $l > m_I + n_I$ s.t. $b_l = 1$. Each case corresponds to a different cancellation in the integral.

The case $b_l = 0$ for some $l \leq m_I$. By the definition of the reduced form, $a_l = 1$ and therefore the index l appears only once in I_{2k} and never in J_{2k} thus being present exactly once in the product inside the integral. Assume that $i_\nu = l$ for some ν so

$$\begin{aligned} & \int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} dX \\ &= \int_{\mathbb{D}^{N-1}} \frac{\bar{\rho}_N}{\rho(x_{i_\nu})} \prod_{\nu' \neq \nu} G^{i_{\nu'}, j_{\nu'}} \left(\int_{\mathbb{D}} \rho(x_{i_\nu}) G^{i_\nu, j_\nu} dx_{i_\nu} \right) \prod_{\nu' \neq \nu} dx_{i_{\nu'}}. \end{aligned}$$

Now it is enough to remark that for any i and j

$$\int_{\mathbb{D}} \rho(x_i) G^{ij} dx_i = 0, \quad (7.18)$$

as

$$\begin{aligned} \int_{\mathbb{D}} \rho(x_i) G^{ij} dx_i &= \int_{\mathbb{D}} \rho(x_i) \sum_{h,l=1}^d R_{hl}(x_i) \delta V_{hl}^{ij} dx_i - \int_{\mathbb{D}} \rho(x_i) \Delta^{ij} dx_i \\ &= \sum_{h,l=1}^d \int_{\mathbb{D}} \partial_{x_i^l} \partial_{x_i^h} \rho(x_i) [V_{hl}(x_i - x_j) - V_{hl} \star \rho(x_i)] dx_i - \int_{\mathbb{D}} \rho(x_i) \Delta^{ij} dx_i \\ &= \int_{\mathbb{D}} \rho(x_i) [(\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i)] dx_i - \int_{\mathbb{D}} \rho(x_i) \Delta^{ij} dx_i \\ &= 0, \end{aligned}$$

where we do integration by parts twice from the 2nd line to the 3rd line and recall that

$$\sum_{h,l=1}^d \partial_{x^h} \partial_{x^l} V_{hl} = \operatorname{div}_x K$$

and

$$\Delta^{ij} = (\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i).$$

The case $b_l = 1$ for some $l > m_I + n_I$. Again by definition, this means that $a_l = 0$.

The index l appears only once in J_{2k} and never in I_{2k} . Again it is present exactly once in the product inside the integral. Assume that $j_\nu = l$ for some ν so

$$\begin{aligned} & \int_{\mathbb{D}^N} \bar{\rho}_N G^{i_1, j_1} \dots G^{i_{2k}, j_{2k}} \, dX \\ &= \int_{\mathbb{D}^{N-1}} \frac{\bar{\rho}_N}{\rho(x_{j_\nu})} \prod_{\nu' \neq \nu} G^{i_{\nu'}, j_{\nu'}} \left(\int_{\mathbb{D}} \rho(x_{j_\nu}) G^{i_\nu, j_\nu} \, dx_{j_\nu} \right) \prod_{\nu' \neq \nu} dx_{j_{\nu'}}. \end{aligned}$$

The results then follows from the fact that for $i \neq j$

$$\int_{\mathbb{D}} \rho(x_j) G^{ij} \, dx_j = 0. \quad (7.19)$$

Indeed,

$$\begin{aligned} & \int_{\mathbb{D}} \rho(x_j) R_{hl}(x_i) (V_{hl}(x_i - x_j) - V_{hl} \star \rho(x_i)) \, dx_j \\ &= R_{hl}(x_i) \left[\int_{\mathbb{D}} \rho(x_j) V_{hl}(x_i - x_j) \, dx_j - V_{hl} \star \rho(x_i) \right] = 0, \end{aligned}$$

by the definition of the convolution

$$V_{hl} \star \rho(x_i) = \int_{\mathbb{D}} V_{hl}(x_i - x_j) \rho(x_j) \, dx_j,$$

while similarly

$$\int_{\mathbb{D}} \rho(x_j) ((\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i)) \, dx_j = 0.$$

And from the definition of G^{ij} in (7.7),

$$\begin{aligned} & \int_{\mathbb{D}} \rho(x_j) G^{ij} \, dx_j = \sum_{h,l=1}^d \int_{\mathbb{D}} \rho(x_j) R_{hl}(x_i) (V_{hl}(x_i - x_j) - V_{hl} \star \rho(x_i)) \, dx_j \\ &+ \int_{\mathbb{D}} \rho(x_j) ((\operatorname{div}_x K)(x_i - x_j) - (\operatorname{div}_x K) \star \rho(x_i)) \, dx_j = 0. \end{aligned}$$

□ From the cancellations obtained in Lemma 19, it is enough to bound $|\mathcal{J}_{m,n}|$ and for each m and n the cardinal of $\{I_{2k} \mid m_I = m, n_I = n\}$. Indeed by (7.16) and (7.17), one has

$$\begin{aligned} \int_{\mathbb{D}^N} \bar{\rho}_N \left| \frac{1}{N} \sum_{i,j=1}^N G^{ij} \right|^{2k} dX &\leq 2^{2k} [\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty}]^{2k} \frac{1}{N^{2k}} \\ &\cdot \left(\sum_{a_1+\dots+a_N=2k, a_1 \geq 0, \dots, a_N \geq 0} |\mathcal{J}_{m_a, n_a}| |\{I_{2k} \mid \Phi_{N,2k}(I_{2k}) = (a_1, \dots, a_N)\}| \right) \\ &\cdot \int_{\mathbb{D}^N} \bar{\rho}_N (R(x_1) + 1)^{a_1} \dots (R(x_N) + 1)^{a_N} dX, \end{aligned}$$

where we denote $m_a = m_{(a_1, \dots, a_N)} = |\{l \mid a_l = 1\}|$, $n_a = n_{(a_1, \dots, a_N)} = |\{l \mid a_l > 1\}|$.

Recall that $M_p^p = \int (|R(x)| + 1)^p \rho(x) dx$ and thus

$$\int_{\mathbb{D}^N} \bar{\rho}_N (R(x_1) + 1)^{a_1} \dots (R(x_N) + 1)^{a_N} dX \leq e^{2k} \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{2k} a_1! \dots a_N!.$$

On the other hand by Lemma 13

$$|\{I_{2k} \mid \Phi_{N,2k}(I_{2k}) = (a_1, \dots, a_N)\}| \leq \frac{(2k)!}{a_1! \dots a_N!},$$

which implies that

$$\begin{aligned} &\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N \left| \frac{1}{N} \sum_{i,j=1}^N G^{ij} \right|^{2k} dX \\ &\leq \frac{(2e)^{2k} (\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty})^{2k}}{N^{2k}} \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{2k} \sum_{a_1+\dots+a_N=2k, a_1 \geq 0, \dots, a_N \geq 0} |\mathcal{J}_{m_a, n_a}|. \end{aligned}$$

The missing estimate is given by

Lemma 20 *One has that for some universal constant C*

$$|\mathcal{J}_{m,n}| \leq C^k N^{k-m/2} k^{k+m/2},$$

where C can be chosen as $512e$ or roughly 1400 .

