Master's Thesis

Elite Master Course Theoretical and Mathematical Physics
Ludwig-Maximilians-Universität München
Department Mathematisches Institut

The Vlasov equation for multiple particle types

Johannes Sedlmeir

June 22, 2018

supervised by Prof. Dr. Peter Pickl

Abstract

Many physical systems, e.g. plasmas, dust clouds, galaxies or galaxy clusters, contain so many substituents, referred to as *particles*, that it is neither analytically nor numerically possible to compute their exact time evolution. However, it is often the case or at least a reasonable simplification that there are only a few different types of particles, and for each species there are numerous representants. One can therefore imagine the initial state of the system as a collection of *empirical probability measures* on phase space, i.e. the system at time 0 is construed as realization of independent random variables where for each species the particles are identically distributed with respect to a corresponding probability measure. The latter should be thought of as being given by a smooth probability density and depicting a smearing of the original point particle distribution. The law of large numbers then suggests that one might be able to approximately describe the time evolution of the true system by a suitable time evolution of the smooth densities. This is highly desirable because it would allow us to reduce the numerical complexity of the problem significantly. Consequently, our main objective is to motivate and prove what is known in the literature as *propagation of chaos*: If the initial distribution of every type of particles is close to some associated smooth initial probability density, then the true time evolution of the particles *typically* stays close to an appropriate time evolution of the densities in a physically meaningful measure of distance. One can also say that *statistical independence of the particles is almost conserved.*

First, we will heuristically derive a coupled system of PDE's for the time evolution of the above-mentioned densities in a *weak coupling regime*, namely the so-called *Vlasov equation*, first introduced in similar form by A. Vlasov in 1938. Next, we prove some basic results on existence and uniqueness of solutions in case the interaction forces are all bounded and Lipschitz continuous. Fortunately, the proof also quickly leads us to the result that propagation of chaos holds in a very strong sense under these assumptions. This extends the corresponding, well-known result for one type of particles, as, for instance, treated in Spohn's monograph [37, p. 77-82]. We also briefly discuss how to generalize the most important existence and uniqueness results for the Coulomb interaction case to multiple types of particles by giving reference to the relevant literature. Finally, we prove propagation of chaos for the Coulomb case with a cut-off depending on the particle number, generalizing a recent paper by Pickl and Lazarovici ([28]). En route, we increase the degree of detailedness and the level of mathematical rigor both for the bounded Lipschitz and the Coulomb case compared to how it is usually treated, developing the required mathematical framework in the appendix. Most notably, we prove various generalizations of Grønwall's lemma, a special high order Markov inequality and an extension of Liouville's theorem to log-Lipschitz interaction forces.

Acknowledgement

I want to give special thanks to Peter for his inspiring lectures and for granting me the opportunity to select the topics contained in this thesis autonomously according to my interests and ideas. Also, I want to particularly acknowledge my whole family for the constant both material and mental backing throughout all the years of my studies, and Andreas for the insightful discussions regarding both research and non-research issues. Of course, I also owe great thanks to all other people who contributed directly or indirectly to the final success of this work but who are not mentioned explicitly here.

Furthermore, I want to thank the Elitenetzwerk Bayern, in particular the Max-Weber-Programme, for the many years of support and the diversity of events I was allowed to take part at. They were a significant enrichment both of my education and my personal environment. Also, I highly appreciated the delightful and fertile framework for studying in the TMP programme.

Finally, I want to express my deepest gratitude to Pia for all her love and the encouragement to spend additional hours on working until late at night.

Contents

1 Motivation

1.1 Introduction and basic notions

The Vlasov equation is an effective description for the dynamics of physical systems consisting of many constituents of similar type. More precisely, it is an equation for the time evolution of some probability measures, typically given by smooth probability densities, on phase space. The solution of this equation is meant to reflect certain properties of the physical system under consideration, namely phase-space averages of a large class of classical observables, in an adequate way. Typical examples where the Vlasov equation finds application are plasmas or stellar systems, the particles being electrons and ions resp. dust particles, stars or galaxies.

We assume that the constituents, which will be called *particles* in the following, can be described by point masses/charges and clustered into a comparably small number of types, each present in a great many of representants. We will always consider the classical, non-relativistic situation with a two-body interaction depending only on the relative coordinates of the particles. However, it turns out that external forces can be added without any complication provided the existence theory of solutions in chapter 2 works out nicely.

Let n denote the number of particle types. Two particles are regarded as being of the same type if both they have equal mass and exhibit the same interaction force with any other particle in the system. For $k \in \{1, \dots, n\}$, we denote by N_k the number of particles of type k and introduce the sets $\Gamma_k := \{i \in \mathbb{N} : 1 + \sum_{l=1}^{k-1} N_l \leq i \leq \sum_{l=1}^{k} N_l\}$. Then obviously $|\Gamma_k| = N_k$ for all $k \in [n]$, where we use the common notation $[d] := \{1, \dots, d\}$ for $d \in \mathbb{N}$. Finally, $N := \sum_{k=1}^{n} N_k$ is the total number of particles.

We now arrange the numeration of the particles such that these of type k have indices in Γ_k and write the trajectory of the system on phase space via

$$X(t) := \big(Q(t), P(t)\big) := \big(Q_1(t), \dots, Q_n(t), P_1(t), \dots, P_n(t)\big) \in \mathbb{R}^{6N},$$

where for $k \in [n]$,

$$\begin{aligned} Q_k(t) : &= \big(Q_k^1(t), \dots, Q_k^{N_k}(t)\big) := \big(q_{\min \Gamma_k}(t), \dots, q_{\max \Gamma_k}(t)\big) \in \mathbb{R}^{3N_k}, \\ P_k(t) : &= \big(P_k^1(t), \dots, P_k^{N_k}(t)\big) := \big(p_{\min \Gamma_k}(t), \dots, p_{\max \Gamma_k}(t)\big) \in \mathbb{R}^{3N_k}. \end{aligned}$$

The vectors $q_i(t), p_i(t) \in \mathbb{R}^3$, $i \in [N]$, represent the position respectively the momentum of the i-th particle at time t. Let the particles of type k have mass $m_k > 0$. By Newton's equations, translated into a first order system (i.e. considered on *phase space*), the components of the trajectory $X(t)$ satisfy the coupled system of autonomous, first-order ordinary differential equations (ODEs)

$$\dot{q}_i(t) = \frac{p_i(t)}{m_k}, \qquad \dot{p}_i(t) = \sum_{l \in [n]} \sum_{\substack{j \in \Gamma_l \\ j \neq i}} f_{k,l}\big(q_i(t) - q_j(t)\big) =: F_i(Q(t)), \qquad k \in [n],\ i \in \Gamma_k. \tag{1.1}$$

Here, for $k, l \in [n]$, $f_{k,l}$ denotes the pair interaction force between particles of type k and l.

We are interested in the case where all N_k are very large, while n is relatively small. For example, in a plasma in a fusion reactor, there are a few grams of hydrogen, rather to be considered as a collection of protons and electrons, leading to $n = 2$ and $N_1, N_2 \approx 10^{23}$. Unfortunately, for N that big it is practically impossible to solve (1.1), both analytically and numerically: An analytic solution typically cannot be obtained even for $N = 3$, and the largest N-body simulation of a gravitational system carried out by recent supercomputers yet can handle only $N \approx 10^{12}$ particles. Therefore, we have to come up with some new concepts in order to derive an approximation to the solution of (1.1) which at least describes some of the true system's physical properties satisfactorily. The crucial idea is to consider the system in a probabilistic way, representing the discrete particle distribution by smooth probability densities and replacing the pair interactions by an external field. This allows us to apply powerful tools such as multivariable calculus or the law of large numbers and thus leads to an approximation of the system which is accessible to numerical computation.

Let us for the moment assume that the interaction forces $f_{k,l}$ are bounded and smooth with $f_{k,l}(0) = 0$. In this case, we may include the summands $f_{k,l}(q_i(t) - q_i(t))$ on the r.h.s. of (1.1). Since pair-interactions are usually radially symmetric, the latter hypothesis is actually not a big additional restriction. Moreover, it is easily seen that even the case $f_{k,l}(0) \neq 0$ should not be a problem to deal with because one can *shift* $f_{k,l}(0)$ into a constant external force, i.e. replace $f_{k,l}$ by $\tilde{f}_{k,l} := f_{k,l} - f_{k,l}(0)$, which then obviously satisfies $\tilde{f}_{k,l}(0) = 0$, and compensate for this by adding the constant external force $f_{k,l}(0)$, which only particles of type k are coupled to, in (1.1). Of course, for the physically most interesting case, namely the Coulomb interaction, the maps $f_{k,l}$ are not bounded; they are not even defined in $0 \in \mathbb{R}^3$. However, in the motivational part, we want to stick to the mathematically less troublesome case in order to focus on the physical motivation. The rigorous treatment will then enter in chapters 2 and 3, where we work our way through from *easy* pair interactions, namely bounded Lipschitz forces, towards the Coulomb case.

Under the above-mentioned preliminary assumptions, it is well-known that for every initial condition $Z := X(0) \in \mathbb{R}^{6N}$, there is a unique, global solution $X(t)$ to the system (1.1). In order to emphasize the dependence on the initial condition, we will denote this solution by

$$\Psi_t(Z) := \left(\Psi_t^1(Z), \Psi_t^2(Z)\right) = \left(\Psi_{1,t}^1(Z), \ldots, \Psi_{n,t}^1(Z), \Psi_{1,t}^2(Z), \ldots, \Psi_{n,t}^2(Z)\right),$$

where for $k \in [n]$, $(\Psi_{k,t}^1(Z), \Psi_{k,t}^2(Z)) = (Q_k(t), P_k(t))$ and

$$\begin{aligned}
\left(\Psi_{k,t}^{1,1}(Z), \ldots, \Psi_{k,t}^{1,N_k}(Z)\right) &:= \left(Q_k^1(t), \ldots, Q_k^{N_k}(t)\right) = \Psi_{k,t}^1(Z), \\
\left(\Psi_{k,t}^{2,1}(Z), \ldots, \Psi_{k,t}^{2,N_k}(Z)\right) &:= \left(P_k^1(t), \ldots, P_k^{N_k}(t)\right) = \Psi_{k,t}^2(Z).
\end{aligned}$$

For fixed initial condition $Z \in \mathbb{R}^{6N}$ and $k \in [n]$, we now introduce the time-dependent *empirical probability measure* for particles of type k on phase space $\mathbb{R}^6$, namely the map

$$\mu_{\mathrm{emp},k}^Z : \mathbb{R} \to \mathcal{P}(\mathbb{R}^6), \qquad t \mapsto \mu_{\mathrm{emp},k,t}^Z := \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{\left(\Psi_{k,t}^{1,i}(Z), \Psi_{k,t}^{2,i}(Z)\right)}. \tag{1.2}$$

Here, $\mathcal{P}(\mathbb{R}^6)$ denotes the space of all probability measures on $\mathbb{R}^6$, and for $y \in \mathbb{R}^6$, δ_y is the *Dirac measure* with mass at y. Let us also decompose the initial state $Z = \Psi_0(Z)$ accordingly, denoting $Z_k^i := (\Psi_{k,0}^{1,i}(Z), \Psi_{k,0}^{2,i}(Z))$. Now, the probabilistic image enters the scene: If we imagine $(Z_k^i)_{i \in \mathbb{N}}$ as independent random variables distributed according to the law $\mu_{k,0}$ for some (initial) probability measure $\mu_{k,0}$ on $\mathbb{R}^6$, then for $N_k \to \infty$, by the law of large numbers we expect some kind of convergence of the empirical probability measure, i.e.

$$\mu^Z_{\mathrm{emp},k,0} = \frac{1}{N_k}\sum_{i=1}^{N_k}\delta_{\left(\Psi^{1,i}_{k,0}(Z),\Psi^{2,i}_{k,0}(Z)\right)} = \frac{1}{N_k}\sum_{i=1}^{N_k}\delta_{Z^i_k} \xrightarrow{\text{some sense}} \mu_{k,0}.$$

The precise notion of convergence is not important at this stage, however, the interested reader might want to catch a glimpse at chapter 4.5 already. To get the connection between the empirical probability measure and the Newtonian equations of motion, note that given the solution $\Psi_t(Z)$, we can rewrite (1.1) with initial condition $X(0) = Z \in \mathbb{R}^{6N}$ via the (now no more autonomous) first-order system

$$\dot{q}_i(t) = \frac{p_i(t)}{m_k}, \qquad \dot{p}_i(t) = N \cdot F_k^{\mu^Z_{\mathrm{emp}}}(q_i(t),t), \qquad k \in [n],\ i \in \Gamma_k, \tag{$*$}$$

where, using $f_{k,l}(0) = 0$ and thus including the summands for $j = i$ in the force term in (1.1),

$$\begin{aligned} F_k^{\mu^Z_{\mathrm{emp}}}(q,t) &= \frac{1}{N}\sum_{l\in[n]}\sum_{j\in\Gamma_l} f_{k,l}\big(q - q_j(t)\big) = \sum_{l=1}^{n}\frac{N_l}{N}\cdot\frac{1}{N_l}\sum_{j=1}^{N_l} f_{k,l}\big(q - \Psi^{1,j}_{l,t}(Z)\big) \\ &= \sum_{l=1}^{n}\frac{N_l}{N}\cdot\frac{1}{N_l}\sum_{j=1}^{N_l}\int_{\mathbb{R}^6} f_{k,l}(q-\tilde{q})\,\mathrm{d}\delta_{\left(\Psi^{1,j}_{l,t}(Z),\Psi^{2,j}_{l,t}(Z)\right)}(\tilde{q},\tilde{p}) \\ &= \sum_{l=1}^{n}\frac{N_l}{N}\cdot\int_{\mathbb{R}^6} f_{k,l}(q-\tilde{q})\,\mathrm{d}\mu^Z_{\mathrm{emp},l,t}(\tilde{q},\tilde{p}) \\ &=: \sum_{l=1}^{n}\alpha_l\cdot(f_{k,l} *_q \mu^Z_{\mathrm{emp},l,t})(q). \end{aligned} \tag{1.3}$$

Here, for $l \in [n]$, $\alpha_l := \frac{N_l}{N}$ denotes the relative number of particles of type l in the system, and in a slight abuse of notation, $f_{k,l}$ is also regarded as a function on phase space $\mathbb{R}^6$ in the obvious way. More details on this and the operation $*_q$ can be found in section 4.3, see particularly remark 4.30. For simplification, we will omit the superscript Z when writing down the empirical measures from now on, having in mind that just like solutions of (1.1), they depend on the choice of initial condition.