Assuming the above lemma true for the time being, we may now conclude the proof of the Proposition 18 as

$$\begin{aligned} &\frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N \left| \frac{1}{N} \sum_{i,j=1}^N G^{ij} \right|^{2k} dX \\ &\leq \frac{(2e)^{2k} (\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty})^{2k}}{N^{2k}} \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{2k} \sum_{a_1+\dots+a_N=2k, a_1 \geq 0, \dots, a_N \geq 0} C^k N^{k-m_a/2} k^{k+m_a/2}. \end{aligned}$$

Consider any (a_1, \dots, a_N) with exactly p coefficients $a_l \geq 1$. Up to $\binom{N}{p}$ permutations, we can actually assume that $a_1, \dots, a_p \geq 1$. All the other a_l are 0. Since we have $m_a + n_a = p$ and $m_a + 2n_a \leq 2k$ then $m_a \geq 2(p-k)$. As $N \geq k$ then

$$N^{k-m_a/2} k^{k+m_a/2} \leq N^{k-(p-k)_+} k^{k+(p-k)_+}.$$

Hence

$$\begin{aligned} \sum_{a_1+\dots+a_N=2k} N^{k-m_a/2} k^{k+m_a/2} &= \sum_{p=1}^{2k} \binom{N}{p} \sum_{a_1, \dots, a_p \geq 1, a_1+\dots+a_p=2k} N^{k-(p-k)_+} k^{k+(p-k)_+} \\ &\leq \sum_{p=1}^{2k} \binom{N}{p} \binom{2k-1}{p-1} N^{k-(p-k)_+} k^{k+(p-k)_+}, \end{aligned}$$

by Lemma 12. Simply bound $\binom{2k-1}{p-1} \leq 2^{2k}$ and now for $p \leq k$ since obviously $\binom{N}{p}$ is maximum for $p = k$

$$\sum_{p=1}^k \binom{N}{p} \binom{2k-1}{p-1} N^{k-(p-k)_+} k^{k+(p-k)_+} \leq 2^{2k} k \binom{N}{k} N^k k^k \leq (8e)^k N^{2k},$$

by Lemma 11. Similarly for $p > k$,

$$\binom{N}{p} N^{k-(p-k)_+} k^{k+(p-k)_+} \leq e^p N^p p^{-p} N^{2k-p} k^p = e^p N^{2k}.$$

Hence again

$$\sum_{p=k+1}^{2k} \binom{N}{p} \binom{2k-1}{p-1} N^{k-(p-k)_+} k^{k+(p-k)_+} \leq k 2^{2k} e^{2k} N^{2k} < \frac{1}{2} (8e^2)^k N^{2k}.$$

Finally,

$$\begin{aligned} \frac{1}{(2k)!} \int_{\mathbb{D}^N} \bar{\rho}_N \left| \frac{1}{N} \sum_{i,j=1}^N G^{ij} \right|^{2k} dX &\leq (32e^4 C)^k \left(\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty} \right) \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{2k} \\ &\leq (1600)^{2k} \left(\|V\|_{L^\infty} + \|\operatorname{div} K\|_{L^\infty} \right) \left(\sup_{p \geq 1} \frac{M_p}{p} \right)^{2k}. \end{aligned}$$

concluding the proof of Proposition 18.

Now we give the proof of the above lemma. *Proof of Lemma 20* One simply has to impose that $b_l \geq 1$ for $l \leq m$ and $b_l = 0, 2, 3, \dots$ for $l > m+n$. Let us

distinguish further between those $l > m + n$ where $b_l = 0$ and those for which $b_l \geq 2$.

Choose first $p = 0, 1, \dots, \lfloor \frac{2k-m}{2} \rfloor$ and choose then p indices l_1, \dots, l_p between $m + n + 1$ and N which will correspond to $b_l \geq 2$. There are $\binom{N-m-n}{p}$ such possibilities.

Once these l_1, \dots, l_p have been chosen, the set of possible multiplicities for $J_{2k} \in \mathcal{J}_{m,n}$ is given by

$$\mathcal{B}_{m,n,p,l_1,\dots,l_p} = \left\{ (b_1, \dots, b_N) \mid b_1, \dots, b_m \geq 1, b_{l_1}, \dots, b_{l_p} \geq 2, \right. \\ \left. b_l = 0 \text{ if } l > m + n \text{ and } l \neq l_1, \dots, l_p, \text{ and } b_1 + b_2 + \dots + b_N = 2k \right\}.$$

Applying the invariance by permutation, one may assume that $l_1 = m + n + 1$, $l_2 = m + n + 2 \dots$. Denoting the partial sums $s_m = b_1 + \dots + b_m$ and $s_n = b_{m+n+1} + \dots + b_{m+n+p}$, one has

$$|\mathcal{J}_{m,n}| = \sum_{p=0}^{k-m/2} \binom{N-m-n}{p} \sum_{s_m=m}^{2k-2p} \sum_{b_1 \dots b_m \geq 1, b_1 + \dots + b_m = s_m} \\ \sum_{s_n=2p}^{2k-s_m} \sum_{b_{m+n+1}, \dots, b_{m+n+p} \geq 2, b_{m+n+1} + \dots + b_{m+n+p} = s_n} \\ \sum_{b_{m+1}, \dots, b_n \geq 0, b_{m+1} + \dots + b_{m+n} = 2k - s_m - s_n} \frac{(2k)!}{b_1! \dots b_{m+n+p}!}.$$

Using the standard multinomial summation (5.2), one can easily calculate the last sum to obtain

$$|\mathcal{J}_{m,n}| = \sum_{p=0}^{k-m/2} \binom{N-m-n}{p} \\ \sum_{s_m=m}^{2k-2p} \sum_{b_1 \dots b_m \geq 1, b_1 + \dots + b_m = s_m} \sum_{s_n=2p}^{2k-s_m} \frac{n^{2k-s_m-s_n}}{(2k-s_m-s_n)!} \\ \sum_{b_{m+n+1}, \dots, b_{m+n+p} \geq 2, b_{m+n+1} + \dots + b_{m+n+p} = s_n} \frac{(2k)!}{b_1! \dots b_m! b_{m+n+1}! \dots b_{m+n+p}!}.$$

Now bound the sum on $b_1 \dots b_m$ by the sum starting at $b_1, \dots, b_m = 0$ and similarly

for the sum on $b_{m+n+1} \dots b_{m+n+p}$ to obtain

$$|\mathcal{J}_{m,n}| \leq \sum_{p=0}^{k-m/2} \binom{N-m-n}{p} \sum_{s_m=m}^{2k-2p} \sum_{s_n=2p}^{2k-s_m} \frac{(2k)! n^{2k-s_m-s_n} m^{s_m} p^{s_n}}{(2k-s_m-s_n)! s_m! s_n!}.$$

We recall the obvious bound $\binom{a}{b} \leq 2^a$ so that

$$\begin{aligned} \frac{(2k)!}{(2k-s_m-s_n)! s_m! s_n!} &= \frac{(2k-s_m)!}{(2k-s_m-s_n)! s_n!} \frac{(2k)!}{(2k-s_m)! s_m!} \\ &= \binom{2k}{s_m} \binom{2k-s_m}{s_n} \leq 2^{4k}. \end{aligned}$$

Furthermore by the Stirling's formula, more precisely by lemma 11 as $m+n \leq N/2$,

$\binom{N-m-n}{p} \leq e^p N^p p^{-p}$. Thus

$$|\mathcal{J}_{m,n}| \leq 2^{4k} \sum_{p=0}^{k-m/2} e^p N^p \sum_{s_m=m}^{2k-2p} \sum_{s_n=2p}^{2k-s_m} n^{2k-s_m-s_n} p^{s_n-p} m^{s_m}.$$

Note that $2k - s_m - s_n \geq 0$ and $s_n - p \geq 0$ and $m, n, p \leq 2k$ so

$$n^{2k-s_m-s_n} p^{s_n-p} m^{s_m} \leq (2k)^{2k-p}.$$

Therefore finally

$$\begin{aligned} |\mathcal{J}_{m,n}| &\leq 2^{6k} e^k (2k)^2 \sum_{p=0}^{k-m/2} N^p k^{2k-p} \\ &\leq 2^{6k} e^k (2k)^2 k N^{k-m/2} k^{k+m/2} < (2^9 e)^k N^{k-m/2} k^{k+m/2}. \end{aligned}$$

as since $N \geq k$, the maximum of $N^p k^{2k-p}$ is attained for the maximal value of p . \square

Appendix A: Preliminary about entropy and relative entropy

A.1 Definitions

Consider a Polish (complete separable metric) space E . For instance in this thesis we set it as $\mathbb{D} \times \mathbb{R}^d$ in the 2nd order case (1.1) or \mathbb{D} in the 1st order case (1.2). There are two important quantities that we use to quantify (Kac's) chaos: the Boltzmann entropy and the Fisher information. Denote by $f_N, g_N \in \mathcal{P}(E^N)$. And recall we denote by $Z = (z_1, \dots, z_N) \in E^N$ with $z \in E$ in general. The (scaled) entropy is defined as

$$H_N(f_N) = \frac{1}{N} \int_{E^N} f_N \log f_N \, dz_1 \cdots dz_N. \quad (\text{A.1})$$

The Fisher information is

$$I_N(f_N) = \frac{1}{N} \int_{E^N} \frac{|\nabla f_N|^2}{f_N} \, dz_1 \cdots dz_N. \quad (\text{A.2})$$

The relative entropy of f_N w.r.t. g_N is defined as

$$H_N(f_N|g_N) = \frac{1}{N} \int_{E^N} f_N \log \frac{f_N}{g_N} \, dz_1 \cdots dz_N. \quad (\text{A.3})$$

Upon normalization with the factor $1/N$, both the entropy and the relative entropy have the very famous tensorized property

Lemma 21 (Tensorization properties) *If f_N is tensorized or (Boltzmann-)chaotic,*

i.e. there exists a probability measure f on E s.t. $f_N = f^{\otimes N}$, then

$$\begin{cases} H_N(f_N) = H_1(f) = H(f) = \int_E f \log f \, dz, \\ I_N(f_N) = I_1(f) = I(f) = \int_E \frac{|\nabla f|^2}{f} \, dz. \end{cases}$$

Similarly, if in addition g_N is tensorized, with $g_N = g^{\otimes N}$, then

$$H_N(f_N|g_N) = H_1(f|g) = H(f|g) = \int_E f \log \frac{f}{g} \, dz.$$

Simply checking the definitions will give the proof.