In order to have a shorter notion at hand for letting $N_k \to \infty$ for all $k \in [n]$, we introduce the number of particles of the type with the least representants, $S := \min\{N_k : k \in [n]\}$, which will be of substantial importance later on. For large S, we can imagine that the factors α_l stabilize; we regard them as the share of particles of the corresponding type in the system. Since we expect that as $S \to \infty$, for every $k \in [n]$ the initial empirical probability measure $\mu_{\mathrm{emp},k,0}$ converges to $\mu_{k,0}$ in some sense, there is hope that for fixed time t, every $\mu_{\mathrm{emp},k,t}$ converges to a corresponding probability measure $\mu_{k,t}$ in the very same sense. This in turn would imply some kind of convergence of the true force $F_k^{\mu_{\mathrm{emp}}}$ to a *mean field force* F_k^{μ}. Our goal in this thesis is to find the right notions for this idea and make things rigorous. However, note that so far, the pre-factor N in $(*)$ destroys any hope for convergence in the sense we just described. Hence, heuristically we need a pre-factor N^{-1} on the r.h.s. of (1.1), i.e. a dampening of the forces proportional to the total number of particles in the system. Actually, one can readily convince oneself that a scaling of the time coordinate by the factor $N^{\frac{1}{2}}$ yields this pre-factor. That corresponds to considering the system in slow-motion, therefore decreasing accelerations and consequently the perceived strength of forces. However, this approach is not desirable for the following reason: As already indicated, our goal will be to prove that in the scaled system, an appropriate time evolution of the probability measures $\mu_{k,t}$, which is yet to be determined, will stay close to the true time evolution, represented by $\mu_{\text{emp},k,t}$, for finite times t provided they are all close initially, i.e. in the limit $S \to \infty$. However, in unscaled (*physical*) time, this would then only hold for very short times, namely for $N^{-\frac{1}{2}}t$, making our results practically useless since also $N \to \infty$. Fortunately, there is another change of coordinates yielding the pre-factor N^{-1} which works out particularly well for a system with gravitational or electrostatic interactions: Assume that all $f_{k,l}$ are homogeneous of degree -2, i.e. $f_{k,l}(\lambda\,\cdot\,) = \lambda^{-2} f_{k,l}$ for all $\lambda > 0$. This is precisely the case for the Coulomb force $f(q) \sim q \cdot |q|^{-3}$. Then obviously the force F_i acting on any particle i in (1.1) is homogeneous of degree -2, too. Consequently, describing the original physical system by coordinates $\tilde{X}(t) = (\tilde{Q}(t), \tilde{P}(t))$ and making the scaling $X(t) := N^{-\frac{1}{3}} \tilde{X}(t)$, we see by the chain rule that

$$
\begin{aligned}
\dot{q}_i(t) &= N^{-\frac{1}{3}} \cdot \dot{\tilde{q}}_i(t) = N^{-\frac{1}{3}} \cdot \frac{\tilde{p}_i(t)}{m_k} = \frac{p_i(t)}{m_k}, \\
\dot{p}_i(t) &= N^{-\frac{1}{3}} \cdot \dot{\tilde{p}}_i(t) = N^{-\frac{1}{3}} \cdot F_i(\tilde{Q}(t)) = N^{-\frac{1}{3}} \cdot N^{-\frac{2}{3}} F_i(N^{-\frac{1}{3}} \tilde{Q}(t)) = N^{-1} \cdot F_i(Q(t)),
\end{aligned}
$$

i.e. we obtain precisely the desired pre-factor N^{-1}. This coordinate transform corresponds to multiplying lengths (and also momenta) in the physical system by a factor $N^{\frac{1}{3}}$ and consequently to enlarging spatial volumes by a factor $(N^{\frac{1}{3}})^3 = N$. Hence, we can interpret the limit $S \to \infty$ with $\frac{N_k}{N} \to \alpha_k$ for appropriate $\alpha_k \in (0,1)$ as a macroscopic limit, increasing the size of the physical system but leaving the number of particles of each type per unit of volume, i.e. the average particle densities, constant. Thus, the system which we want to start our investigations from is not (1.1), but rather

$$
\dot{q}_i(t) = \frac{p_i(t)}{m_k}, \qquad \dot{p}_i(t) = \frac{1}{N} \sum_{l \in [n]} \sum_{\substack{j \in \Gamma_l \\ j \neq i}} f_{k,l}\big(q_i(t) - q_j(t)\big), \qquad k \in [n],\ i \in \Gamma_k. \tag{1.4}
$$

Note once more that by our temporary working hypothesis, we may include the summands $j = i$ on the r.h.s. of (1.4). In the literature, physical systems described by a system of ODEs of this form are usually called *weakly coupled*, see, for instance, [10]. This refers to the pre-factor N^{-1}, which at first glance seems artificial and unphysical. However, having in mind the preceding discussion, it does physically make perfect sense for systems with gravitational and electrostatic interactions. Sometimes, one also finds the expression *weak* or *long range interactions* in this context, e.g. in [37]. Nevertheless, we are going to analyze (1.4) also for pair interaction forces $f_{k,l}$ which are not homogeneous of degree -2. We do this not only because it is an interesting problem from the mathematical perspective, but also because it is reasonable to find concepts and explore possible theorems for easier, e.g. bounded, forces before turning to the more challenging, physically interesting case (note that functions which are homogeneous of degree -2 are either identically 0 or unbounded at the origin).

As already explained, we intend to find an approximation to (1.4) which is accessible to numerical computation for large S. We are going to use the above-mentioned ideas and regard the point particle distribution as a realization of probability measures (ultimately, rather smooth probability densities) on phase space, which in turn induce an external field, to be thought of as an approximation of the true interaction forces. This, in turn will allow us to find some *product structure* for what we expect to describe an approximate time evolution of the physical system, therefore virtually reducing the phase space dimension from $6N$ to $6n$. However, the price we have to pay is that finally we get from a system of $6N$ coupled, autonomous first order ODEs to a system of n coupled, first-order PDEs. But for fixed n, the complexity of the PDE will not change with the number of particles in the system, so when S is large we clearly obtain a big gain.

1.2 Heuristic derivation of the Vlasov equation

Looking at (1.3), it is a reasonable conjecture that in general, the force field $F_k^{\nu_l}(q)$ exerted on a *test particle* of type k at position q and generated by an arbitrary mass (charge) distribution of particles of type l, which is in turn represented by a probability measure ν_l on one-particle phase space $\mathbb{R}^6$, is in the weakly coupled system (1.4) given by $F_k^{\nu_l}(q) = (f_{k,l} *_q \nu_l)(q)$. One can justify this also as follows: Assume that ν_l has a density (Radon-Nikodym derivative) v_l w.r.t Lebesgue measure. This suffices for physical considerations because by a computation analogous to the one carried out for $F_k^{\mu_{\mathrm{emp}}}$ in (1.3), the statement already holds for arbitrary discrete probability measures representing a single-type point particle system. Then the expected number ΔN_l of particles with position in a small subset $A \subset \mathbb{R}^3$ (meaning A is contained in a ball with small radius) is in good approximation given by

$$\Delta N_l = \nu_l(A \times \mathbb{R}^3) = \int_{A\times\mathbb{R}^3} v_l(\overline{q},\overline{p})\,\mathrm{d}\overline{q}\,\mathrm{d}\overline{p} = \int_A \left(\int_{\mathbb{R}^3} v_l(\overline{q},\overline{p})\,\mathrm{d}\overline{p} \right) \mathrm{d}\overline{q} \approx |A| \cdot \rho_l(\tilde{q}),$$

where $\rho_l(q) := \int_{\mathbb{R}^3} v_l(q,\overline{p})\,\mathrm{d}\overline{p}$ is the *spatial density* of particles of type l, which for simplification is assumed to be continuous (at least, this is true for v_l continuous with compact support in the p-variable), and $\tilde{q} \in A$ can be chosen arbitrarily. Consequently, the force exerted on our test particle by particles of type l in this region is roughly given by $f_{k,l}(q-\tilde{q}) \cdot \Delta N_l$. Splitting position space $\mathbb{R}^3$ into small, disjoint sets and summing things up, one is left with a Riemann sum which in the limit of high granularity converges to

$$\int_{\mathbb{R}^3} f_{k,l}(q-\tilde{q}) \cdot \rho_l(\tilde{q})\,\mathrm{d}\tilde{q} = \int_{\mathbb{R}^6} f_{k,l}(q-\tilde{q}) \cdot v_l(\tilde{q},\tilde{p})\,\mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p} = \int_{\mathbb{R}^6} f_{k,l}(q-\tilde{q})\,\mathrm{d}\nu_l(\tilde{q},\tilde{p}) = (f_{k,l} *_q \nu_l)(q).$$

For those readers who do not approve of this derivation of $F_k^{\nu_l}$, the approach to the formula under consideration by means of the marginals, as introduced in section 4.3 might be more insightful. Now, we go one step further and assume that we were given time-dependent probability measures $\nu_{l,t}$, $l \in [n]$, representing the distribution of particles of the corresponding type at all times. Taking into account the shares α_l of the types w.r.t. the total particle number N, this would impose the time-dependent external force

$$F_k^{\nu}(q,t) := \sum_{l\in[n]} \alpha_l \cdot (f_{k,l} *_q \nu_{l,t})(q) \tag{1.5}$$

on a test particle of type k at position q and time t. Hence, under the purely external force F_k^{ν}, the N-particle system would evolve according to

$$\dot{q}_i(t) = \frac{p_i(t)}{m_k}, \qquad \dot{p}_i(t) = F_k^{\nu}(q_i(t),t), \qquad k \in [n],\ i \in \Gamma_k. \tag{1.6}$$

For nice regularity properties of all $f_{k,l}$ and ν_l, also the F_k^{ν} should behave nicely, and therefore it is reasonable to assume that (1.6) admits a unique, global flow Φ_t^{ν}. As for Ψ_t, we decompose Φ_t^{ν} into

$$\Phi_t^{\nu} = \left(\Phi_t^{\nu,1}, \Phi_t^{\nu,2}\right) = \left(\Phi_{1,t}^{\nu,1}, \dots, \Phi_{n,t}^{\nu,1}, \Phi_{1,t}^{\nu,2}, \dots, \Phi_{n,t}^{\nu,2}\right).$$

By the special structure of (1.6), we immediately see that for $k \in [n]$,

$$\Phi_{k,t}^{\nu,1} = \prod_{i=1}^{N_k} \varphi_{k,t}^{\nu,1}, \qquad \Phi_{k,t}^{\nu,2} = \prod_{i=1}^{N_k} \varphi_{k,t}^{\nu,2},$$

where $\varphi^\nu_{k,t} := (\varphi^{\nu,1}_{k,t}, \varphi^{\nu,2}_{k,t})$ denotes the *one-particle flow* for particles of type k, i.e. the flow for the first order system

$$\dot{q}(t) = \frac{p(t)}{m_k}, \qquad \dot{p}(t) = F^\nu_k(q(t), t). \tag{1.7}$$

This the product structure that has been announced before. Hence, for every $k \in [n]$ we obtain a flow $\varphi^\nu_{k,t}$ on $\mathbb{R}^6$ which tells us the motion of test particles of type k in the force field created by the collection of time-dependent probability measures $\nu_{l,t}$ on phase space. This, in turn, tells us how an initial *test distribution* of particles of type k, represented by a probability measure $\mu_{k,0}$ on phase space, should evolve in time: Heuristically, the probability measures $\mu_{k,t}$ *count* particles, i.e for every measurable $A \subset \mathbb{R}^6$ and $t \in \mathbb{R}$, $\mu_{k,t}(A)$ is the expected number of particles of type k which at time t are located in the subset A of phase space. However, since the particles move in the external field F^ν_k, i.e. according to the flow $\varphi^\nu_{k,t}$, these are exactly the particles which at time 0 have been in the phase-space region $(\varphi^\nu_{k,t})^{-1}(A)$, i.e. $\mu_{k,t}$ is given by the *pushforward* or *image measure* of $\mu_{k,0}$ under the flow $\varphi^\nu_{k,t}$:

$$\mu_{k,t} = \varphi^\nu_{k,t}\#\mu_{k,0} = \mu_{k,0} \circ (\varphi^\nu_{k,t})^{-1}, \qquad k \in [n]. \tag{$**$}$$

In case the reader is not yet familiar with the concept of a pushforward measure and/or its characteristic property with respect to integration, she is recommended to briefly scroll to the beginning of section 4.3 because we will heavily rely on lemma 4.23 in the sequel.

Looking for a reasonable time evolution of some given initial probability measures $\mu_{k,0}$, we now naturally require that for all $k \in [n]$, the probability measures evolve precisely according to the force field which they generate themselves, i.e. for all $k \in [n]$, we set $\mu_{k,t} = \nu_{k,t}$ in ($**$). This already yields the **Vlasov equation in integral form** for n types of particles, namely

$$\mu_{k,t} = \mu_{k,0} \circ \left(\varphi^\mu_{k,t}\right)^{-1}, \qquad k \in [n]. \tag{1.8}$$

This form of the Vlasov equation is yet very general, including the possibility of both discrete and continuous parts in the probability distributions. In fact, we will see in section 2.1 that the collection of empirical distributions $(\mu_{\mathrm{emp},1,t}, \dots, \mu_{\mathrm{emp},n,t})$ does indeed solve 1.8 provided $f_{k,l}(0) = 0$, so in this case, the Vlasov equation is nothing but a reformulation of (1.4) in terms of the empirical measure. In particular, (1.8) does not yet seem to be more accessible to numerical computation than (1.4). However, things change if we focus on initial probability distributions $\mu_{k,0}$ which have smooth probability densities $u_{k,0}$ w.r.t. Lebesgue measure because under this additional assumption, we can show that the time-evolved measures $\mu_{k,t}$, i.e. solutions of the Vlasov equation in integral form (1.8), stay absolutely continuous w.r.t. Lebesgue measure, and determine a differential equation for the corresponding densities $u_{k,t}$: by Liouville's theorem, $\varphi^\mu_{k,t}$ preserves 6-dimensional Lebesgue measure provided the forces F^μ_k exhibit some quite weak regularity assumptions, for more details on this see section 4.9. It follows that for every $A \in \mathcal{B}(\mathbb{R}^6)$,

$$\begin{aligned}
\mu_{k,t}(A) = \mu_{k,0}\big((\varphi^\mu_{k,t})^{-1}(A)\big) &= \int_{\mathbb{R}^6} \mathbb{1}_{(\varphi^\mu_{k,t})^{-1}(A)} \cdot u_{k,0}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^6} \big(\mathbb{1}_A \circ \varphi^\mu_{k,t}\big) \cdot \big(u_{k,0} \circ (\varphi^\mu_{k,t})^{-1} \circ \varphi^\mu_{k,t}\big)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^6} \mathbb{1}_A \cdot \big(u_{k,0} \circ (\varphi^\mu_{k,t})^{-1}\big)\,\mathrm{d}x.
\end{aligned} \tag{1.9}$$

Here we used (4.33), which is basically the substitution or transformation formula for a measure preserving change of coordinates, thus not containing a term for volume distortion. We now immediately see that $\mu_{k,t}$ has the density $u_{k,0} \circ (\varphi^{\mu}_{k,t})^{-1}$ w.r.t. 6-dimensional Lebesgue measure. In our notation, we will often identify the probability measures $\mu_{k,t}$ with their corresponding probability densities under these circumstances and write $\varphi^{u}_{k,t} := \varphi^{\mu}_{k,t}$ and $F^{u}_{k} := F^{\mu}_{k}$.