A.2 Monotonicity of the relative entropy

Recall that the k -marginal of f_N is defined as

$$f_{N,k}(z_1, \dots, z_k) = \int_{E^{N-k}} f_N(Z) \, dz_{k+1} \cdots dz_N.$$

One has the following key observation as per

Proposition 19 (General form of Prop. 1) *Assume that $f_N \in \mathcal{P}(E^N)$ (not necessarily symmetric) and $f_{N-1} \in \mathcal{P}(E^{N-1})$ with the assumption*

$$\int_E f_N(z_1, \dots, z_N) \, dz_i = f_{N-1}(z_1, \dots, z_{i-1}, \hat{z}_i, \dots, z_N), \quad i = 1, \dots, N,$$

where \hat{z}_i means the variable z_i is taken away at that position. Then one has

$$H_{N-1}(f_{N-1}) \leq H_N(f_N).$$

Consequently, provided that $f_N \in \mathcal{P}_{Sym}(E^N)$, $f_{N,k}$ the k -marginal of f_N and that $f \in \mathcal{P}(E)$, one has

$$H_k(f_{N,k}) \leq H_N(f_N), \quad H_k(f_{N,k}|f^{\otimes k}) \leq H_N(f_N|f^{\otimes N}). \quad (\text{A.4})$$

The proof of the above proposition relies on a consequence or a variant of the very famous General Hölder inequality (See Appendix B g in [52]) as per

Lemma 22 (General Hölder inequality) *For N functions f_i defined on E^{N-1} ($N \geq 2$), one defines a function $f : E^N \rightarrow \mathbb{R}$ as*

$$f(z_1, \dots, z_N) = \prod_{i=1}^N f_i(z_1, \dots, z_{i-1}, \hat{z}_i, z_{i+1}, \dots, z_N),$$

where \hat{z}_i means we omit the variable z_i at that position. Then one has

$$\|f\|_{L^1} \leq \prod_{i=1}^N \|f_i\|_{L^{N-1}}. \quad (\text{A.5})$$

Proof. The case with $N = 2$ can be easily show by Fubini's theorem. For $N \geq 3$, it can be proved by induction and the usual Hölder inequality. \square

Proof of Proposition 19. By the symmetry,

$$\begin{aligned} H_{N-1}(f_{N-1}) &= \frac{1}{N-1} \frac{1}{N} \sum_{i=1}^N \int_{E^N} f_N \log f_{N-1}(z_1, \dots, z_{i-1}, \hat{z}_i, z_{i+1}, \dots, z_N) dZ \\ &= \frac{1}{N} \int_{E^N} f_N \log G_N(Z) dX, \end{aligned}$$

where

$$G_N(Z) = \prod_{i=1}^N (f_{N-1}(\hat{z}_i))^{\frac{1}{N-1}}.$$

Applying Lemma 22 , one has

$$\|G_N\|_{L^1} \leq \prod_{i=1}^N \| (f_{N-1}(\hat{z}_i))^{\frac{1}{N-1}} \|_{L^{N-1}} = \|f_{N-1}\|_{L^1}^{\frac{N}{N-1}} \leq 1.$$

Consequently,

$$\begin{aligned} H_N(f_N) - H_{N-1}(f_{N-1}) &= \frac{1}{N} \int_{E^N} f_N \log \frac{f_N}{G_N} dZ \\ &= \frac{1}{N} \int G_N \left(\frac{f_N}{G_N} \log \frac{f_N}{G_N} + 1 - \frac{f_N}{G_N} \right) dZ - \frac{1}{N} \int G_N dZ + \frac{1}{N} \int f_N dZ \geq 0, \end{aligned}$$

where the first integral in the second line is non-negative since the function $h(x) = x \log x + 1 - x \geq 0$ for any $x \geq 0$. The second part of this theorem is a direct consequence of the first part. \square

A.3 Csiszár-Kullback-Pinsker(CKP) inequality

The classical Csiszár-Kullback-Pinsker(CKP) inequality can be illustrated by the following elementary calculation.

Lemma 23 *Assume that $\rho, \bar{\rho} \in \mathcal{P}(\mathbb{T}^d) \cap L^1(\mathbb{T}^d) \cap L^\infty(\mathbb{T}^d)$. Then one has a baby version of CKP inequality*

$$\|\rho - \bar{\rho}\|_{L^1} \leq \|\rho - \bar{\rho}\|_{L^2} \leq \sqrt{2(\|\rho\|_{L^\infty} + \|\bar{\rho}\|_{L^\infty})} \sqrt{H(\rho|\bar{\rho})}.$$

Proof Let $g(x) = x \log x$ for $x \geq 0$ with the convention that $g(0) = 0$. Then Taylor's expansion near $\bar{\rho}$ gives

$$g(\rho) - g(\bar{\rho}) = (1 + \log \bar{\rho})(\rho - \bar{\rho}) + \frac{1}{2}g''(\xi)(\rho - \bar{\rho})^2,$$

with ξ chosen between ρ and $\bar{\rho}$. Taking the integral of both sides leads to

$$H(\rho|\bar{\rho}) = \int_{\mathbb{T}^d} \rho(z) \log \frac{\rho(z)}{\bar{\rho}(z)} dz = \frac{1}{2} \int_{\mathbb{T}^d} \frac{1}{\xi(z)} (\rho(z) - \bar{\rho}(z))^2 dz.$$

Then

$$\|\rho - \bar{\rho}\|_{L^2}^2 \leq 2 \left(\sup_z |\xi(z)| \right) H(\rho|\bar{\rho}) \leq 2(\|\rho\|_{L^\infty} + \|\bar{\rho}\|_{L^\infty}) H(\rho|\bar{\rho}).$$

The control of the L^1 norm by the L^2 norm is ensured by the Cauchy-Schwarz inequality. \square

A.4 Lower bounds for the entropy

For $E = \mathbb{T}^d$ where $\int_E 1 dz = 1$, the entropy is always non-negative, *i.e.*

$$\int_{E^N} f_N \log f_N dZ = \int (f_N \log f_N + 1 - f_N) dZ \geq 0,$$

thanks to the function $h(x) = x \log x + 1 - x \geq 0$ for any $x > 0$.

If we work on the unbounded space for instance $E = \mathbb{D} \times \mathbb{R}^d$ in the 2nd order case, a uniform bound w.r.t. N for the moments of f_N up to time T is required, for instance

$$\sup_{N \geq 2} \sup_{0 \leq t \leq T} \frac{1}{N} \int \sum_{i=1}^N |z_i|^2 f_N dZ < \infty.$$

We assume a finite moment for initial distribution f_N^0 and then we can propagate the bound such that it is still uniform in N up to time T .

Now we can give a lower bound for $H_N(f_N)$ as

$$\begin{aligned} H_N(f_N) &= \frac{1}{N} \int f_N \log f_N dZ \\ &= \frac{1}{N} \int G_N \left(\frac{f_N}{G_N} \log \frac{f_N}{G_N} - \frac{f_N}{G_N} + 1 \right) + \frac{1}{N} \int f_N \log G_N \\ &\geq \frac{1}{N} \int f_N \log G_N \end{aligned}$$

where for instance if $E = \mathbb{R}^d$ then $G_N \in \mathcal{P}(E)$ is usually chosen as the Gaussian

$$G_N(X) \equiv \frac{1}{\pi^{\frac{dN}{2}}} \exp\left(-\sum_{i=1}^N |z_i|^2\right),$$

this explains why the 2nd moment bound is important.

Appendix B:

B.1 Weak-strong uniqueness on Eq. (1.3) and the proof of Theorem

4

Assume that f and \tilde{f} solve Vlasov equation (1.3) in weak sense. Assume that f satisfies (1.10). By density we may assume that f is smooth, C^1 , and decays at infinity without ever vanishing; just consider any such sequence f_n satisfying uniformly the bound (1.10) and pass to the limit $f_n \rightarrow f$ at the end of the argument.

Consider for any $t \in [0, T]$ and decompose

$$\begin{aligned} H(t) &= \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} \log\left(\frac{\tilde{f}}{f}\right) dx dv = \int \tilde{f} \log \tilde{f} - \int \tilde{f} \log f \\ &\leq \int \tilde{f}^0 \log \tilde{f}^0 - \sigma \int_0^t \int \frac{|\nabla_v \tilde{f}|^2}{\tilde{f}} - \int \tilde{f} \log f, \end{aligned}$$

with $\tilde{f}^0 = \tilde{f}(t=0)$ and per the assumption of dissipation of entropy for \tilde{f} in Theorem 4.

By our assumption f is smooth and $\log f$ can hence be used as a test function. Thus since \tilde{f} is a solution to the Vlasov equation (1.3) in the sense of distribution, one has that

$$\begin{aligned} \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} \log f &= \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f}^0 \log f^0 \\ &+ \int_0^t \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f}(s, x, v) (\partial_t \log f + v \cdot \nabla_x \log f + K \star \tilde{\rho} \cdot \nabla_v \log f + \sigma \Delta_v \log f). \end{aligned}$$

Since f is a strong solution to the Vlasov equation, this leads to

$$\begin{aligned} \int \tilde{f} \log f &= \int \tilde{f}^0 \log f^0 + \int_0^t \int \tilde{f}(s, x, v) R \, dx \, dv \, ds \\ &+ \sigma \int_0^t \int \tilde{f}(s, x, v) \left(\frac{\Delta_v f}{f} + \Delta_v \log f \right) \, dx \, dv \, ds, \end{aligned}$$

where we define

$$R := \nabla_v \log f(x, v) \cdot \{K \star \tilde{\rho}(x) - K \star \rho(x)\}.$$

Observe now that, with usual entropy estimates

$$\begin{aligned} & - \int \tilde{f}(s, x, v) \left(\frac{\Delta_v f}{f} + \Delta_v \log f \right) \, dx \, dv - \int \frac{|\nabla_v \tilde{f}|^2}{\tilde{f}} \, dx \, dv \\ &= \int \left(-\tilde{f} \frac{|\nabla_v f|^2}{f^2} + 2 \frac{\nabla_v \tilde{f} \cdot \nabla_v f}{f} - \frac{|\nabla_v \tilde{f}|^2}{\tilde{f}} \right) \, dx \, dv \\ &= - \int \tilde{f} |\nabla_v \log \frac{\tilde{f}}{f}|^2 \, dx \, dv \leq 0. \end{aligned}$$

Therefore

$$H(t) \leq H(0) - \int_0^t \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} R \, dx \, dv \, ds. \quad (\text{B.1})$$

Note that by the definition of R

$$\int_{\mathbb{D} \times \mathbb{R}^d} f R \, dx \, dv = \int \nabla_v f (K \star \tilde{\rho} - K \star \rho) \, dx \, dv = 0,$$

as $K \star \rho$ and $K \star \tilde{\rho}$ do not depend on v . Hence

$$\int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} R \, dx \, dv = \int_{\mathbb{D} \times \mathbb{R}^d} (\tilde{f} - f) R \, dx \, dv.$$

Simply bound

$$\left| \int_{\mathbb{D} \times \mathbb{R}^d} \tilde{f} R \, dx \, dv \right| \leq \|K \star (\tilde{\rho} - \rho)\|_{L^\infty} \int_{\mathbb{D} \times \mathbb{R}^d} |\nabla_v \log f| |\tilde{f} - f| \, dx \, dv.$$

Observe that

$$\|K \star (\tilde{\rho} - \rho)\|_{L^\infty} \leq \|K\|_{L^\infty} \|\tilde{\rho} - \rho\|_{L^1} \leq \|K\|_{L^\infty} \|\tilde{f} - f\|_{L^1},$$

so that

$$H(t) \leq H(0) + \|K\|_{L^\infty} \int_0^t \|\tilde{f} - f\|_{L^1} \left[\int_{\mathbb{D} \times \mathbb{R}^d} |\nabla_v \log f| |\tilde{f} - f| dx dv \right] ds.$$

Use the weighted Csiszár-Kullback-Pinsker inequality in Theorem 1 in [21] with $\varphi(x, v) = |\nabla_v \log f|$ to obtain

$$\int |\nabla_v \log f| |\tilde{f} - f| dx dv \leq \frac{2}{\lambda_f} \left(\frac{3}{2} + \log \int e^{\lambda_f |\nabla_v \log f|} f dx dv \right) \left(\sqrt{H} + \frac{1}{2}H \right).$$

Recall the notation

$$\theta_f = \sup_{t \in [0, T]} \int e^{\lambda_f |\nabla_v \log f|} f dx dv < \infty,$$

by the assumption (1.10). This leads to

$$H(t) \leq H(0) + C (1 + \log \theta_f) \|K\|_{L^\infty} \int_0^t \|f - \tilde{f}\|_{L^1} \left(\sqrt{H} + \frac{H}{2} \right) ds.$$

Simply use now the classical Csiszár-Kullback-Pinsker inequality (see [138]) to find

$$H(t) \leq H(0) + C (1 + \log \theta_f) \|K\|_{L^\infty} \int_0^t \left(H + \frac{H^{3/2}}{2} \right) ds. \quad (\text{B.2})$$

As long as $H(t) \leq 1$, then $H^{\frac{3}{2}} \leq H$. Eq. (B.2) gives a Gronwall's inequality which proves Theorem 4.

B.2 Proof of Proposition 10

We first denote the linear operator for a fixed $\rho(t, x)$ as

$$L = v \cdot \nabla_x f + K \star \rho \cdot \nabla_v.$$

To show the existence of a smooth solution over a short time, it is sufficient to propagates some norms of $|\nabla f|$.

Step I: Propagate $\|\nabla f\|_{L^1}$ and $\|\nabla f\|_{L^\infty}$. It is easy to check that

$$\{\partial_t(\nabla_x f) + L(\nabla_x f) = \sigma \Delta_v(\nabla_x f) - (K \star \nabla_x \rho) \cdot \nabla_v f, \partial_t(\nabla_v f) + L(\nabla_v f) = \sigma \Delta_v(\nabla_v f) - \nabla_x f.\} \quad (\text{B.3})$$

In the following, we also write

$$\nabla f = \begin{pmatrix} \nabla_x f \\ \nabla_v f \end{pmatrix}.$$

Hence the equation (B.3) can be written as

$$\partial_t(\nabla f) + L(\nabla f) = \sigma \Delta_v(\nabla f) - \begin{pmatrix} (K \star \nabla_x \rho) \cdot \nabla_v f \\ \nabla_x f \end{pmatrix}.$$

The evolution of $\|\nabla f\|_{L^1}$ is given by

$$\begin{aligned} \frac{d}{dt} \|\nabla f\|_{L^1} &\leq (\|K \star \nabla_x \rho\|_{L^\infty} + 1) \|\nabla f\|_{L^1} (\|K\|_{L^\infty} \|\nabla \rho\|_{L^1} + 1) \|\nabla f\|_{L^1} \\ &\leq (\|K\|_{L^\infty} \|\nabla f\|_{L^1} + 1) \|\nabla f\|_{L^1}. \end{aligned}$$

This is a closed inequality as the right-hand side only depends on $\|\nabla f\|_{L^1}$. This may blow-up in finite time because of the $\|\nabla f\|_{L^1}^2$. However there exists $T > 0$ which depends only on $\|\nabla f^0\|_{L^1}$ s.t. $\sup_{t \leq T} \|\nabla f\|_{L^1} < \infty$. This is the time interval over which Prop. 10 holds.

By the maximum principle, we can now bound $\|\nabla f\|_{L^\infty}$ up to this time T .