Let us now regard the time-dependent densities $u_{k,t} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ as maps $u_k : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ by the obvious prescription $u_k(q,p,t) := u_{k,t}(q,p)$. Then our above calculations show that for all $(q,p,t) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}$, $u_k(q,p,t) = u_k((\varphi^{\mu}_{k,t})^{-1}(q,p), 0)$, or, equivalently by bijectivity of $\varphi^{\mu}_{k,t}$, $u_k(q,p,0) = u_k(\varphi^{\mu}_{k,t}(q,p), t)$. Expecting continuous differentiability of u_k w.r.t. all arguments, applying the chain rule we obtain that for all $(q,p,t) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}$,

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}t}\, u_k(q,p,0) = \frac{\mathrm{d}}{\mathrm{d}t}\, u_k(\varphi^{u}_{k,t}(q,p), t) = \frac{\mathrm{d}}{\mathrm{d}t}\, u_k(\varphi^{u,1}_{k,t}(q,p), \varphi^{u,2}_{k,t}(q,p), t) \\
&= \partial_t u_k(\varphi^{u}_{k,t}(q,p),t) + \nabla_q u_k(\varphi^{u}_{k,t}(q,p),t) \cdot \partial_t \varphi^{u,1}_{k,t}(q,p) + \nabla_p u_k(\varphi^{u}_{k,t}(q,p),t) \cdot \partial_t \varphi^{u,2}_{k,t}(q,p) \\
&= \partial_t u_k(\varphi^{u}_{k,t}(q,p),t) + \nabla_q u_k(\varphi^{u}_{k,t}(q,p),t) \cdot \frac{\varphi^{u,2}_{k,t}(q,p)}{m_k} + \nabla_p u_k(\varphi^{u}_{k,t}(q,p),t) \cdot F^{u}_{k}(\varphi^{u,1}_{k,t}(q,p),t).
\end{aligned} \tag{1.10}$$

In the last step, we used that $\varphi^{u}_{k,t}$ is the one-partice flow for the external force field induced by u defined by (1.7). But clearly, for any $(q,p,t) \in \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}$, we can find a flow line for $\varphi^{\mu}_{k,t}$ passing through (q,p) at time t, namely the one with initial condition $(\varphi^{u}_{k,t})^{-1}(q,p)$. Consequently, we obtain that u_k satisfies the partial differential equation (PDE)

$$\partial_t u_k(q,p,t) + \tfrac{p}{m_k} \cdot \nabla_q u_k(q,p,t) + F^{u}_{k}(q,t) \cdot \nabla_p u_k(q,p,t) = 0.$$

Together with the expression from (1.5) for F^{u}_{k}, we thus finally arrive at the **Vlasov equation in differential form** for n types of particles, i.e. the coupled system of (quasi-linear, first-order) PDEs which written in short form read

$$\partial_t u_k + \tfrac{p}{m_k} \cdot \nabla_q u_k + F^{u}_{k} \cdot \nabla_p u_k = 0, \qquad F^{u}_{k} = \sum_{l \in [n]} \alpha_l \cdot (f_{k,l} *_q u_{l,t}), \qquad k \in [n]. \tag{1.11}$$

In the following chapter, we are going to analyze both the Vlasov equation in differential and integral form. We will concentrate on the case where the interaction forces $f_{k,l}$ are bounded and Lipschitz continuous. First, we give precise definitions of being a solution to the Vlasov equation in its different forms, establish connections between these and digress on some interesting properties of solutions such as energy conservation. Afterwards, we prove existence and uniqueness of solutions. We also briefly mention how to proceed in the Coulomb case, giving references to the relevant literature. Afterwards, in chapter 3 we show that solutions of the Vlasov equation are indeed a good approximation for the true time evolution of the (weakly coupled) system 1.4 in a physically reasonable sense. Hence, fortunately it turns out that the Vlasov equation can indeed fulfil the purpose which we constructed it for. The proof that the Vlasov equation describes some aspects of the system on a macroscopic scale can therefore be regarded as a *rigorous derivation* of the Vlasov equation from the microscopic time evolution, however with the blemish that yet we need a regularization of the Coulomb force at the singularity.

2 Solutions to the Vlasov equation

2.1 Rigorous formulation and first results

We now want to develop precise definitions for being a solution to the Vlasov equation (1.8) resp. (1.11). In the sequel, the parameters n, N_k, N and thus α_k are unless otherwise stated fixed and as introduced in section 1.1. We start with treating the Vlasov equation in integral form. In order to make sense of the considerations presented in the heuristics from the first chapter and in particular of (1.8), we need to ensure existence of the flows Φ_t^μ for (1.6), i.e. of the one-particle flows $\varphi_{k,t}^\mu$ for (1.7), which is of course closely related to the regularity of the forces F_k^μ. From the theory of ODEs, it is well-known that a unique, global solution of (1.7) exists provided all F_k^μ are continuous maps which are uniformly in t Lipschitz continuous in q, see also section 4.8 where we prove a generalization of this statement. On the other hand, from the physical perspective, the forces should not get arbitrarily large. This suggests restricting to pair interactions $f_{k,l}$ such that all $F_k^\mu(\,\cdot\,, t) \in \mathrm{BL}(\mathbb{R}^3; \mathbb{R}^3)$, where $\mathrm{BL}(\mathbb{R}^3; \mathbb{R}^3)$ denotes the space of *bounded Lipschitz functions*, which together with a suitable norm is introduced in section 4.4. Note that the special form of equation (1.8) makes sure that a solution which is initially a collection of probability measures remains a collection of probability measure for all times provided the flows $\varphi_{k,t}^\mu$ exist and are measurable for every t. Expecting that this quite weak hypothesis is satisfied, one can readily check that boundedness resp. Lipschitz continuity of all $f_{k,l}$ ensures boundedness resp. Lipschitz continuity of the $F_k^\mu(\,\cdot\,, t)$ uniformly in t. So, the only thing we still have to care about is continuity of the mean field forces F_k^μ w.r.t. time. A condition which is obviously sufficient is that for every $q \in \mathbb{R}^3$, the maps $t \mapsto (f_{k,l} *_q \mu_{k,t})(q)$ are continuous. Consequently, we see that *weak continuity* of the measures $\mu_{k,t}$ is a good notion to guarantee continuous dependence of all F_k^μ on t, where we call a curve $\mu : I \to \mathcal{P}(\mathbb{R}^6)$ from some subset $I \subset \mathbb{R}$ into the space of probability measures weakly continuous if for all $g \in \mathrm{BL}(\mathbb{R}^6)$, the map $t \mapsto \int_{\mathbb{R}^6} g \, \mathrm{d}\mu(t)$ is continuous. We denote by $\mathcal{C}^*(I; (\mathcal{P}(\mathbb{R}^6))^n)$ the set of all *vector valued weakly continuous curves*, i.e. every component of $\mu \in \mathcal{C}^*(I; (\mathcal{P}(\mathbb{R}^6))^n)$ is a weakly continuous curve into $\mathcal{P}(\mathbb{R}^6)$. A formal definition of weak continuity which is slightly more general, as well as some important completeness properties for spaces of weakly continuous curves, which we will heavily rely on in the existence and uniqueness proof in section 2.2, can be found in chapter 4.4. Finally, let us mention that lemma 2.9 basically gives a detailed proof for those arguments above which were only sketched briefly.

We are going to concentrate on solutions of the Vlasov equation on finite time intervals $[0, T]$ where $T > 0$. If we can prove existence and uniqueness of solutions for arbitrary $T > 0$, we will also have established global existence and uniqueness: for positive times, this is clear. On the other hand, it is physically obvious and mathematically checkable from the still due, final definitions both for being a solution of the integral and differential form of the Vlasov equation that going back in time is the same as going forward in time with reversed initial momenta: given initial probability measures $\mu_{k,0}$, one can introduce set functions $\tilde{\mu}_{k,0}$ by defining $\tilde{\mu}_{k,0}(A_1 \times A_2) := \mu_{k,0}(A_1 \times (-A_2))$ on products, i.e. for $A_1, A_2 \in \mathcal{B}(\mathbb{R}^3)$. Employing standard measure theoretic arguments which are also abundantly used in section 4.3, one can readily show that $\tilde{\mu}_{k,0}$ extends uniquely to a probability measure $\tilde{\mu}_{k,0}$ on $\mathcal{B}(\mathbb{R}^6) = \mathcal{B}(\mathbb{R}^3) \otimes \mathcal{B}(\mathbb{R}^3)$, and that in case $\mu_{k,0}$ has probability density $u_{k,0}$, it holds that $\tilde{\mu}_{k,0}$ has probability density $\tilde{u}_{k,0}(q,p) = u_{k,0}(q,-p)$ almost everywhere. It is then easy to prove that a solution $\tilde{\mu}_t$ resp. $\tilde{u}_t$ on $[0,T]$ to the Vlasov equation with initial state $\tilde{\mu}_0$ resp. $\tilde{u}_0$ induces a solution μ_t resp. u_t of the Vlasov equation on $[-T,0]$ to the initial (rather: final) condition μ_0 resp. u_0 via $\mu_t := \tilde{\mu}_{-t}$ resp. $u_t := \tilde{u}_{-t}$. The key insights in the formal proof are the fact that $F^{\mu} = F^{\tilde{\mu}}$ and for $(\varphi_t^1, \varphi_t^2)$ a solution of (1.6), also $(\varphi_{-t}^1, -\varphi_{-t}^2)$ is a solution by the chain rule. Similar arguments are applicable for the differential case, where the substitution $(q,p,t) \to (q,-p,-t)$ gives a total minus sign in (1.11) and therefore does not change the property of being a solution. A more sophisticated proof of the ideas sketched here might be a good exercise at the end of the this section, when we will finally have introduced all the precise definitions. Anyways, the moral of this short digression is that it suffices to concentrate on forward time evolution in this work. Hopefully, the following formal definition is by now sufficiently motivated:

Definition 2.1. *We say that $\mu : [0,T] \to (\mathcal{P}(\mathbb{R}^6))^n$ is a solution to the Vlasov equation in integral form for bounded Lipschitz pair interactions $f_{k,l} \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ if $\mu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ and*

$$\mu_{k,t} = \mu_{k,0} \circ \left(\varphi_{k,t}^{\mu}\right)^{-1} \qquad \forall k \in [n],\ t \in [0,T], \tag{2.1}$$

where $\varphi_{k,t}^{\mu}$ is the unique, global flow to the ODE

$$\dot{q}(t) = \frac{p(t)}{m_k}, \qquad \dot{p}(t) = F_k^{\mu}(q(t),t) = \sum_{l\in[n]} \alpha_k \cdot (f_{k,l} *_q \mu_{l,t})(q(t)). \tag{2.2}$$

We will often write (2.1) in an aggregate way, namely via $\mu_t = \mu_0 \circ (\varphi_t^{\mu})^{-1}$ where φ_t^{μ} is the flow corresponding to the force field F^{μ}, to be interpreted component-wise.

Remark 2.2. Lemma 2.9, together with our results in chapter 4.8, shows that under the hypotheses $\mu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ and $f_{k,l} \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$, existence and uniqueness of global flows $\varphi_{k,t}^{\mu}$ for (2.2) are secured, so everything is indeed well-defined.

Let us first check that as announced in the motivational part, for a large class of pair interactions $f_{k,l}$, the empirical probability measure corresponding to a solution of the weakly coupled system 1.4 is a solution of the Vlasov equation in integral form. This shows that in fact, the Vlasov equation in the sense of definition 2.1 can be regarded as a generalization of the Newtonian equations of motion to not necessarily discrete particle distributions.

Lemma 2.3. *Provided that $f_{k,l} \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ with $f_{k,l}(0) = 0$ for all $k,l \in [n]$, the empirical distribution $\mu_{\mathrm{emp}} := (\mu_{\mathrm{emp},1}, \dots, \mu_{\mathrm{emp},n})$ defined by* (1.2) *and corresponding to solutions of* (1.4) *is a solution of the Vlasov equation in integral form in the sense of definition 2.1.*

Proof. We fix an initial condition $Z = (Z_1, \dots, Z_n) \in \mathbb{R}^{6N}$ and therefore suppress Z in the notation of the empirical distribution from now on. Let us first prove that $\mu_{\mathrm{emp}} \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$. Only weak continuity is not obvious. However, from continuity (even continuous differentiability) of the map $t \mapsto (\Psi_{k,t}^{1,i}(Z), \Psi_{k,t}^{2,i}(Z))$, where Ψ_t denotes the unique flow for the weakly coupled system (1.4), we see that for every $k \in [n]$ and $g \in \mathrm{BL}(\mathbb{R}^6) \subset C(\mathbb{R}^6)$, the map

$$[0,T] \to \mathbb{R}, \qquad t \mapsto \int_{\mathbb{R}^6} g \,\mathrm{d}\mu_{\mathrm{emp},k,t} = \frac{1}{N_k} \sum_{i=1}^{N_k} g\big(\Psi_{k,t}^{1,i}(Z), \Psi_{k,t}^{2,i}(Z)\big)$$

is continuous. This already proves weak continuity of μ_{emp}.