Indeed

$$\frac{d}{dt} \|\nabla f\|_{L^\infty} \leq (\|K\|_{L^\infty} \|\nabla f\|_{L^1} + 1) \|\nabla f\|_{L^\infty} \leq C \|\nabla f\|_{L^\infty}.$$

Observe that there cannot be any blow-up in $\|\nabla\|_{L^\infty}$ before there is blow-up in $\|\nabla\|_{L^1}$.

To conclude this step, we have obtained a time $T > 0$, s.t.

$$\|\nabla f\|_{L^1} \leq C, \quad \|\nabla f\|_{L^\infty} \leq C, \quad \forall t \leq T,$$

where C depends on $\|K\|_{L^\infty}$, $\|\nabla f^0\|_{L^1}$ and $\|\nabla f^0\|_{L^\infty}$.

Step II: Define the variable quantity

$$\Theta_f(t, \lambda) := \int_{\mathbb{D} \times \mathbb{R}^d} f \exp(\lambda |\nabla \log f|) dx dv.$$

The main object below is to bound $\Theta_f(t, \lambda)$ in $[0, T]$ for some λ as the estimate required for weak-strong uniqueness argument is

$$\sup_{t \in [0, T]} \int f \exp(\lambda |\nabla_v \log f|) dz < \infty.$$

First, we derive the equation for $\exp(\lambda |\nabla \log f|)$. Denote

$$\vec{N} = \nabla \log f = \begin{pmatrix} \vec{N}_x \\ \vec{N}_v \end{pmatrix} = \begin{pmatrix} \nabla_x \log f \\ \nabla_v \log f \end{pmatrix}, \quad \vec{n} = \frac{\vec{N}}{|\vec{N}|}.$$

By Eq. (B.3), one has that

$$\begin{aligned} & (\partial_t + L) \exp(\lambda |\nabla \log f|) = \lambda \exp(\lambda |\nabla \log f|) \vec{n} \cdot (\partial_t + L) \vec{N} \\ & = \lambda \exp(\lambda |\nabla \log f|) \vec{n} \cdot \begin{pmatrix} -(K \star \nabla_x \rho) \cdot \nabla_v \log f + \frac{\sigma}{f} (\Delta_v(\nabla_x f) - \nabla_x \log f \Delta_v f) \\ -\nabla_x \log f + \frac{\sigma}{f} (\Delta_v(\nabla_v f) - \nabla_v \log f \Delta_v f) \end{pmatrix} \\ & \leq C \lambda \exp(\lambda |\nabla \log f|) |\nabla \log f| \\ & \quad + \sigma \lambda \frac{1}{f} \exp(\lambda |\nabla \log f|) \vec{n} \cdot \begin{pmatrix} \Delta_v(\nabla_x f) - \nabla_x \log f \Delta_v f \\ \Delta_v(\nabla_v f) - \nabla_v \log f \Delta_v f \end{pmatrix}. \end{aligned}$$

Thus

$$\begin{aligned}
& \partial_t(f \exp(\lambda|\nabla \log f|)) + L(f \exp(\lambda|\nabla \log f|)) \\
& \leq C\lambda f \exp(\lambda|\nabla \log f|)|\nabla \log f| + \sigma \exp(\lambda|\nabla \log f|)\Delta_v f \\
& \quad + \sigma \lambda \exp(\lambda|\nabla \log f|)\vec{n} \cdot \begin{pmatrix} \Delta_v(\nabla_x f) - \nabla_x \log f \Delta_v f \\ \Delta_v(\nabla_v f) - \nabla_v \log f \Delta_v f \end{pmatrix}.
\end{aligned}$$

Hence, by integration by parts,

$$\frac{d}{dt} \int_{\mathbb{D} \times \mathbb{R}^d} f \exp(\lambda|\nabla \log f|) dz \leq C\lambda \int f \exp(\lambda|\nabla \log f|)|\nabla \log f| + Q_\sigma,$$

where Q_σ is an extra term due to the diffusion,

$$\begin{aligned}
Q_\sigma = & \sigma \lambda \int \frac{\exp(\lambda|\nabla \log f|)}{|\nabla \log f|} \left(\nabla_x \log f \cdot \Delta_x(\nabla_x f) - |\nabla_x \log f|^2 \Delta_v f + \right. \\
& \left. \nabla_v \log f \cdot \Delta_v(\nabla_v f) - |\nabla_v \log f|^2 \Delta_v f \right) + \sigma \int \exp(\lambda|\nabla \log f|)\Delta_v f.
\end{aligned}$$

Notice that

$$\begin{aligned}
(\nabla_x \log f) \cdot \Delta_v(\nabla_x f) &= |\nabla_x \log f|^2 \Delta_v f + 2(\nabla_x \log f) \cdot (\nabla_v f \cdot \nabla_v)(\nabla_x \log f) \\
&+ f \nabla_x \log f \cdot \Delta_v(\nabla_x \log f),
\end{aligned}$$

and

$$\begin{aligned}
(\nabla_v \log f) \cdot \Delta_v(\nabla_v f) &= |\nabla_v \log f|^2 \Delta_v f + 2(\nabla_v \log f) \cdot (\nabla_v f \cdot \nabla_v)(\nabla_v \log f) \\
&+ f \nabla_v \log f \cdot \Delta_v(\nabla_v \log f).
\end{aligned}$$

We hence obtain that

$$\begin{aligned}
Q_\sigma &= 2\lambda\sigma \int f \exp(\lambda|\nabla \log f|) \vec{n} \cdot (\vec{N}_v \cdot \nabla_v) \vec{N} + \lambda\sigma \int f \exp(\lambda|\nabla \log f|) \vec{n} \cdot \Delta_v \vec{N} \\
&\quad + \sigma \int \exp(\lambda|\nabla \log f|) \Delta_v f \\
&= \lambda\sigma \int f \exp(\lambda|\nabla \log f|) \vec{N}_v (\nabla_v \vec{N} \vec{n}) + \lambda\sigma \int f \exp(\lambda|\nabla \log f|) \vec{n} \cdot \Delta_v \vec{N} \\
&= \lambda\sigma \int f \exp(\lambda|\nabla \log f|) \vec{N}_v (\nabla_v \vec{N} \vec{n}) - \lambda\sigma \sum_{i=1}^{2d} \int f \exp(\lambda|\nabla \log f|) \nabla_v N_i \cdot \nabla_v n_i \\
&\quad - \lambda\sigma \sum_{i=1}^{2d} \int f \exp(\lambda|\nabla \log f|) (\vec{N}_v + \lambda \nabla_v \vec{N} \vec{n}) n_i \nabla_v N_i \\
&= -\lambda^2 \sigma \int f \exp(\lambda|\nabla \log f|) |\nabla_v \vec{N} \vec{n}|^2 - \lambda\sigma \int f \exp(\lambda|\nabla \log f|) \nabla_v \vec{N} \cdot \nabla_v \vec{n} \\
&\leq 0.
\end{aligned}$$

Hence,

$$\frac{d}{dt} \int_{\mathbb{D} \times \mathbb{R}^d} f \exp(\lambda|\nabla \log f|) dz \leq C\lambda \int f \exp(\lambda|\nabla \log f|) |\nabla \log f|.$$

That is

$$\partial_t \Theta_f - C\lambda \partial_\lambda \Theta_f \leq 0.$$

The characteristic equation is given by $\lambda(t) = \lambda_0 e^{-Ct}$ which implies

$$\Theta_f(t, \lambda(t)) \leq \Theta_f(0, \lambda_0) = \int f \exp(\lambda_0 |\nabla \log f|) < \infty.$$

Hence we get

$$\int f \exp(\lambda_0 e^{-Ct} |\nabla \log f|) \leq \Theta_f(0) < \infty.$$

Consequently (1.10) holds for $\lambda_f < \lambda_0 e^{-CT}$, where $C = \|K \star \nabla_{x\rho}\|_{L^\infty} + 1 < \infty$.

In the case $\sigma = 0$, we can easily propagate the bound for $|\nabla \log f|$ by tracing back the characteristics.

Bibliography

- [1] S. J. Aarseth. Gravitational N-body Simulations. Cambridge University Press, Cambridge, UK, 2010.
- [2] S. Adams, N. Dirr, M. Peletier and J. Zimmer, Large derivations and gradient flows. *Phil Trans R Soc A* **371**: 20120341.
- [3] S. M. Ahn, S.-Y. Ha, Stochastic flocking dynamics of the Cucker-Smale model with multiplicative white noise. *J. Math. Physics* **51** (2010) 103301.
- [4] L. Ambrosio, Transport equation and Cauchy problem for BV vector fields. *Invent. Math.* **158** (2004) 227–260.
- [5] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics, ETH Zurich. Birkhauser-Verlag, Basel, 2005.
- [6] L. Ambrosio, G. Crippa, Continuity equations and ODE flows with non-smooth velocity. *Proc. Roy. Soc. Edinburgh* **144A** (2014) 1191–1244.
- [7] A. A. Arsen'ev, Existence in the large of a weak solution of Vlasov's system of equations. *Zh. Vychisl. Mat. Mat. Fiz.* **15** (1975) 136–147.
- [8] C. Bardos, P. Degond, Global existence for the Vlasov-Poisson equation in 3 space variables with small initial data. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **2** (1985) 101–118.
- [9] D. Benedetto, E. Caglioti, J. A. Carrillo, M. Pulvirenti, A non Maxwellian steady distribution for one-dimensional granular media. *J. Stat. Phys.* **91** (1998) 979–990.

- [10] R. J. Berman and M. Önnheim, Propagation of chaos, Wasserstein gradient flows and toric Kähler-Einstein Metrics. Preprint, arXiv: 1501.07820, 2016.
- [11] R. J. Berman and M. Önnheim, Propagation of chaos for a class of first order models with singular mean field interactions. Preprint, arXiv: 1610.04327, 2016.
- [12] A. Blanchet and P. Degond, Topological interactions in a Boltzmann-type framework. *J. Stat. Phys.* **163** (2016) 41–60.
- [13] S. Bobkov and M. Ledoux, One-dimensional empirical measure, order of statistics, and Kantorovich transport distance. To appear in *Memoirs of the AMS*.
- [14] T. Bodineau, I. Gallagher, L. Saint-Raymond, The Brownian motion as the limit of a deterministic system of hard-spheres. *Invent. Math.* **203**, (2016) 493–553.
- [15] T. Bodineau, I. Gallagher and L. Saint-Raymond, Derivation of the Ornstein-Uhlenbeck process for a massive particle in a rarified gas of particles. Private Communication. 2017.
- [16] N. Boers, P. Pickl, On Mean Field Limits for Dynamical Systems. Preprint arXiv:1307.2999. 2013.
- [17] F. Bolley, J. A. Cañizo and J. A. Carrillo, Stochastic mean-field limit: non-Lipschitz forces and swarming. *Math. Mod. Meth. App. S.* **21** (2011) 2179–2210.
- [18] F. Bolley, J. A. Cañizo and J. A. Carrillo, Mean-field limit for the stochastic Vicsek model. *Appl. Math. Lett.* **25** (2012) 339–343.
- [19] F. Bolley, A. Guillin and F. Malrieu, Trend to equilibrium and particle approximation for a weakly self-consistent Vlasov-Fokker-Planck equation. *Math. Model. Numer. Anal.* **44** (2010) 867–884.
- [20] F. Bolley, A. Guillin and C. Villani, Quantitative concentration inequalities for empirical measures on non-compact space. *Probab. Theory Relat. Fields* **137** (2007) 541–593.
- [21] F. Bolley, C. Villani, Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des sciences de Toulouse* **14** (2005) 331–352.
- [22] F. Bouchut, Renormalized solutions to the Vlasov equation with coefficients of bounded variation. *Arch. Ration. Mech. Anal.* **157** (2001) 75–90.

- [23] F. Bouchut, G. Crippa, Lagrangian flows for vector fields with gradient given by a singular integral. *J. Hyperbolic Differ. Equ.* **10**(2) (2013) 235–282.
- [24] W. Braun and K. Hepp, The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Commun. Math. Phys.* **56** (1977) 125–146.
- [25] A. Bressan and T. Nguyen, Non-existence and non-uniqueness for multidimensional sticky particle systems. *Kinet. Relat. Models* **7** (2014) 205–218.
- [26] A. Budhiraja and R. Wu, Moderate Deviation Principles for Weakly Interacting Particle Systems. *Probab. Theory Relat. Fields* (2016). doi:10.1007/s00440-016-0723-3.
- [27] E. Caglioti, P. L. Lions, C. Marchioro, M. Pulvirenti, A special class of two-dimensional Euler Equations: A statistical mechanics description. *Commun. Math. Phys.* **143** (1992) 501–525.
- [28] E. Caglioti, P. L. Lions, C. Marchioro, M. Pulvirenti, A special class of two-dimensional Euler Equations: A statistical mechanics description. Part II. *Commun. Math. Phys.* **174** (1995) 229–260.
- [29] J. A. Carrillo, Y.-P. Choi and M. Hauray, The derivation of swarming models: Mean Field limit and Wasserstein distances. In: *Collective Dynamics from Bacteria to Crowds*, Volume 553 of CISM International Center for Mechanical Sciences, pp 1–46. Springer Vienna, 2014.
- [30] J. A. Carrillo, Y.-P. Choi, M. Hauray and S. Salem, Mean-field limit for collective behavior models with sharp sensitivity regions. *J. Eur. Math. Soc.* to appear.
- [31] J. A. Carrillo, Y.-P. Choi, P. B. Mucha and J. Peszek, Sharp conditions to avoid collisions in singular Cucker-Smale interactions. Preprint, arXiv: 1609.03447 (2016).
- [32] J. A. Carrillo, M. G. Delgadino, A. Mellet, Regularity of local minimizers of the interaction energy via obstacle problems. *Commun. Math. Phys.* **343** (2016) 747–781.
- [33] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent and D. Slepcev, Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.* **156** (2011) 229–271.

- [34] J. A. Carrillo, M. Fornasier, G. Toscani, F. Vecil, Particle, kinetic, and hydrodynamic models of swarming. In: *Mathematical modeling of collective behavior in socio-economic and life sciences*. pp. 297-336, *Model. Simul. Sci. Eng. Technol.*, Birkhauser Boston, Inc., Boston, MA, 2010.
- [35] J. A. Carrillo, S. Lisini and E. Mainini, Gradient flows for non-smooth interaction potentials. *Nonlinear Anal.* **100** (2014) 122–147.
- [36] P. Cattiaux, A. Guillin and F. Malrieu, Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Relat. Fields* **140** (2008) 19–40.
- [37] C. Cercignani, R. Illner and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*. Springer-Verlag, New York, 1994.
- [38] N. Champagnat and P.-E. Jabin, Strong solutions to stochastic differential equations with rough coefficients. Preprint, arXiv 1303.2611, (2013).
- [39] P.-H. Chavanis, Hamiltonian and Brownian systems with long-range interactions: Stochastic kinetic equations and theory of fluctuations. *Phys. A.* **387** (2008) 5716–5740.
- [40] T. Chen, C. Hainzl, N. Pavlović and R. Seiringer, Unconditional Uniqueness for the Cubic Gross-Pitaevskii Hierarchy via Quantum de Finetti. *Comm. Pure Appl. Math.* **68** (2015) 1845–1884.
- [41] G.-H. Cottet, J. Goodman, T. Y. Hou, Convergence of the grid-free point vortex method for the three-dimensional Euler equations. *SIAM J. Numer. Anal.* **28** (1991) 291–307.
- [42] F. Cucker and S. Smale, On the mathematics of emergence. *Japan. J. Math.* **1** (2007) 197–227.
- [43] P. Degond, A. Frouvelle and J.-G. Liu, Macroscopic limits and phase transition in a system of self-propelled particles. *J. Nonlinear Sci.* **23** (2013) 427–456.
- [44] C. De Lellis, Notes on hyperbolic systems of conservation laws and transport equations. In: *Handbook of differential equations, Evolutionary equations*, Vol. 3, 2007.
- [45] S. Dereich, M. Scheutzow and R. Schottstedt, Constructive quantization: approximation by empirical measures. *Ann. Inst. H. Poincaré Probab. Stat.* **49** (2013), 1183–1203.