Next, recall from (1.3) that for all $k \in [n]$, $q \in \mathbb{R}^3$ and $t \in [0,T]$,

$$\frac{1}{N} \sum_{l \in [n]} \sum_{j \in \Gamma_l} f_{k,l}\big(q - \Psi^{1,j}_{l,t}(Z)\big) = \sum_{l \in [n]} \alpha_l \cdot (f_{k,l} *_q \mu_{\mathrm{emp},l,t})(q) = F_k^{\mu_{\mathrm{emp}}}(q,t).$$

For $k \in [n]$, let $\psi_{k,t} = (\psi^1_{k,t}, \psi^2_{k,t})$ denote the unique global flow for the (non-autonomous) ODE

$$\dot{q}(t) = \frac{p(t)}{m_k}, \qquad \dot{p}(t) = F_k^{\mu_{\mathrm{emp}}}(q(t),t), \tag{$*$}$$

with initial time 0. Note that global existence and uniqueness is again provided by remark 2.2. Then for initial conditions from $Z_k = (Z_k^1, \ldots, Z_k^{N_k})$, $\psi_{k,t}$ coincides with the component of the trajectory for the corresponding particle, i.e. $\psi_{k,t}(Z_k^i) = (\Psi^{1,i}_{k,t}(Z), \Psi^{2,i}_{k,t}(Z))$ for all $i \in [N_k]$: Clearly, one has $\psi_{k,0}(Z_k^i) = Z_k^i = (\Psi^{1,i}_{k,0}(Z), \Psi^{2,i}_{k,0}(Z))$, and by $(*)$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\, \psi^1_{k,t}(Z_k^i) &= \frac{\psi^2_{k,t}(Z_k^i)}{m_k}, \\
\frac{\mathrm{d}}{\mathrm{d}t}\, \psi^2_{k,t}(Z_k^i) &= F_k^{\mu_{\mathrm{emp}}}\big(\psi^1_{k,t}(Z_k^i), t\big) = \frac{1}{N} \sum_{l \in [n]} \sum_{j \in \Gamma_l} f_{k,l}\big(\psi^1_{k,t}(Z_k^i) - \Psi^{1,j}_{l,t}(Z)\big).
\end{aligned}$$

It follows that $(\Psi^1_t(Z), \Psi^2_t(Z))$ still satisfies (1.4) if we replace the component $(\Psi^{1,i}_{k,t}(Z), \Psi^{2,i}_{k,t}(Z))$ by $(\psi^1_{k,t}(Z_k^i), \psi^2_{k,t}(Z_k^i))$. By uniqueness, it follows that $(\Psi^{1,i}_{k,t}(Z), \Psi^{2,i}_{k,t}(Z)) = (\psi^1_{k,t}(Z_k^i), \psi^2_{k,t}(Z_k^i))$. Consequently, for all $k \in [n]$, $t \in [0,T]$,

$$\mu_{\mathrm{emp},k,t} = \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{\left(\Psi^{1,i}_{k,t}(Z), \Psi^{2,i}_{k,t}(Z)\right)} = \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{\psi_{k,t}(Z_k^i)} = \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{Z_k^i} \circ \psi_{k,t}^{-1} = \mu_{\mathrm{emp},k,0} \circ \psi_{k,t}^{-1}.$$

This proves that μ_{emp} does indeed solve (2.1). □

Remark 2.4. For the case that all forces $f_{k,l}$ are radially symmetric, solutions of (2.1) conserve total energy: Under this assumption, $f_{k,l}$ admit potentials $V_{k,l}$, i.e. differentiable maps $V_{k,l} : \mathbb{R}^3 \to \mathbb{R}$ such that $-\nabla V_{k,l} = f_{k,l}$, which we can construct as follows: Let $f_{k,l}(q) = g_{k,l}(|q|) \cdot \frac{q}{|q|}$, then clearly $g_{k,l} : [0,\infty) \to \mathbb{R}$ is also bounded and Lipschitz continuous, and for $G_{k,l}$ an indefinite integral of $g_{k,l}$, $-G_{k,l} \circ |\cdot|$ is a potential for $f_{k,l}$, as an easy computation with the chain rule shows. Let us define the *kinetic energy* $T(t)$ and the *potential energy* $V(t)$ by

$$\begin{aligned}
T(t) &:= \frac{1}{2} \sum_{k \in [n]} \alpha_k \cdot \frac{1}{2m_k} \int_{\mathbb{R}^6} p^2 \,\mathrm{d}\mu_{k,t}(q,p), \\
V(t) &:= \frac{1}{2} \sum_{k,l \in [n]} \alpha_k \alpha_l \int_{\mathbb{R}^6 \times \mathbb{R}^6} V_{k,l}(q - \tilde{q}) \,\mathrm{d}\mu_{k,t}(q,p) \,\mathrm{d}\mu_{l,t}(\tilde{q},\tilde{p}).
\end{aligned}$$

Then formally, using $\mu_{k,t} = \mu_{k,0} \circ (\varphi^\mu_{k,t})^{-1}$ and lemma 4.23 (integration w.r.t. the image measure),

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\, T(t) &= \sum_{k \in [n]} \alpha_k \cdot \frac{1}{2m_k} \int_{\mathbb{R}^6} \frac{\mathrm{d}}{\mathrm{d}t} \left(\varphi^{\mu,2}_{k,t}(q,p)\right)^2 \mathrm{d}\mu_{k,0}(q,p) \\
&= \sum_{k \in [n]} \alpha_k \cdot \frac{1}{m_k} \int_{\mathbb{R}^6} \varphi^{\mu,2}_{k,t}(q,p) \cdot F_k^\mu\big(\varphi^{\mu,1}_{k,t}(q,p), t\big) \,\mathrm{d}\mu_{k,0}(q,p) \\
&= \sum_{k \in [n]} \alpha_k \cdot \frac{1}{m_k} \int_{\mathbb{R}^6} p \cdot F_k^\mu(q,t) \,\mathrm{d}\mu_{k,t}(q,p).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} V(t) &= \frac{1}{2} \sum_{k,l\in[n]} \alpha_k\alpha_l \int_{\mathbb{R}^6\times\mathbb{R}^6} \frac{\mathrm{d}}{\mathrm{d}t} \Big(V_{k,l}\big(\varphi^{\mu,1}_{k,t}(q,p) - \varphi^{\mu,1}_{l,t}(\tilde{q},\tilde{p})\big)\Big) \,\mathrm{d}\mu_{k,0}(q,p)\,\mathrm{d}\mu_{l,0}(\tilde{q},\tilde{p}) \\
&= -\frac{1}{2} \sum_{k,l\in[n]} \alpha_k\alpha_l \int_{\mathbb{R}^6\times\mathbb{R}^6} f_{k,l}\big(\varphi^{\mu,1}_{k,t}(q,p) - \varphi^{\mu,1}_{l,t}(\tilde{q},\tilde{p})\big)\cdot \frac{\varphi^{\mu,2}_{k,t}(q,p)}{m_k} \,\mathrm{d}\mu_{k,0}(q,p)\,\mathrm{d}\mu_{l,0}(\tilde{q},\tilde{p}) \\
&\qquad + \frac{1}{2} \sum_{k,l\in[n]} \alpha_k\alpha_l \int_{\mathbb{R}^6\times\mathbb{R}^6} f_{k,l}\big(\varphi^{\mu,1}_{k,t}(q,p) - \varphi^{\mu,1}_{l,t}(\tilde{q},\tilde{p})\big)\cdot \frac{\varphi^{\mu,2}_{l,t}(\tilde{q},\tilde{p})}{m_l} \,\mathrm{d}\mu_{k,0}(q,p)\,\mathrm{d}\mu_{l,0}(\tilde{q},\tilde{p}) \\
&= -\sum_{k\in[n]} \alpha_k\alpha_l \int_{\mathbb{R}^6\times\mathbb{R}^6} f_{k,l}(q-\tilde{q})\cdot\frac{p}{m_k} \,\mathrm{d}\mu_{k,t}(q,p)\,\mathrm{d}\mu_{l,t}(\tilde{q},\tilde{p}) \\
&= -\sum_{k\in[n]} \alpha_k\cdot\frac{1}{m_k}\int_{\mathbb{R}^6} p\cdot\left[\sum_{l\in[n]}\alpha_l\cdot\int_{\mathbb{R}^6} f_{k,l}(q-\tilde{q})\,\mathrm{d}\mu_{l,t}(\tilde{q},\tilde{p})\right]\mathrm{d}\mu_{k,t}(q,p) \\
&= -\sum_{k\in[n]} \alpha_k\cdot\frac{1}{m_k}\int_{\mathbb{R}^6} p\cdot F^{\mu}_k(q,t)\,\mathrm{d}\mu_{k,t}(q,p).
\end{aligned}$$

We used that $-\nabla V_{k,l} = f_{k,l}$ and that $f_{k,l} = f_{l,k}$ is radially symmetric and therefore in particular antisymmetric, i.e. $f_{k,l}(q-\tilde{q}) = -f_{l,k}(\tilde{q}-q)$, in order to see that we may simultaneously interchange the roles of k, q, p and $l, \tilde{q}, \tilde{p}$ at the cost of a minus-sign in the third line of the calculation. It follows directly that $\frac{\mathrm{d}}{\mathrm{d}t}(T(t)+V(t)) = 0$ and thus $E(t) := T(t)+V(t) = T(0)+V(0) = \text{const}$.

However, it remains to justify some of the previous formal computations. Of course, we assume that the integrals occurring in $T(0)$ and $V(0)$ exist and are finite - otherwise the whole discussion would not make sense at all. We only sketch the relevant arguments briefly since we will not need the result in our further investigations. Some of the mentioned arguments will probably be much clearer after having studied section 2.2 because methods we use to derive the estimates there are in the same spirit. In particular, it will be clear that the constant C in the following paragraph can be chosen as $\frac{1}{m}+\|f\|_\infty$, to be explained later. From the ODEs which the flows $\varphi^{\mu}_{k,t}$ satisfy and uniform boundedness of all $f_{k,l}$ and therefore of all $F^{\mu}_k(\,\cdot\,,t)$ uniformly in t, by Grønwall's lemma, one can prove that $|\varphi^{\mu,2}_{k,t}(q,p)| \leq |p| + C|t|$ for all $(q,p)\in\mathbb{R}^6$. Consequently, using that $p\in\mathcal{L}^1(\mathbb{R}^6;\mathrm{d}\mu_{k,0})$ (which in turn follows from our assumption $p\in\mathcal{L}^2(\mathbb{R}^6;\mathrm{d}\mu_{k,0})$ by Hölder's inequality because $\mu_{k,0}$ is a finite measure), we can conclude that $\varphi^{\mu,2}_{k,t}\cdot F^{\mu}_k(\,\cdot\,,t)\in\mathcal{L}^1(\mathbb{R}^6;\mathrm{d}\mu_{k,0})$ for all $k\in[n]$, i.e. $T(t)$ is well-defined for all times. Now, a short computation, using the mean value theorem of differentiation, shows that we can bound difference quotients (with respect to t and for $|h|<1$) of the integrand $(\varphi^{\mu,2}_{k,t})^2$ by $2(|p|+C(|t|+1))\cdot C$ uniformly in (q,p), and hence by dominated convergence (constant maps are integrable w.r.t finite measures), interchanging integration and differentiation in the first step of the computation of $\frac{\mathrm{d}}{\mathrm{d}t}T(t)$ is justified. Similar arguments apply for $V(t)$. The crucial observation there is that since the $f_{k,l} = -\nabla V_{k,l}$ are radially symmetric and bounded, the potentials $V_{k,l}$ are by construction Lipschitz continuous with $\|V_{k,l}\|_{\mathrm{L}} \leq \|f_{k,l}\|_\infty < C$, and hence for all $(q,p),(\tilde{q},\tilde{p})\in\mathbb{R}^6$, $k,l\in[n]$,

$$\begin{aligned}
\big|V_{k,l}\big(\varphi^{\mu,1}_{k,t}(q,p) - \varphi^{\mu,1}_{l,t}(\tilde{q},\tilde{p})\big)\big| &\leq |V_{k,l}(q-\tilde{q})| + C\cdot\big(\big|\varphi^{\mu,1}_{k,t}(q,p)-q\big| + \big|\varphi^{\mu,1}_{l,t}(\tilde{q},\tilde{p})-\tilde{q}\big|\big) \\
&\leq |V_{k,l}(q-\tilde{q})| + 2C\cdot C(|pt| + Ct^2).
\end{aligned}$$

In the last step, we used a Grønwall type estimate for $\varphi^{\mu,1}_{k,t}$, which one directly gets from the bound on $|\varphi^{\mu,2}_{k,t}|$. Existence of $V(0)$ and $T(0)$ then guarantee that $V(t)$ is well-defined for all times, and as for $T(t)$, with the mean value theorem one can find an integrable (constant) majorant for difference quotients of the integrand w.r.t. time. Finally, the application of Fubini's theorem is justified because we have already argued that all (double) integrals exist.

Now, let us turn to the definition of a solution of the Vlasov equation in differential form. Recapping the heuristics in section 1.2, every solution of the Vlasov equation in differential form should induce a solution of the Vlasov equation in integral form for bounded Lipschitz pair interactions. We try with the probably most intuitive approach:

Definition 2.5. *A* ***classical solution*** *of the Vlasov equation in differential form with bounded Lipschitz pair interactions is a map $u \in \mathcal{C}(\mathbb{R}^3 \times \mathbb{R}^3 \times [0,T]; ([0,\infty))^n) \cap \mathcal{C}^1(\mathbb{R}^3 \times \mathbb{R}^3 \times (0,T); ([0,\infty))^n)$ such that $u_k(\,\cdot\,,t) \in L^1(\mathbb{R}^3 \times \mathbb{R}^3)$ with $\|u_k(\,\cdot\,,t)\|_1 = 1$ for all $k \in [n]$, $t \in [0,T]$ and the system of coupled PDEs*

$$\partial_t u_k + \tfrac{p}{m_k} \cdot \nabla_q u_k + F_k^u \cdot \nabla_p u_k = 0, \qquad F_k^u = \sum_{l \in [n]} \alpha_l \cdot (f_{k,l} *_q u_{l,t}), \qquad k \in [n] \tag{2.3}$$

is satisfied.