- [46] D.A. Dewsbury, Dominance rank, copulatory behavior, and differential reproduction. *Q. Rev. Biol.* **57** (1982) 135–159.
- [47] R.J. DiPerna, P.L. Lions, Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.* **98** (1989) 511–547.
- [48] Dirr, N., Stamatakis, M., Zimmer, J.: Entropic and gradient flow formulations for nonlinear diffusion. *J. Math. Physics* **57** 081505 (2016).
- [49] R. L. Dobrusin. Vlasov equations. *Funktsional. Anal. i Prilozhen.* **13**(2) (1979) 48–56.
- [50] M. R. D’Orsogna, Y. L. Chuang, A. L. Bertozzi and L. Chayes, Self-propelled particles with soft-core interactions: patterns, stability and collapse. *Phys. Rev. Lett.* **96** (2016) 104302.
- [51] M. Duerinckx, Mean-field limits for some Riesz interaction gradient flows. *SIAM J. Math. Anal.* **48** (2016) 2269-2300.
- [52] L. C. Evans, Partial Differential Equations. Graduate Studies in Mathematics, Vol. 19, 2nd Edition, American Mathematical Society, Providence, Rhode Island, 2010.
- [53] L. Erdős, B. Schlein and H.-T. Yau, Derivation of the cubic non-linear Schrödinger equation from quantum dynamics of many-body systems. *Invent. Math.* **167** (2007) 515–614.
- [54] L. Erdős, B. Schlein and H.-T. Yau, Derivation of the Gross-Pitaevskii equation for the dynamics of Bose-Einstein condensate. *Ann. of Math.* **172** (2010) 291–370.
- [55] A. Figalli, Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.* **254** (2008) 109–153.
- [56] R.C. Fetecau and W. Sun, First-order aggregation models and zero inertia limits. *J. Differential Equations* **259** (2015) 6774-6802.
- [57] F. Flandoli, M. Gubinelli and E. Priola, Well-posedness of the transport equation by stochastic perturbation. *Invent. Math.* **180** (2010) 1–53.
- [58] F. Flandoli, M. Gubinelli and E. Priola, Full well-posedness of point vortex dynamics corresponding to stochastic 2D Euler equations. *Stoch. Process. Appl.* **121** (2011) 1445–1463.

- [59] J. Fontbona, B. Jourdain, A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations. *Ann. Probab.* **44** (2016) 131–170.
- [60] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** (2015), 707–738.
- [61] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Relat. Fields* **162** (2015) 707–738.
- [62] N. Fournier and M. Hauray, Propagation of chaos for the Landau equations with moderately soft potential. *Ann. Probab.* **44** (2016) 3581–3660.
- [63] N. Fournier, M. Hauray and S. Mischler, Propagation of chaos for the 2d viscous vortex model. *J. Eur. Math. Soc.* **16** (2014) 1425–1466.
- [64] N. Fournier and B. Jourdain, Stochastic particle approximation of the Keller-Segel Equation and two-dimensional generalization of Bessel process. Preprint, arXiv: 1507.01087 (2015).
- [65] I. Gallagher, L. Saint-Raymond and B. Texier, From Newton to Boltzmann: hard spheres and short-range potentials. In: Zurich Advanced Lectures in Mathematics Series, 2014.
- [66] K. Ganguly, H. D. Victory, Jr., On the convergence of particle methods for multidimensional Vlasov-Poisson systems. *SIAM J. Numer. Anal.* **26** (1989) 249–288.
- [67] K. Ganguly, J. T. Lee, H. D. Victory, Jr., On simulation methods for Vlasov-Poisson systems with particles initially asymptotically distributed. *SIAM J. Numer. Anal.* **28** (1991) 1574–1609.
- [68] D. Godinh and C. Quininao, Propagation of chaos for a sub-critical Keller-Segel Model. *Ann. Inst. H. Poincaré Probab. Statist.* **51** (2015) 965–992.
- [69] F. Golse, On the dynamics of large particle systems in the mean field limit. In: Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field limits and Ergodicity. In: Lecture Notes Appl. Math. Mech., Vol. 3, Springer, 2016, pp. 1–144.
- [70] F. Golse, C. Mouhot, T. Paul, On the mean field and classical limits of quantum mechanics. *Comm. Math. Phys.* **343** (2016) 165–205.

- [71] F. Golse, C. Mouhot, V. Ricci, Empirical measures and Vlasov hierarchies. *Kinet. Relat. Models* **6** (2013) 919–943.
- [72] J. Goodman, T. Y. Hou and J. Lowengrub, Convergence of the point vortex method for the 2-D Euler Equations. *Comm. Pure Appl. Math.* **43** (1990) 415–430.
- [73] M. G. Grillakis, M. Machedon and D. Margetis, Second-order corrections to mean field evolution of weakly interacting bosons. I. *Commun. Math. Phys.* **294** (2010) 273–301.
- [74] M. Z. Guo, G. C. Papanicolaou and S. R. S. Varadhan, Nonlinear diffusion limit for a system with nearest interactions. *Commun. Math. Phys.* **118** (1988) 31–59.
- [75] S.-Y. Ha, K. Lee, D. Levy, Emergence of time-asymptotic flocking in a stochastic Cucker-Smale system. *Commun. Math. Sci.* **7** (2009) 453–469.
- [76] S.-Y. Ha and E. Tadmor, From particle to kinetic and hydrodynamic description of flocking. *Kinet. Relat. Models* **1** (2008) 415–435.
- [77] D. Han-Kwan and T. T. Nguyen, Instabilities in the Mean Field Limit. *J. Stat. Phys.* **162** (2016) 1639–1653.
- [78] J. Haskovec, Flocking dynamics and mean-field limit in the Cucker- Smale-type model with topological interactions. *Physica D* **261** (2013) 42–51.
- [79] M. Hauray, Wasserstein distances for vortices approximation of Euler-type equations. *Math. Model. Meth. Appl. Sci.* **19** (2009) 1357–1384.
- [80] M. Hauray, Mean field limit for the one dimensional Vlasov-Poisson equations. *Séminaire Laurent Schwartz- EDP et applications*, **21**, 2012-2013 2013.
- [81] M. Hauray, P.-E. Jabin, N -particles approximation of the Vlasov equations with singular potential. *Arch. Ration. Mech. Anal.* **183**(3), (2007) 489–524.
- [82] M. Hauray, P.-E. Jabin, Particle Approximation of Vlasov Equations with Singular Forces. *Ann. Scient. Ecole Norm. Sup.* **48** (2015) 891–940.
- [83] M. Hauray and S. Mischler, On Kac’s chaos and related problems. *Journal of Functional Analysis* **266**(10) (2014) 6055–6157.
- [84] M. Hauray and S. Salem, Propagation of chaos for the Vlasov-Poisson-Fokker-Planck system in 1D. Preprint, arXiv 1510.06260.

- [85] C. K. Hemelrijk and H. Hildenbrandt, Self-Organized Shape and Frontal Density of Fish Schools. *Ethology* **114** (2008) 245–254.
- [86] H. Hildenbrandt, C. Carere, and C. K. Hemelrijk, Self-organised complex aerial displays of thousands of starlings: a model. *Behavioral Ecology* **21** (2010) 1349–1359.
- [87] Holding, T.: Propagation of chaos for Hölder continuous interaction kernels via Glivenko-Cantelli. Manuscript. Preprint, arXiv: 1608.02877 (2016).
- [88] K. Itô, On stochastic differential equations. *Memoirs of the American Mathematical Society* **4** (1951) 1–51.
- [89] G. Iyer and J. Mattingly, A stochastic-Lagrangian particle system for the Navier-Stokes equations. *Nonlinearity* **21** 2537 (2008).
- [90] P.-E. Jabin and F. Otto, Identification of the dilute regime in particle sedimentation. *Comm. Math. Phys.* **250** (2004) 415–432.
- [91] P.-E. Jabin, A review of mean field limits for Vlasov equations. *Kinet. Relat. Models* **7** (2014) 661–711.
- [92] P.E. Jabin and N. Masmoudi, DiPerna-Lions flow for relativistic particles in an electromagnetic field. *Arch. Ration. Mech. Anal.* **215**(3), (2015) 1029–1067.
- [93] P.-E. Jabin and Z. Wang, Mean field limit and propagation of chaos for Vlasov Systems with bounded forces. *J. Funct. Anal.* **271** (2016) 3588–3627.
- [94] P.-E. Jabin and Z. Wang, Mean field limit for stochastic particle systems. In: *Active Particles, Volume 1: Theory, Models, Applications*, Birkhauser-Springer (Boston), series *Modelling and Simulation in Science Engineering and Technology*, 2017.
- [95] P.-E. Jabin and Z. Wang, Propagation of chaos for stochastic first order systems with almost bounded stream functions. Preprint.
- [96] J. H. Jeans, On the theory of star-streaming and the structure of the universe. *Monthly Notices of the Royal Astronomical Society* **76** (1915) 70–84.
- [97] M. Kac, Foundations of kinetic theory. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, Vol. III*, pp. 171-197. University of California Press, Berkeley, 1956.