Lemma 2.6. *For bounded Lipschitz interaction forces $f_{k,l}$, every classical solution of the Vlasov equation in differential form is also a solution of the Vlasov equation in integral form in the following sense: For $u = (u_1, \ldots, u_n)$ a solution of (2.3) and $\mu = (\mu_1, \ldots, \mu_n) : [0,T] \to (\mathcal{P}(\mathbb{R}^6))^n$ where $\mu_{k,t}$ is the probability measure with probability density $u_k(\,\cdot\,,t)$ w.r.t. Lebesgue measure for all $k \in [n]$, $t \in [0,T]$, μ solves (2.1)*

Proof. We first claim that μ is weakly continuous. Indeed, let $k \in [n]$, $g \in \mathrm{BL}(\mathbb{R}^6)$ and $t \in [0,T]$, then for every $x \in \mathbb{R}^6$ and $(t_n)_{n\in\mathbb{N}} \subset [0,T]$ with $t_n \xrightarrow{n\to\infty} t$, by continuity of u_k, it holds that $\lim_{n\to\infty} u_k(x,t_n) = u_k(x,t)$. Denoting $h_n := u_k(\,\cdot\,,t_n)$ and $h := u_k(\,\cdot\,,t)$, we see that $h, h_n \in L^1(\mathbb{R}^6)$ with $\|h_n\| = 1 = \|h\|$ for all $n \in \mathbb{N}$, and $h_n \xrightarrow{n\to\infty} h$ pointwise everywhere in $\mathbb{R}^6$. By theorem 1.9 in [29] with $p = 1$ (or a direct proof using Fatou's lemma), we obtain that $h_n \to h$ in $L^1(\mathbb{R}^6)$, i.e. $\lim_{n\to\infty} \int_{\mathbb{R}^6} |h_n - h| \,\mathrm{d}x = 0$. Consequently,

$$\begin{aligned}\lim_{n\to\infty} \left| \int_{\mathbb{R}^6} g \,\mathrm{d}\mu_{k,t_n} - \int_{\mathbb{R}^6} g \,\mathrm{d}\mu_{k,t} \right| &\leq \lim_{n\to\infty} \int_{\mathbb{R}^6} \big| g(x) \cdot u_k(x,t_n) - g(x) \cdot u_k(x,t) \big| \,\mathrm{d}x \\ &\leq \lim_{n\to\infty} \|g\|_\infty \cdot \int_{\mathbb{R}^6} \big| h_n(x) - h(x) \big| \,\mathrm{d}x = 0.\end{aligned}$$

Hence, $\lim_{n\to\infty} \int_{\mathbb{R}^6} g \,\mathrm{d}\mu_{k,t_n} = \int_{\mathbb{R}^6} g \,\mathrm{d}\mu_{k,t}$, which proves weak continuity of μ. By lemma 2.9, the corresponding mean field forces F_k^u are continuous and uniformly in t Lipschitz continuous in q. From section 4.8, we thus obtain that (2.2) does indeed have unique, global flows $\varphi_{k,t}^u$ for all $k \in [n]$ which are also homeomorphisms for fixed $t \in [0,T]$. Moreover, section 4.9 shows that these $\varphi_{k,t}^u$ are measure preserving w.r.t 6-dimensional Lebesgue measure. Using the chain rule and the fact that $\varphi_{k,t}^u$ is the flow for (1.7), for $(q,p) \in \mathbb{R}^6$ and $t \in (0,T)$, with the short notation $u_{k,t}(q,p) := u_k(q,p,t)$, we obtain

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t} (u_{k,t} \circ \varphi_{k,t}^u)(q,p) = \frac{\mathrm{d}}{\mathrm{d}t} u_k(\varphi_{k,t}^u(q,p),t) \\ &= \partial_t u_k(\varphi_{k,t}^u(q,p),t) + \frac{\varphi_{k,t}^{u,2}(q,p)}{m_k} \cdot \nabla_q u_k(\varphi_{k,t}^u(q,p),t) + F_k^u(\varphi_{k,t}^{u,1}(q,p),t) \cdot \nabla_q u_k(\varphi_{k,t}^u(q,p),t) = 0,\end{aligned}$$

where in the last step we employed (2.3). Consequently,

$$(u_{k,t} \circ \varphi^u_{k,t})(q,p) = (u_{k,0} \circ \varphi^u_{k,0})(q,p) = u_{k,0}(q,p) \qquad \forall t \in [0,T].$$

Hence, for every $k \in [n]$ and $A \in \mathcal{B}(\mathbb{R}^6)$, using that $\varphi^u_{k,t}$ is bijective and measure preserving and thus applying (4.9),

$$\begin{aligned} \mu_{k,0}(A) &= \int_{\mathbb{R}^6} \mathbb{1}_A \cdot u_{k,0} \, \mathrm{d}x = \int_{\mathbb{R}^6} \left(\mathbb{1}_{\varphi^u_{k,t}(A)} \circ \varphi^u_{k,t} \right) \cdot \left(u_{k,t} \circ \varphi^u_{k,t} \right) \, \mathrm{d}x = \int_{\mathbb{R}^6} \mathbb{1}_{\varphi^u_{k,t}(A)} \cdot u_{k,t} \, \mathrm{d}x \\ &= \mu_{k,t} \left(\varphi^u_{k,t}(A) \right). \end{aligned}$$

Since $\varphi^u_{k,t}$ is invertible with continuous and therefore in particular measurable inverse, we conclude that $\mu_{k,0}((\varphi^u_{k,t})^{-1}(A)) = \mu_{k,t}(A)$, i.e. $\mu_{k,t} = \mu_{k,0} \circ (\varphi^u_{k,t})^{-1}$. This proves that μ is a solution of the Vlasov equation in integral form, as claimed. □

Remark 2.7. One might ask under which conditions a solution μ of the Vlasov equation in integral form yields a solution to the Vlasov equation in differential form. Imagine that $\mu_{k,0}$ has densities $u_{k,0} \in \mathcal{C}^1(\mathbb{R}^6)$ for all $k \in [n]$ and that $f_{k,l} \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3) \cap \mathcal{C}^1(\mathbb{R}^3;\mathbb{R}^3)$. Then for every $t \in [0,T]$, $F^u_k \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3) \cap \mathcal{C}^1(\mathbb{R}^3;\mathbb{R}^3)$, as an easy computation in the spirit of lemma 2.9 using the mean value theorem of differentiation and the dominated convergence theorem (the derivatives of the $f_{k,l}$ are uniformly bounded by the Lipschitz constant for the $f_{k,l}$) shows. By standard ODE theory, it follows that $\varphi^u_{k,t}, (\varphi^u_{k,t})^{-1} = \varphi^u_{k,-t} \in \mathcal{C}^1(\mathbb{R}^6;\mathbb{R}^6)$ for all $t \in [0,T]$. Hence, continuity of $u_k(q,p,t)$ is obvious, and the chain rule shows that the densities $u_k(q,p,t) = (u_{k,0} \circ (\varphi^u_{k,t})^{-1}))(q,p)$ are in fact continuously partially differentiable on $\mathbb{R}^3 \times \mathbb{R}^3 \times (0,T)$. By the computation (1.10) and the arguments thereafter, we may conclude that $u := (u_1, \dots, u_n)$ is a classical solution of the Vlasov equation in differential form.

Finally, let us briefly mention how to proceed in the case where all $f_{k,l}$ are proportional to the Coulomb force, i.e. $f_{k,l}(q) = c_k c_l \cdot k(q)$ where c_k, c_l are the *coupling constants* (masses, charges) and $k(q) = \pm q \cdot |q|^{-3}$. Looking at (1.1), we see that for a gravitational system, we need the minus sign, whereas for the electrostatic case, we need the plus sign. In the literature, for $n = 1$, this system is then usually referred to as **Vlasov-Poisson system**, and we will adapt this term also for $n > 1$ in the above-described scenario. This time, we cannot hope to find a good notion for the Vlasov equation in integral form which includes probability measures with a discrete part because the Coulomb force is well-defined only almost everywhere, which means that we cannot make sense of integrating k w.r.t. a probability measure which has a discrete part in case the *atom*, i.e. the support of the δ-measure, sits in the singularity at some time. This is of course closely related to the fact that there are initial conditions for which (1.1) resp. (1.4) do not have (global) solutions for the Coulomb interaction case. Consequently, it suffices to define solutions for the differential form. Unfortunately, by contrast to what we saw for bounded Lipschitz forces, it is not clear that the force fields F^u induced by formal solutions of (1.11) give rise to unique, global flows φ^u_t (more details on this will be given in section 2.3). Hence, it is reasonable to take care that a solution u will also admit a reasonable flow φ^u_t explicitly in the definition. However, there are quite a few different approaches in the literature, which are also all formulated for a single type of particles only. Thus, some research concerning the generalization to multiple particle types and the relationship between the typical definitions remains to be done. Hopefully, the discussion on the bounded Lipschitz case, together with the generalization of Liouville's theorem given in section 4.9 prepare for a rigorous and yet comprehensive discussion. In the eyes of the author of this work, by far the best reference for the Vlasov-Poisson system is [33]. Some further remarks concerning the Vlasov-Poisson system will be made in section 2.3.

2.2 Existence and uniqueness: the bounded Lipschitz case

In this section, we want to prove that in case all interaction forces are bounded Lipschitz functions, given any initial distribution $\mu_0 \in (\mathcal{P}(\mathbb{R}^6))^n$, there is a unique, global solution μ for the Vlasov equation in integral form (2.1), i.e. we can say that the IVP is well-posed. We are thus looking for $\mu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ satisfying $\mu_t = \mu_0 \circ (\varphi_t^\mu)^{-1}$. We can regard a solution to this IVP as fixed point of a map G which takes a weakly continuous curve ν and maps it to the curve $t \mapsto \nu_0 \circ (\varphi_t^\nu)^{-1}$. Hence, one is immediately tempted to check whether the Banach Fixed Point Theorem is applicable, since this would yield existence and uniqueness of solutions immediately. However, we have to choose a metric wisely in order to make the space $\mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ complete. Indeed, in section 4.4, we introduce the bounded Lipschitz distance d_{BL} as a metric on spaces of finite measures and derive a quite natural family of metrics $\overline{d}_{\mathrm{BL}}$ on $\mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ which make $\left(\mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n), \overline{d}_{\mathrm{BL}}\right)$ a complete metric space. Hence, our first big goal is to prove that the restriction of G to a suitable, closed subspace is in fact a contraction for an appropriate choice $\overline{d}_{\mathrm{BL}}$ in this family. On the other hand, the aim of this work is to prove that solutions of the Vlasov equation with initial probabilities close to the initial distribution of particles stay good approximations to the empirical distribution for finite times. Since the latter has turned out to be a solution of (2.1) as well for a large class of pair interactions (see lemma 2.3), we want to show that if two solutions of the Vlasov equation are close initially, then also their time evolutions do not move apart too fast. This will be the second important result of this section. The motivation for both main statements and their proofs stems from chapter 5 in [37]. However, besides extending the result to multiple particle types, the treatment here is more systematic and detailed.

Remark 2.8. From the discussion in remark 2.7, we know that for *nice* initial probability densities and pair interaction forces, a solution of the Vlasov equation in integral form yields a classical solution of the Vlasov equation in differential form. On the other hand, lemma 2.6 shows that every classical solution of the Vlasov equation in differential form yields a solution to the Vlasov equation in integral form. Thus, proving that a unique solution of the Vlasov equation in integral form does always exist also proves that for nice initial conditions and pair interactions, there is a unique classical solution for the Vlasov equation in differential form given by the time-dependent probability density of the corresponding solution of the Vlasov equation in integral form. In particular, all our results for solutions of the Vlasov equation in integral form apply to the solution of the Vlasov equation in differential form as well.

In the remainder of this section, unless mentioned otherwise, $|\cdot|$ will denote the maximum norm on $\mathbb{R}^d$. Moreover, we are going to use the following shorthand notations: $f \in \mathrm{BL}(\mathbb{R}^3;(\mathbb{R}^3)^{n^2})$ means that $f_{k,l} \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ for all $k,l \in [n]$, and we will write

$$\begin{aligned}
\|f\|_\infty &:= \max\left\{\|f_{k,l}\|_\infty : k,l \in [n]\right\},\\
\|f\|_{\mathrm{L}} &:= \max\left\{\|f_{k,l}\|_{\mathrm{L}} : k,l \in [n]\right\},\\
\|f\|_{\mathrm{BL}} &:= \max\left\{\|f\|_\infty, \|f\|_{\mathrm{L}}\right\}.
\end{aligned}$$

Note that this is coherent with the picture of regarding $f := (f_{1,1},\dots,f_{n,n})$ as a map $\mathbb{R}^3 \to (\mathbb{R}^3)^{n^2}$ and computing $\|f\|_{\infty,\mathrm{L},\mathrm{BL}}$ in the usual sense resp. in the sense of definition 4.36 when using the maximum norm on product spaces. Likewise, we denote $\varphi_t^\mu := \left(\varphi_{1,t}^\mu,\dots,\varphi_{n,t}^\mu\right) : \mathbb{R}^6 \to (\mathbb{R}^6)^n$ and $F^\mu := (F_1^\mu,\dots,F_n^\mu) : \mathbb{R}^3 \times [0,T] \to (\mathbb{R}^3)^n$, with $\|\varphi_t^\mu\|_{\infty,\mathrm{L},\mathrm{BL}}$ and $\|F^\mu(\,\cdot\,,t)\|_{\infty,\mathrm{L},\mathrm{BL}}$ being defined in the same spirit. Last but not least, we introduce $m := \min\{m_k : k \in [n]\}$, the mass of the *lightest* particles.

Let us mention that one can readily check that in all of the following arguments, it is not important that the physical dimension equals 3, i.e actually, the proof works for arbitrary dimensions d of the underlying physical space, by contrast to the Vlasov-Poisson system, for which the relationship between the spatial dimension and the order of the singularity is crucial.

We first collect some estimates on the external forces F_k^μ coming from a particle distribution $\mu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$. This, in turns, gives us estimates on the induced flows φ_t^μ. In what follows, d_{BL} denotes the *bounded Lipschitz distance* between (vector valued) measures. A short introduction to d_{BL} with a digression on its most important elementary properties is given in section 4.4.

Lemma 2.9. *Let* $\mu, \nu \in \mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big)$. *Then*

(i) for all $t \in [0,T]$, $F^\mu(\,\cdot\,,t) \in \mathrm{BL}\big(\mathbb{R}^3;(\mathbb{R}^3)^n\big)$ *with* $\|F^\mu(\,\cdot\,,t)\|_{\mathrm{BL}} \le \|f\|_{\mathrm{BL}}$. *More precisely,* $\|F^\mu(\,\cdot\,,t)\|_\infty \le \|f\|_\infty$ *and* $\|F^\mu(\,\cdot\,,t)\|_{\mathrm{L}} \le \|f\|_{\mathrm{L}}$ *for all* $t \in [0,T]$.

(ii) $F^\mu \in C_b(\mathbb{R}^3 \times [0,T];(\mathbb{R}^3)^n)$.

(iii) for all $t \in [0,T]$, $\|F^\mu(\,\cdot\,,t) - F^\nu(\,\cdot\,,t)\|_\infty \le \|f\|_{\mathrm{BL}} \cdot \mathrm{d}_{\mathrm{BL}}(\mu_t,\nu_t)$.

Proof. We will often use without explicit mentioning that $\mu_{k,t}, \nu_{k,t}$ are probability measures on $\mathbb{R}^6$ for all $k \in [n]$, $t \in [0,T]$. Moreover, note that for all three claims it suffices to prove the statement component-wise. For (i), we compute for $t \in [0,T]$, $k \in [n]$ and $q, q' \in \mathbb{R}^3$

$$\big|F_k^\mu(q,t)\big| \le \sum_{l=1}^n \alpha_l \cdot \int_{\mathbb{R}^6} \big|f_{k,l}(q-\tilde{q})\big| \,\mathrm{d}\mu_{l,t}(\tilde{q}) \le \sum_{l=1}^n \alpha_l \cdot \int_{\mathbb{R}^6} \|f_{k,l}\|_\infty \,\mathrm{d}\mu_{l,t} \le \|f\|_\infty \cdot \sum_{l=1}^n \alpha_l = \|f\|_\infty\,,$$

where we used $\sum_{l=1}^n \alpha_l = 1$, and similarly

$$\begin{aligned}\big|F_k^\mu(q,t) - F_k^\mu(q',t)\big| &\le \sum_{l=1}^n \alpha_l \cdot \int_{\mathbb{R}^6} \big|f_{k,l}(q-\tilde{q}) - f_{k,l}(q'-\tilde{q})\big| \,\mathrm{d}\mu_{l,t}(\tilde{q},\tilde{p}) \\ &\le \sum_{l=1}^n \alpha_l \cdot \int_{\mathbb{R}^6} \|f\|_{\mathrm{L}} \cdot |(q-\tilde{q}) - (q'-\tilde{q})| \,\mathrm{d}\mu_{l,t}(\tilde{q},\tilde{p}) \\ &\le \|f\|_{\mathrm{L}} \cdot |q-q'|\,.\end{aligned}$$

For (ii), observe that continuity of F^μ in the first component $q \in \mathbb{R}^3$ is clear by (i). That F^μ is also continuous w.r.t. t holds because all μ_k are weakly continuous: It is easy to check that for $g \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ and $q \in \mathbb{R}^3$, the map $g_q : \mathbb{R}^6 \to \mathbb{R}^3$, $(\tilde{q},\tilde{p}) \mapsto g(q-\tilde{q})$ satisfies $g_q \in \mathrm{BL}(\mathbb{R}^6;\mathbb{R}^3)$ with $\|g_q\|_{\mathrm{BL}} = \|g\|_{\mathrm{BL}}$. Consequently, the maps $t \mapsto (f_{k,l} *_q \mu_{k,t})(q)$ are continuous, and hence $F_k^\mu(q,\,\cdot\,)$ as a finite sum of continuous maps, too. Finally,

$$\begin{aligned}\big|F_k^\mu(q,t) - F_k^\nu(q,t)\big| &= \left|\sum_{l=1}^n \alpha_l \cdot \int_{\mathbb{R}^6} f_{k,l}(q-\tilde{q})\, \big(\mathrm{d}\mu_{l,t}(\tilde{q},\tilde{p}) - \mathrm{d}\nu_{l,t}(\tilde{q},\tilde{p})\big)\right| \\ &\le \sum_{l=1}^n \alpha_l \cdot \big(\|f_{k,l}\|_{\mathrm{BL}} \cdot d_{\mathrm{BL}}(\mu_{l,t},\nu_{l,t})\big) \\ &= \|f\|_{\mathrm{BL}} \cdot d_{\mathrm{BL}}(\mu_t,\nu_t).\end{aligned}$$

□

Lemma 2.10. *Under the same hypotheses, φ_t^μ is Lipschitz continuous with*

$$\|\varphi_t^\mu\|_{\mathrm{L}} \leq e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)t} \qquad \forall t \in [0,T].$$

Proof. For $t \in (0,T)$, $x, y \in \mathbb{R}^6$ and $k \in [n]$, using $\|F_k^\nu(\,\cdot\,,t)\|_{\mathrm{L}} \leq \|f\|_{\mathrm{L}}$ by lemma 2.9 (i) and applying corollary 4.19 to the $\mathcal{C}^1$-curve $\gamma(t) := \varphi_{k,t}^\mu(x) - \varphi_{k,t}^\mu(y)$, we obtain

$$\begin{aligned}
\partial_t^+\left|\varphi_{k,t}^\mu(x) - \varphi_{k,t}^\mu(y)\right| &\leq \left|\partial_t\big(\varphi_{k,t}^{\mu,1}(x), \varphi_{k,t}^{\mu,2}(x)\big) - \partial_t\big(\varphi_{k,t}^{\mu,1}(y), \varphi_{k,t}^{\mu,2}(y)\big)\right| \\
&= \max\left\{\left|\tfrac{1}{m_k}\,\varphi_{k,t}^{\mu,2}(x) - \tfrac{1}{m_k}\,\varphi_{k,t}^{\mu,2}(y)\right|, \left|F_k^\mu\big(\varphi_{k,t}^{\mu,1}(x),t\big) - F_k^\mu\big(\varphi_{k,t}^{\mu,1}(y),t\big)\right|\right\} \\
&\leq \tfrac{1}{m}\left|\varphi_{k,t}^{2,\mu}(x) - \varphi_{k,t}^{2,\mu}(y)\right| + \left|F_k^\mu\big(\varphi_{k,t}^{1,\mu}(x),t\big) - F_k^\mu\big(\varphi_{k,t}^{1,\mu}(y),t\big)\right| \\
&\leq \tfrac{1}{m}\left|\varphi_{k,t}^{\mu}(x) - \varphi_{k,t}^{\mu}(y)\right| + \|f\|_{\mathrm{L}}\cdot\left|\varphi_{k,t}^{1,\mu}(x) - \varphi_{k,t}^{1,\mu}(y)\right| \\
&\leq \big(\tfrac{1}{m} + \|f\|_{\mathrm{L}}\big)\cdot\left|\varphi_{k,t}^\mu(x) - \varphi_{k,t}^\mu(y)\right|.
\end{aligned}$$

Now, theorem 4.3 (Grønwall's lemma) with $\varphi_{k,0}^\mu = \mathrm{id}_{\mathbb{R}^6}$ and thus $|\varphi_{k,0}^\mu(x) - \varphi_{k,0}^\mu(y)| = |x-y|$ yields

$$\left|\varphi_{k,t}^\mu(x) - \varphi_{k,t}^\mu(y)\right| \leq e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)t}\cdot|x-y| \qquad \forall t\in[0,T].$$

□

Lemma 2.11. *Again under the hypotheses of lemma 2.9,*

$$\|\varphi_t^\mu - \varphi_t^\nu\|_\infty \leq \|f\|_{\mathrm{BL}}\cdot\int_0^t e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)(t-s)}\cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s \quad \forall t\in[0,T].$$

Proof. Let $k \in [n]$. Another time using corollary 4.19, for $t \in (0,T)$ and $x \in \mathbb{R}^6$,

$$\begin{aligned}
\partial_t^+\left|\varphi_{k,t}^\mu(x) - \varphi_{k,t}^\nu(x)\right| &\leq \tfrac{1}{m}\left|\varphi_{k,t}^{2,\mu}(x) - \varphi_{k,t}^{2,\nu}(x)\right| + \left|F_k^\mu\big(\varphi_{k,t}^{1,\mu}(x),t\big) - F_k^\nu\big(\varphi_{k,t}^{1,\nu}(x),t\big)\right| \\
&\leq \tfrac{1}{m}\left|\varphi_{k,t}^\mu(x) - \varphi_{k,t}^\nu(x)\right| + (\mathrm{I}).
\end{aligned}$$

With the triangle inequality and lemma 2.9 (i), (iii),

$$\begin{aligned}
(\mathrm{I}) &\leq \left|F_k^\mu\big(\varphi_{k,t}^{1,\mu}(x),t\big) - F_k^\mu\big(\varphi_{k,t}^{1,\nu}(x),t\big)\right| + \left|F_k^\mu\big(\varphi_{k,t}^{1,\nu}(x),t\big) - F_k^\nu\big(\varphi_{k,t}^{1,\nu}(x),t\big)\right| \\
&\leq \|f\|_{\mathrm{L}}\cdot\left|\varphi_{k,t}^{1,\mu}(x) - \varphi_{k,t}^{1,\nu}(x)\right| + \|f\|_{\mathrm{BL}}\cdot d_{\mathrm{BL}}(\mu_t,\nu_t) \\
&\leq \|f\|_{\mathrm{L}}\cdot|\varphi_t^\mu(x) - \varphi_t^\nu(x)| + \|f\|_{\mathrm{BL}}\cdot d_{\mathrm{BL}}(\mu_t,\nu_t).
\end{aligned}$$

Taking the maximum over all components and using lemma 4.15, we arrive at

$$\partial_t^+\left|\varphi_t^\mu(x) - \varphi_t^\nu(x)\right| \leq \big(\tfrac{1}{m} + \|f\|_{\mathrm{L}}\big)\cdot\left|\varphi_t^\mu(x) - \varphi_t^\nu(x)\right| + \|f\|_{\mathrm{BL}}\cdot d_{\mathrm{BL}}(\mu_t,\nu_t). \tag{2.4}$$

Since the l.h.s. is right-continuous (lemma 4.18) and the r.h.s. is measurable and bounded (lemma 4.44), we may integrate on both sides and use our fundamental theorem of calculus for right-continuous maps (lemma 4.14). With $\left|\varphi_0^\mu(x) - \varphi_0^\nu(x)\right| = 0$, we obtain

$$\left|\varphi_t^\mu(x) - \varphi_t^\nu(x)\right| \leq \big(\tfrac{1}{m} + \|f\|_{\mathrm{L}}\big)\cdot\int_0^t\left|\varphi_s^\mu(x) - \varphi_s^\nu(x)\right|\mathrm{d}s + \|f\|_{\mathrm{BL}}\cdot\int_0^t d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s.$$

By Grønwall's lemma in integral form (theorem 4.80), it follows that for all $t \in [0,T]$,

$$\begin{aligned}
\left|\varphi_t^\mu(x) - \varphi_t^\nu(x)\right| &\leq e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)t}\cdot\int_0^t e^{-\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)s}\cdot\big(\|f\|_{\mathrm{BL}}\cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\big)\,\mathrm{d}s \\
&= \|f\|_{\mathrm{BL}}\cdot\int_0^t e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)(t-s)}\cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s.
\end{aligned}$$

□

We are now ready to introduce the announced map G.

Theorem 2.12. *Define*

$$G : \mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big) \to \mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big), \qquad \mu \mapsto \mu_0 \circ (\varphi_t^\mu)^{-1}.$$

Then for $\mu, \nu \in \mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big)$, *it holds that*

$$d_{\mathrm{BL}}\big((G[\mu])_t, (G[\nu])_t\big) \leq 2e^{Kt} \cdot \left[d_{\mathrm{BL}}(\mu_0,\nu_0) + \|f\|_{\mathrm{BL}} \cdot \int_0^t e^{-Ks} \cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s\right] \qquad \forall t \in [0,T], \tag{2.5}$$

where $K := \frac{1}{m} + \|f\|_{\mathrm{L}}$

Proof. Let $\mu,\nu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$, $h \in \mathrm{BL}(\mathbb{R}^6)$ with $\|h\|_{\mathrm{BL}} \leq 1$, and $t \in [0,T]$. Then for any $k \in [n]$, with lemma 4.23 and the previous lemmata 2.10 and 2.11, we obtain

$$\begin{aligned}
&\left|\int_{\mathbb{R}^6} h\,\big(\mathrm{d}(G[\mu])_{k,t} - \mathrm{d}(G[\nu])_{k,t}\big)\right| \\
&= \left|\int_{\mathbb{R}^6} h\,\mathrm{d}\big(\mu_{k,0}\circ(\varphi_{k,t}^\mu)^{-1}\big) - \int_{\mathbb{R}^6} h\,\mathrm{d}\big(\nu_{k,0}\circ(\varphi_{k,t}^\nu)^{-1}\big)\right| \\
&= \left|\int_{\mathbb{R}^6} h\circ\varphi_{k,t}^\mu\,\mathrm{d}\mu_{k,0} - \int_{\mathbb{R}^6} h\circ\varphi_{k,t}^\nu\,\mathrm{d}\nu_{k,0}\right| \\
&\leq \left|\int_{\mathbb{R}^6} \big(h\circ\varphi_{k,t}^\mu - h\circ\varphi_{k,t}^\nu\big)\,\mathrm{d}\mu_{k,0}\right| + \left|\int_{\mathbb{R}^6} (h\circ\varphi_{k,t}^\nu)\,(\mathrm{d}\mu_{k,0} - \mathrm{d}\nu_{k,0})\right| \\
&\leq \int_{\mathbb{R}^6} \left|\varphi_{k,t}^\mu(x) - \varphi_{k,t}^\nu(x)\right|\mathrm{d}\mu_{k,0}(x) + \left\|h\circ\varphi_{k,t}^\nu\right\|_{\mathrm{BL}} \cdot d_{\mathrm{BL}}(\mu_{k,0},\nu_{k,0}) \\
&\leq \|f\|_{\mathrm{BL}} \cdot \int_0^t e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)(t-s)} \cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s + \left(1 + e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)t}\right) \cdot d_{\mathrm{BL}}(\mu_0,\nu_0) \\
&\leq 2e^{Kt} \cdot \left[d_{\mathrm{BL}}(\mu_0,\nu_0) + \|f\|_{\mathrm{BL}} \cdot \int_0^t e^{-Ks} \cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s\right].
\end{aligned}$$

Note that we used our observation from remark 4.64 (d) to see that $\left\|h\circ\varphi_{k,t}^\nu\right\|_{\mathrm{L}} \leq 1 + e^{\left(\frac{1}{m}+\|f\|_{\mathrm{L}}\right)t}$. It follows that

$$d_{\mathrm{BL}}\big((G[\mu])_t, (G[\nu])_t\big) \leq 2e^{Kt} \cdot \left[d_{\mathrm{BL}}(\mu_0,\nu_0) + \|f\|_{\mathrm{BL}} \cdot \int_0^t e^{-Ks} \cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s\right] \qquad \forall t \in [0,T]$$

and hence the claim. □

One can now hope that one of the metrics $\overline{d}_{\mathrm{BL}}$ from section 4.4 will make G a contraction. However, despite the fact that it looks difficult to find such a metric, it is also not necessary in this generality. What we really need in order to obtain a solution of the Vlasov equation 2.1 is that for a fixed initial condition $\eta \in (\mathcal{P}(\mathbb{R}^6))^n$, $A_\eta := \big\{\nu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n) : \nu_0 = \eta\big\}$ is a closed subset of $\mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big)$ w.r.t. the chosen $\overline{d}_{\mathrm{BL}}$, and the restriction of G onto A_η is a contraction w.r.t. this metric. Observe that in this case, the summand $d_{\mathrm{BL}}(\mu_0,\nu_0)$ in (2.5) vanishes for $\mu,\nu \in A_\eta$. Let us for $\alpha > 0$ define the metric

$$\begin{aligned}
\overline{d}_{\mathrm{BL}}^{\alpha} : \mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big) \times \mathcal{C}^*\big([0,T];(\mathcal{P}(\mathbb{R}^6))^n\big) &\to [0,\infty), \\
(\mu,\nu) &\mapsto \sup\big\{e^{-\alpha t} \cdot d_{\mathrm{BL}}(\mu_t,\nu_t) : 0 \leq t \leq T\big\}.
\end{aligned}$$

Then for $\eta \in (\mathcal{P}(\mathbb{R}^6))^n$, $A_\eta \subset \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ is indeed closed w.r.t. $\overline{d}_{\mathrm{BL}}^{\alpha}$: for an arbitrary sequence $(\mu_n)_{n\in\mathbb{N}} \subset A_\eta$ with $\mu_n \xrightarrow{n\to\infty} \mu$ w.r.t. $\overline{d}_{\mathrm{BL}}^{\alpha}$, by definition of $\overline{d}_{\mathrm{BL}}^{\alpha}$, $\eta = \mu_{n,0} \xrightarrow{n\to\infty} \mu_0$ w.r.t. d_{BL}, so we see that $\mu_0 = \eta$ and thus $\mu \in A_\eta$.