- [98] C. Kipnis and C. Landim, Scaling limit of interacting particle systems. In: *Grundlehren der mathematischen Wissenschaften* 320, Springer, 1999.
- [99] U. Krause, A discrete nonlinear and non-autonomous model of consensus formation. In: *Communications in Difference Equations, Proceedings of the Fourth International Conference on Difference Equations*, pp. 227–236, CRC Press 2000.
- [100] Y. Kuramoto, Self-entrainment of a population of coupled non-linear oscillators. In: *Lecture Notes in Physics, International Symposium on Mathematical Problems in Theoretical Physics, Volume 39*, pp. 420–422, Springer-Verlag, New York, 1975.
- [101] O.E. Lanford III, Time evolution of large classical systems. In *Dynamical systems, theory and applications (Recontres, Battelle Res. Inst., Seattle, Wash., 1974)*, pp 1-111. *Lecture Notes in Phys.*, Vol. 38. Springer, Berlin, 1975.
- [102] D. Lazarovici, Mean field limit for charged particles. PhD Thesis, 2015.
- [103] D. Lazarovici, The Vlasov-Poisson dynamics as the mean-field limit of extended charges. *Commun. Math. Phys.* **347** (2016) 271–289.
- [104] D. Lazarovici, P. Pickl, A Mean-field limit for the Vlasov-Poisson system. Preprint, arXiv 1502.04608 (2015).
- [105] E.P. Lazear and S. Rosen, Rank-order tournaments as optimum labor contracts. *J. Political Econ.* **89** (1981) 841–864.
- [106] C. Le Bris, P.L. Lions, Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients. *Comm. Partial Differential Equations* **33** (2008) 1272–1317.
- [107] P.-L. Lions, B. Perthame, Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Invent. Math.* **105** (1991) 415–430.
- [108] J.-G. Liu and R. Yang, Propagation of chaos for large Brownian particle system with Coulomb interaction. *Research in the Mathematical Sciences* to appear.
- [109] G. Loeper, Uniqueness of the solution to the Vlasov-Poisson system with bounded density. *J. Math. Pures Appl.* **86** (2006) 68–79.
- [110] D.-G. Long, Convergence of the random vortex model in two dimensional. *J. Amer. Math. Soc.* **1** (1988) 779–804.

- [111] F. Malrieu, Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stoch. Process. Appl.* **95** (2001) 109-132.
- [112] F. Malrieu, Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.* **13** (2003) 540–560.
- [113] C. Marchioro, M. Pulvirenti, Hydrodynamics in two dimensions and vortex theory. *Commun. Math. Phys.* **84** (1982) 483–503.
- [114] H. P. McKean, Jr. , Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*, pp 41-57, Air Force Office Sci. Res., Arlington, Va., 1967.
- [115] S. Méléard, Asymptotic behavior of some interacting particle systems: McKean-Vlasov and Boltzmann models. In *Probabilistic Models for Nonlinear Partial Differential Equations (Lecture Notes in Mathematics)* **1627**, Springer, 1996.
- [116] S. Mischler and C. Mouhot, Kac's Program in Kinetic Theory. *Invent. Math.* **193** (2013) 1–147.
- [117] S. Mischler, C. Mouhot and B. Wennberg, A new approach to quantitative chaos propagation for drift, diffusion and jump process. *Probab. Theory Relat. Fields* **161** (2015) 1–59.
- [118] S. Motsch and E. Tadmor, A new model for self-organized dynamics and its flocking behavior. *J. Stat. Phys.* **144** (2011) 923–947.
- [119] S. Motsch and E. Tadmor, Heterophilious dynamics enhances consensus. *SIAM Rev.* **56** (2014) 577–621.
- [120] P. B. Mucha and J. Peszek, The Cucker-Smale equation: singular communication weight, measure solutions and weak-atomic uniqueness. Preprint, ArXiv: 1509.07673, 2016.
- [121] H. Neunzert, J. Wick, The convergence of simulation methods in plasma physics. In: *Mathematical methods of plasmaphysics (Oberwolfach, 1979)*, Vol. 20 of *Methoden Verfahren Math. Phys.*, Lang, Frankfurt, pp. 271–286, 1980.
- [122] B. Øksendal, Stochastic Differential Equations: An introduction with applications, 5th edn. Springer-Verlag, 2000.

- [123] S. Olla and S.R.S. Varadhan, Scaling limit for interacting Ornstein-Uhlenbeck Processes. *Commun. Math. Phys.* **135** (1991) 355–378.
- [124] H. Osada, Propagation of chaos for the two-dimensional Navier-Stokes equation. In *Probabilistic methods in mathematical physics (Katata/Kyoto, 1985)*, pp 303-334, Academic Press, Boston, MA, 1987.
- [125] C. Pallard, Moment propagation for weak solutions to the Vlasov-Poisson system. *Comm. Partial Differential Equations* **37** (2012) 1273–1285.
- [126] K. Pfaffelmoser, Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data. *J. Differential Equations* **95** (1992) 281–303.
- [127] A. Sarantsev, Competing Brownian Particles. PhD Thesis, 2015.
- [128] J. Schaeffer, Global existence of smooth solutions to the Vlasov-Poisson system in three dimensions. *Comm. Partial Differential Equations* **16** (1991) 1313–1335.
- [129] S. Schochet, The weak vorticity formulation of the 2-D Euler equations and concentration-cancellation. *Comm. Partial Differential Equations*, **20** (1995) 1077–1104.
- [130] S. Schochet, The point-vortex method for periodic weak solutions of the 2D Euler equations. *Commun. Pure Appl. Math.* **49** (1996) 911–965.
- [131] S. Serfaty, Mean field limits of the Gross-Pitaevskii and parabolic Ginzburg-Landau equations. *J. Amer. Math. Soc.* DOI: <https://doi.org/10.1090/jams/872>, (2016).
- [132] H. Spohn, On the Vlasov hierarchy. *Math. Methods Appl. Sci.* **3** (1981) 445–455.
- [133] H. Spohn, Large scales dynamics of interacting particles. Springer Verlag, New York, 1991.
- [134] D.W. Stroock, S.R.S. Varadhan, Multidimensional Diffusion Processes. *Grundlehren Math. Wiss.*, Vol. **233**, Springer-Verlag, Berlin, 1979.
- [135] A.-S. Sznitman, Topics in propagation of chaos. In *Ecole d’été de probabilités de Saint-Flour XIX-1989*, pp 165–251, Springer, 1991.

- [136] A.B. Vasiléva, Asymptotic behavior of solutions of certain problems for ordinary non-linear differential equations with a small parameter multiplying the highest derivatives. *Uspekhi Mat. Nauk* **18** (3(111)) (1963) 15–86.
- [137] T. Vicsek, E. Czirok, E. Ben-Jacob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **75** (1995) 1226–1229.
- [138] C. Villani, Optimal Transport, Old and New. In: Grundlehren der mathematischen Wissenschaften 338. Springer Science & Business Media, 2008.
- [139] Z. Xia, The existence of noncollision singularities in Newtonian systems. *Ann. Math.* **135** (1992) 411–468.
- [140] J. Xue and D. Dolgopyat, Noncollision singularities in a planar two-center-two-body problem. *Commun. Math. Phys.* **345** (2016) 797–879.
- [141] J. Xue, Noncollision singularities in a planar four-body problem. Preprint, arXiv:1409.0048, 2014.
- [142] S. Wollman, On the approximation of the Vlasov-Poisson system by particle methods. *SIAM J. Numer. Anal.* **37** (2000) 1369–1398.
- [143] H.-T. Yau, Relative entropy and hydrodynamics of Ginzburg-Landau models. *Lett. Math. Phys.* **22** (1991) 63–80.
- [144] A. A. Vlasov, The vibration properties of an electron gas. *Sov. Phys. Usp.* **10** (1968) 721–733.
- [145] H. Xia, H. Wang and Z. Xuan, Opinion dynamics: A multidisciplinary review and perspective on future research. *International Journal of Knowledge and Systems Science* **2** (2011) 72–91.