Let $\mu, \nu \in A_\eta$. By definition of $\overline{d}_{\mathrm{BL}}^{\alpha}$, for every $s \in [0,T]$, $e^{-\alpha s} \cdot d_{\mathrm{BL}}(\mu_s, \nu_s) \leq \overline{d}_{\mathrm{BL}}^{\alpha}(\mu,\nu)$, so we obtain that for all $s \in [0,T]$, $d_{\mathrm{BL}}(\mu_s,\nu_s) \leq e^{\alpha s} \cdot \overline{d}_{\mathrm{BL}}^{\alpha}(\mu,\nu)$. Using (2.5), we compute for $t \in [0,T]$ and $\alpha - K > 0$

$$\begin{aligned}
e^{-\alpha t} \cdot d_{\mathrm{BL}}\big((G[\mu])_t, (G[\nu])_t\big) &\leq 2\|f\|_{\mathrm{BL}} \cdot e^{(K-\alpha)\,t} \cdot \int_0^t e^{-Ks} \cdot e^{\alpha s} \cdot \overline{d}_{\mathrm{BL}}^{\alpha}(\mu,\nu)\,\mathrm{d}s \\
&= \overline{d}_{\mathrm{BL}}^{\alpha}(\mu,\nu) \cdot 2\|f\|_{\mathrm{BL}} \cdot \int_0^t e^{(\alpha-K)(s-t)}\,\mathrm{d}s \\
&= \overline{d}_{\mathrm{BL}}^{\alpha}(\mu,\nu) \cdot \frac{2\|f\|_{\mathrm{BL}}}{\alpha - K}\left(1 - e^{-(\alpha-K)\,t}\right) \\
&\leq \frac{2\|f\|_{\mathrm{BL}}}{\alpha-K} \cdot \overline{d}_{\mathrm{BL}}^{\alpha}(\mu,\nu).
\end{aligned}$$

Consequently, for $\alpha > K$ big enough, $\frac{2\|f\|_{\mathrm{BL}}}{\alpha-K} < 1$ and therefore $G|_{A_\eta}$ is a contraction. We have therefore almost proved the following theorem:

Theorem 2.13. *For every $\eta \in (\mathcal{P}(\mathbb{R}^6))^n$, there is a unique solution μ of the Vlasov equation 2.1 satisfying $\mu_0 = \eta$.*

Proof. We just need to apply the Banach Fixed Point theorem (see e.g. [2, p. 350 f.]) to the contraction $G : A_\eta \to A_\eta$ with the induced metric $\overline{d}_{\mathrm{BL}}^{\alpha}|_{A_\eta}$, noting that A_η is complete w.r.t. $\overline{d}_{\mathrm{BL}}^{\alpha}$: by theorem 4.41, $(\mathcal{P}(\mathbb{R}^6), d_{\mathrm{BL}})$ is complete, and lemma 4.45 shows that $\mathcal{C}^*([0,T];\mathcal{P}(\mathbb{R}^6))$ is complete w.r.t. $\overline{d}_{\mathrm{BL}}^{\alpha}$. Consequently, $\big(\mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n), \overline{d}_{\mathrm{BL}}^{\alpha}\big)$ is complete as a product of complete metric spaces w.r.t. the product metric. But we have argued before that $A_\eta \subset \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ is closed w.r.t. $\overline{d}_{\mathrm{BL}}^{\alpha}$, so it is itself complete w.r.t. $\overline{d}_{\mathrm{BL}}^{\alpha}|_{A_\eta}$ as closed subset of a complete metric space. ☐

There is also another very important application of lemma 2.12, which we will heavily rely on when we prove propagation of chaos for the case of bounded Lipschitz interactions in chapter 3.1, namely the announced fact that solutions to the Vlasov equation (2.1) which are initially close do not move away from each other too quickly:

Theorem 2.14. *Let $\mu, \nu \in \mathcal{C}^*([0,T];(\mathcal{P}(\mathbb{R}^6))^n)$ be solutions of the Vlasov equation in the sense of definition 2.1. Then for all $t \in [0,T]$,*

$$d_{\mathrm{BL}}(\mu_t,\nu_t) \leq 2e^{Kt}\left(1 + 2\|f\|_{\mathrm{BL}} \cdot t e^{\frac{2\|f\|_{\mathrm{BL}}}{K} e^{Kt}}\right) \cdot d_{\mathrm{BL}}(\mu_0,\nu_0).$$

Proof. Since μ, ν are solutions of the Vlasov equation, by construction of the map G in theorem 2.14, $G[\mu] = \mu$ and $G[\nu] = \nu$, in particular, $(G[\mu])_t = \mu_t$ and $(G[\nu])_t = \nu_t$ for all $t \in [0,T]$. Consequently, lemma 2.12 yields

$$d_{\mathrm{BL}}\big(\mu_t,\nu_t\big) \leq 2e^{Kt} \cdot d_{\mathrm{BL}}(\mu_0,\nu_0) + 2\|f\|_{\mathrm{BL}} \cdot \int_0^t e^{K(t-s)} \cdot d_{\mathrm{BL}}(\mu_s,\nu_s)\,\mathrm{d}s \qquad \forall t \in [0,T].$$

By Grønwall's lemma in integral form (see theorem 4.80), we obtain that

$$\begin{aligned} d_{\mathrm{BL}}(\mu_t, \nu_t) &\leq 2e^{Kt} \cdot d_{\mathrm{BL}}(\mu_0, \nu_0) + \int_0^t 2e^{Ks} \cdot d_{\mathrm{BL}}(\mu_0, \nu_0) \cdot 2\|f\|_{\mathrm{BL}} \cdot e^{K(t-s)} \cdot e^{\int_s^t 2\|f\|_{\mathrm{BL}} e^{K(t-\tau)} \, \mathrm{d}\tau} \, \mathrm{d}s \\ &\leq \left(2e^{Kt} + 2e^{Kt} \cdot 2\|f\|_{\mathrm{BL}} \cdot t e^{\frac{2\|f\|_{\mathrm{BL}}}{K} e^{Kt}} \right) \cdot d_{\mathrm{BL}}(\mu_0, \nu_0), \end{aligned}$$

where we used that for $s \in [0, T]$,

$$\int_s^t e^{K(t-\tau)} \, \mathrm{d}\tau = -\frac{1}{K} \left[1 - e^{K(t-s)} \right] \leq \frac{1}{K} \, e^{K(t-s)}$$

and consequently

$$\int_0^t e^{\int_s^t 2\|f\|_{\mathrm{BL}} e^{K(t-\tau)} \, \mathrm{d}\tau} \, \mathrm{d}s \leq \int_0^t e^{\frac{2\|f\|_{\mathrm{BL}}}{K} e^{K(t-s)}} \, \mathrm{d}s \leq t \cdot e^{\frac{2\|f\|_{\mathrm{BL}}}{K} e^{Kt}}. \qquad \square$$

2.3 Existence and uniqueness: references for the Coulomb case

We have already briefly discussed at the end of section 2.1 that for k the Coulomb interaction force, one should only consider a differential form of the Vlasov equation. However, since the Coulomb force is neither bounded nor Lipschitz continuous, the *mean field force* $k *_q u_{l,t}$ generated by particles of type l might not be well-defined if the *spatial density* $\rho^u_{l,t} := \int_{\mathbb{R}^3} u_{l,t}(\,\cdot\,, p) \, \mathrm{d}p$ gets unbounded. Moreover, even for bounded spatial densities, $k *_q u_{l,t}$ is in general only bounded and log-Lip-continuous in q (see theorem 4.54). However, by our results from sections 4.8 and 4.9, provided that also continuity of all $k *_q \rho^u_{l,t}$ and thus of the mean field forces F^u_l w.r.t. time is given, this suffices to guarantee the existence of unique, global, measure-preserving flows $\varphi^u_{l,t}$, which are not only desirable from the physical point of view but also helped us a lot in finding connections between the different definitions of solutions to the Vlasov equation and proving existence and uniqueness of solutions for the case of bounded Lipschitz interactions in section 2.2. Hence, what one typically does is demand that a formal solution to the Vlasov equation in differential form (1.11) has additional properties that ensure existence of unique, global, measure preserving flows along which the components $u_{l,t}$ of a solution then turn out to be constant. Using a more elaborate version of the version of Fatou's lemma employed in the proof of lemma 2.6, it might be possible to deduce continuity of the F^u_l in time from boundedness of the $\rho^u_{l,t}$ and continuity of u_l in time, however, this idea requires a thorough analysis an will not pursued further here.

In the literature, there have been two main approaches to local (in time) existence and uniqueness of solutions of the Vlasov-Poisson system for one type of particles: The first approach, pursued mainly by Illner, Neunzert and Horst (see [24],[21],[22]), first regularizes the Coulomb force k by introducing cut-off forces f^ε with $f^\varepsilon \to k$ pointwise as $\varepsilon \to 0$. For reasonable regularizations, namely such that f^ε is bounded and Lipschitz continuous, by our results in section 2.2, there are unique, global solutions u^ε to the regularized Vlasov-Poisson system. Then, one shows that in case the ρ^{u^ε} (and thus the mean-field forces F^{u^ε}) remain uniformly bounded as $\varepsilon \to 0$, the u^ε converge pointwise to some limiting function v as $\varepsilon \to 0$, and v turns out to be a solution to the Vlasov-Poisson system in the sense that it is a formal solution of 1.11 (with $f_{k,l}$ replaced by $c_k c_l \, k$) which induces global, measure preserving flows. However, global-in-time a-priori bounds on the spatial density are difficult to obtain, so with these ideas only the existence of local-in-time solutions could be established. The other approach is via an iterative scheme, introduced in [7], which one might compare with the construction of the solution of first-order ODEs by the Euler method. However, only for spherically symmetric initial conditions, global existence could be proved in this way. In the early 1990s, by different methods, eventually proofs for global existence under quite general assumptions were achieved. The first one, developed by Pfaffelmoser ([32]) and refined by Schaeffer ([35]), is based on getting control on the growth of the support in the momentum component of the densities corresponding to local in time solutions v (and thus only works for initial distributions v_0 which have compact support in the momentum component), obtaining control on the forces F_t^v and thus allowing to extend local solutions v to global solutions (the thoughts are somehow similar to what one does in the theory of ODEs). A different path was taken by Lions and Perthame ([30]), who obtained control on moments of the spatial densities ρ_t^v corresponding to local in time solutions to the Vlasov-Poisson system and bounded the growth of $\|\rho_t^v\|_\infty$ by Sobolev theory; then local extension can be employed again. Finally, basing on the approach of Lions and Perthame, uniqueness of solutions under quite general conditions could be verified by Loeper, see [31]. A highly recommendable aggregation of this development which is also very comprehensive can be found in the already mentioned [33].

Going through the cited literature, it should be straightforward to generalize all the results to the case of multiple particle types. In the above-mentioned paper of Lions and Perthame ([30], page 417 in the journal), it is even mentioned explicitly that all arguments work out for "different species of particles". However, a rigorous, detailed proof would be highly desirable, in particular since their arguments are ingenious, but often lack traceability and detailedness. Since doing all this, in the ideal case also with comparing different definitions for solutions as we did in section 2.1, is not the scope of this thesis for length and time issues, we will be happy with just postulating that the results for one type of particles translate to the ones for multiple particle types, since we will need some of these in section 3.2. So, we will take for granted that if $u_{k,0} \in C_c^\infty(\mathbb{R}^3 \times \mathbb{R}^3; [0, \infty))$ for all $k \in [n]$, a unique, global solution v to the Vlasov-Poisson systems described before does exist, in particular, the spatial densities are uniformly bounded for finite times in the sense that on every compact time interval $[0, T]$, $\max\left\{\sup_{t\in[0,T]} \|\rho_{k,t}^v\|_\infty : k \in [n]\right\} < \infty$, and they induce continuous force fields (which are by theorem 4.54 uniformly in t log-Lip-continuous in q) and therefore unique, global flows $\varphi_{k,t}^v$ (which by our results in section 4.9 are measure preserving). We will also need that the regularized solutions induce uniformly bounded densities, i.e. $\sup_{\varepsilon\in(0,1)} \max\left\{\sup_{t\in[0,T]} \|\rho_{k,t}^{u^\varepsilon}\|_\infty : k \in [n]\right\} < \infty$ for the regularization of the Coulomb force that we introduce in section 3.2, for one type of particles this is also a consequence of [30], and is in fact easy to prove once the solution theory for the unregularized Vlasov-Poisson system for multiple particle types is done. Anyway, we are confident that a sophisticated analysis of the issue for multiple particle types will confirm the statements listed here.

3 Propagation of chaos

In this chapter, we want to prove that under appropriate assumptions and in physically reasonable senses, propagation of chaos holds for bounded Lipschitz and gravitational/electrostatic pair interactions. Unfortunately, for the latter case we need to cut the Coulomb force off at a certain radius around the origin, however, this cut-off radius may be shrinked to 0 as $S \to \infty$.

3.1 The bounded Lipschitz case

As the section header indicates, in this short paragraph we assume that the pair interactions $f_{k,l}$ occurring in the weakly coupled system (1.4) are radially symmetric and bounded Lipschitz, i.e. we make statements about solutions of the Vlasov equation in the sense of definition 2.1. Note that by lemma 2.6, we also get the conclusions of this section for solutions of the Vlasov equation in differential form in the sense of definition 2.5. Let us briefly mention that just like in section 2.2, the arguments we give here are actually independent of the (physical) dimension d, by strict contrast to the Coulomb case which we will tackle in the next section and where the relation between the order of the singularity and the dimension is highly relevant.

Let $\mu_0 := (\mu_{1,0}, \dots, \mu_{n,0}) \in (\mathcal{P}(\mathbb{R}^6))^n$ a collection of (initial) probability measures on phase space $\mathbb{R}^6$ and $(Z_k^i)_{i\in\mathbb{N}}$ i.i.d. random variables distributed to the law $\mu_{k,0}$ for every $k \in [n]$. Recall that this means there is some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a collection of random variables $Z_k^i : \Omega \to \mathbb{R}^6$ where $(k,i) \in [n]\times\mathbb{N}$ such that for every $k \in [n]$, $(Z_k^i)_{i\in\mathbb{N}}$ are independent with $\mathbb{P}\circ\big(Z_k^i\big)^{-1} = \mu_{k,0}$ for all $i \in \mathbb{N}$. From section 4.5, in particular corollary 4.47, we know that the empirical distributions

$$\mu^\omega_{\mathrm{emp},k,0} := \frac{1}{N_k}\sum_{i=1}^{N_k} \delta_{Z_k^i(\omega)}$$

then converge almost surely to $\mu_{k,0}$ in the bounded Lipschitz distance in the sense that

$$\mathbb{P}\left[\left\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^\omega_{\mathrm{emp},k,0}, \mu_{k,0}) \xrightarrow{N_k\to\infty} 0\right\}\right] = 1 \qquad \forall k \in [n].$$

Note that as $S \to \infty$, by definition of S, we have $N_k \to \infty$ for all $k \in [n]$. Consequently, from the definition of d_{BL} on $(\mathcal{P}(\mathbb{R}^6))^n$ (see definition 4.38 and remark 4.39 (d), with the notation $\mu^\omega_{\mathrm{emp},0} := (\mu^\omega_{\mathrm{emp},1,0}, \dots, \mu^\omega_{\mathrm{emp},n,0}) \in (\mathcal{P}(\mathbb{R}^6))^n$, we see that

$$\left\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^\omega_{\mathrm{emp},0}, \mu_0) \xrightarrow{S\to\infty} 0\right\} \supset \bigcap_{k\in[n]} \left\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^\omega_{\mathrm{emp},k,0}, \mu_{k,0}) \xrightarrow{N_k\to\infty} 0\right\}.$$

Since in a finite measure space and thus in a probability space, the intersection of countably many (and hence in particular of a finite number of) sets of full measure has again full measure (this is an easy consequence of σ-additivity by looking at complements and using De Morgan's laws), we obtain that

$$\mathbb{P}\left[\left\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^\omega_{\mathrm{emp},0}, \mu_0)\right\} \xrightarrow{S\to\infty} 0\right] = 1.$$

On the other hand, we have seen in theorem 2.14 that for ν, η solutions of the Vlasov equation in the sense of definition 2.1 and fixed $t \in [0, T]$

$$d_{\mathrm{BL}}(\nu_t, \eta_t) \leq h(t) \cdot d_{\mathrm{BL}}(\nu_0, \eta_0),$$

where $h : [0, T] \to [0, \infty)$ is a continuous (and non-decreasing) map. In particular,

$$\sup_{0 \leq t \leq T} d_{\mathrm{BL}}(\nu_t, \eta_t) \leq h(T) \cdot d_{\mathrm{BL}}(\nu_0, \eta_0) \to 0 \qquad \text{as} \qquad d_{\mathrm{BL}}(\nu_0, \eta_0) \to 0,$$

which shows that

$$\Big\{ \omega \in \Omega : \sup_{0 \leq t \leq T} d_{\mathrm{BL}}(\nu_t, \eta_t) \xrightarrow{S \to \infty} 0 \Big\} \supset \Big\{ \omega \in \Omega : d_{\mathrm{BL}}(\nu_0, \eta_0) \xrightarrow{S \to \infty} 0 \Big\}.$$

Additionally, we have already checked in lemma 2.3 that for radially symmetric pair interactions, the empirical measure $\mu^{\omega}_{\mathrm{emp}}$ defined by

$$\mu^{\omega}_{\mathrm{emp},k,t} := \frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{\left(\Psi^{1,i}_{k,t}(Z(\omega)), \Psi^{2,i}_{k,t}(Z(\omega))\right)},$$

where Ψ_t is the (unique) global flow for (1.4), is a solution to the Vlasov equation in the sense of definition 2.1 with initial condition $\mu^{\omega}_{\mathrm{emp},0}$ (note that $\Psi_0 = \mathrm{id}_{\mathbb{R}^{6N}}$). Combining these three results, we immediately see that for fixed $T > 0$ and μ a solution of the Vlasov equation in integral form (2.1) with initial condition μ_0, under the above-mentioned hypotheses,

$$\mathbb{P}\left[\Big\{ \omega \in \Omega : \sup_{0 \leq t \leq T} d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},t}, \mu_t) \xrightarrow{S \to \infty} 0 \Big\} \right] = 1. \tag{3.1}$$

Thus, we say that *propagation of chaos holds almost surely.* In particular, from the definition of d_{BL} (cf. also remark 4.39 (e) in section 4.5) we see that averages of bounded Lipschitz observables on phase space w.r.t. the true time evolution of an initial state chosen according to some probability measure μ_0 on phase space converge to the expectation value of the observable w.r.t. the solution of the Vlasov equation with initial state μ_0 almost surely as $S \to \infty$. Actually, we saw that solutions to the Vlasov equation stay close deterministically, so the only thing that can go wrong is that the initial conditions of particles are chosen in a *bad* way. In particular, it is not necessary that initial conditions are chosen independently when considering different species; independence is only important within one type of particles because this is necessary to guarantee almost sure convergence of the initial empirical measure to the initial probability distributions. We will see that this changes dramatically in the next chapter.

3.2 The Coulomb case with cut-off

This section is heavily motivated by and oriented at the approach in [28]. Practically all definitions and theorems (albeit adaptions have been made to the multiple species case) are taken from there, as are the ideas for and significant parts of the proofs. However, the extension to multiple particle types is not obvious, so in some places we need additional arguments and ideas. Moreover, it might have struck the reader's eye that in the mentioned paper, at some places not all details of the proofs are given, so in what follows we will provide a solid justification for those arguments whose proof is only sketched or omitted in the reference.

One would hope that the results from the previous section can somehow be extended to a system with Coulomb interactions. However, this case cannot be handled with the results from section 2.2 because the Coulomb kernel k it is not bounded Lipschitz (in fact, it is neither bounded nor Lipschitz continuous). Moreover, spending some thoughts on what might possibly go wrong, it turns out that almost sure convergence of the empirical measure to the solution of the Vlasov equation for finite times cannot hold, mainly for the following reason: Imagine a repulsive situation in the case $n = 1$ where the initial probability densitity u_0 is C_c^∞, but unfortunately the initial conditions for the empirical distribution $\mu_{\text{emp},0}$ are such that the particles accumulate in balls of very small radius, however, the distribution of the balls is such that the empirical distribution converges to the probability measure given by the smooth probability density u_0 ("macroscopically, the small balls cannot be spotted"). Then it is clear that as $S \to \infty$, the initial potential energy goes to ∞ as well, and the particles will move to infinity very quickly as time evolves. On the other hand, the solution to the Vlasov-Poisson equation without cut-off is quite regular, more precisely, as mentioned in section 2.3, the spatial densities and thus the interaction forces are bounded uniformly in time, and so is the situation for the regularized Vlasov-Poisson system. Hence, one cannot expect that the empirical distribution associated with the true time evolution to (1.4) with Coulomb interactions and the solution to the Vlasov-Poisson system are still close even after very short time. In other words, what makes the difference to the case with bounded Lipschitz pair interactions is that the true time evolution and the solution to the Vlasov equation can drift apart rapidly even if they are close initially, i.e. the **deterministic** part of the convergence gets lost.

Spending some more time on this issue, one gets the feeling that in general, the initial conditions under which the empirical measure and the measure induced by the solution to the Vlasov-Poisson system do not stay close even after very short time are **not** a subset of probability 0. However, the painted situation is at least a *conspiracy* because as $S \to \infty$, the probability for an *atypical* initial condition as described above goes to 0. Hence, we cannot expect a result of almost sure convergence such as in section 3.1, but rather convergence in probability, by which we mean that for every $\varepsilon > 0$,

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{0 \le t \le T} d_{\text{BL}}(\nu^\omega_{\text{emp},t}, \nu_t) > \varepsilon\right\}\right] \xrightarrow{S\to\infty} 0,$$

at least for nice initial densities, where $\nu^\omega_{\text{emp},t}$ denotes the empirical measure associated with a solution to (1.4) with Coulomb forces and ν_t denotes the probability measure whose associated probability density is the solution v to the Vlasov-Poisson system (and the sample for $\nu_{\text{emp},0}$ is chosen by the law ν_0, of course). However, there is still a problem, namely that there is no existence and uniqueness result for solutions of (1.4) with Coulomb interactions available, and consequently we cannot really write down the corresponding flow $\Psi_t(Z(\omega))$ resp. the empirical measures $\nu^\omega_{\text{emp},t}$. Moreover, the Coulomb singularity is very strong and thus difficult to deal with because if particles get very close, the error in the mean field approximation is huge. Consequently, it might be a good idea to make things at least a bit simpler for the start (and it seems that this is also the state of the art regarding the Coulomb case because results concerning propagation of chaos for the unmodified Coulomb case are apparently non-existent). Thus, we impose an S-dependent cut-off on the Coulomb force such that in the limit $S \to \infty$, the regularized Coulomb force converges to the true Coulomb force. More precisely, for $\delta > 0$ and $S \in \mathbb{N}$ we define

$$f : \mathbb{R}^3 \to \mathbb{R}^3, \qquad q \mapsto \begin{cases} \frac{q}{|q|^3}, & |q| > S^{-\delta}, \\ S^{3\delta} \cdot q, & \text{else.} \end{cases} \tag{3.2}$$

The cut-off is chosen in this way mainly because it is radially symmetric, bounded, continuous, and in fact Lipschitz continuous (see lemma 4.71 and the corollary thereafter) with minimal bounded Lipschitz norm in the sense that every other Lipschitz continuous, radially symmetric regularization $\tilde{f}$ of the Coulomb force satisfying $\tilde{f} = k$ on $B^c_{S^{-\delta}}(0)$ has the property that $\|f\|_\infty \leq \|\tilde{f}\|_\infty$, $\|f\|_{\mathrm{L}} \leq \|\tilde{f}\|_{\mathrm{L}}$. Note that between $0 \in \mathbb{R}^3$ and the sphere of radius $S^{-\delta}$, we just interpolated linearly, and one can easily see that $\|f\|_{\mathrm{L}} \leq C \cdot S^{3\delta}$, see also lemma 4.73. It is probably not a big surprise that the exact form of the cut-off will not be crucial for what follows, however, there are a few properties of the cut-off which are indeed important, which we generalize to a notion that we will make use of frequently in the sequel, also for other orders of the singularity:

Definition 3.1. *Let $d \in \mathbb{N}$, $\alpha, \delta > 0$, $S \in \mathbb{N}$. We say that a map $h : \mathbb{R}^d \to \mathbb{R}^d$ satisfies a* ***S^α_δ-condition*** *with constant $c > 0$ if*

$$|h(q)| \leq c \cdot \min\left\{S^{\alpha\delta}, |q|^{-\alpha}\right\} \quad \forall q \in \mathbb{R}^d.$$

In the sense of this definition, it is obvious that the regularized Coulomb force (3.2) satisfies a S^2_δ-condition.

In the rest of this section, f will always denote the regularized Coulomb force, and its dependency on S and δ will be suppressed in the notation. The fact that we choose this regularization, which is not even differentiable, might at first be charged with causing difficulties, and in fact, we see that a few results, such as the validity of Liouville's theorem for such interactions or computing Lipschitz constants of convolutions because one may a priori not differentiate, require additional effort. However, it is a modification which one can superbly do explicit calculations with, in particular regarding integration, so it is really worth the additional effort in some areas.

Note that having in mind our scaling of the spatial coordinates from section 1.1 which delivered us the weak-coupling factor N^{-1}, the total density (respecting particles of all types) in the scaled system is proportional to N and therefore roughly proportional to S in case the shares $\alpha_k = \frac{N_k}{N}$ of particles of type k converge. Hence, the average distance between particles in the scaled system is proportional to $S^{-\frac{1}{3}}$. Therefore, cutting off the Coulomb force S-dependently at this distance and proving propagation of chaos would already be a highly desirable result. Unfortunately, it turns out that we cannot yet pass this bar, however, we can get arbitrarily close in the sense that for the cut-off radius decreasing with $S^{-\delta}$ for $0 < \delta < \frac{1}{3}$, we can prove that propagation of chaos holds typically. This is the goal of the current section.

We will rely on the results sketched resp. desired in the short discussion in section 2.3. Hence, we introduce the following definition, which according to this discussion we expect to be satisfied at least if all $u_{k,0}$ are smooth with compact support, but also for other initial densities.

Definition 3.2. *We say that $u_0 : \mathbb{R}^3 \times \mathbb{R}^3 \to ([0, \infty))^n$ satisfies hypothesis A if for every $T > 0$, a unique solution v to the Vlasov-Poisson system with initial condition $v_0 = u_0$ exists on $\mathbb{R}^3 \times \mathbb{R}^3 \times [0, T]$, and this solution induces measure preserving, global flows $\phi^v_{k,t}$ and bounded spatial densities in the sense that $\sup_{0 \leq t \leq T} \sup_{k \in [n]} \|\rho^v_{k,t}\|_\infty < \infty$. Moreover, the unique, global solutions u^S to the regularized Vlasov-Poisson system with initial condition u_0 are supposed to induce uniformly in the cut-off parameter (which corresponds to S for the cut-pff we introduced) bounded densities for the regularized Vlasov-Poisson system, which, in our setting, means that for* *fixed* $T > 0$,

$$\sup_{S\in\mathbb{N}} \sup_{0\le t\le T} \sup_{k\in[n]} \|\rho_{k,t}^{u^S}\|_\infty < \infty.$$

We will denote by Ψ the flow to (1.4) with $f_{k,l} := c_k c_l \cdot f$, i.e. for

$$\dot{q}_i(t) = \frac{p_i(t)}{m_k}, \qquad \dot{p}_i(t) = \frac{1}{N} \sum_{l\in[n]} \sum_{\substack{j\in\Gamma_l \\ j\neq i}} c_k c_l\, f\big(q_i(t) - q_j(t)\big), \qquad k \in [n],\ i \in \Gamma_k, \tag{3.3}$$

which describes a regularized gravitational or electrostatic system, and by μ_{emp} the corresponding empirical measures. μ_t denotes the (curve of) probability measures induced by the solution u to the Vlasov-Poisson system with regularized Coulomb force f, i.e.the solution to

$$\partial_t u_k + \tfrac{p}{m_k} \cdot \nabla_q u_k + F_k^u \cdot \nabla_p u_k = 0, \qquad F_k^u = \sum_{l\in[n]} \alpha_l \cdot c_k c_l \cdot (f_{k,l} *_q u_{l,t}), \qquad k \in [n] \tag{3.4}$$

which induces a corresponding solution to the Vlasov equation in integral form, i.e. for $\mu_{k,t}$ the probability measure given by the density $u_{k,t}$,

$$\mu_{k,t} = \mu_{k,0} \circ \big(\varphi_{k,t}^{\mu}\big)^{-1} \qquad \forall k \in [n],\ t \in [0, T], \tag{3.5}$$

where $\varphi_{k,t}^{\mu}$ is the unique, global flow to the ODE

$$\dot{q}(t) = \frac{p(t)}{m_k}, \qquad \dot{p}(t) = F_k^{\mu}(q(t), t) = \sum_{l\in[n]} \alpha_k \cdot c_k c_l \cdot (f *_q \mu_{l,t})(q(t)). \tag{3.6}$$

Moreover, v is supposed to be a solution to the unmodified Vlasov-Poisson system, i.e. it satisfies (3.4),(3.5),(3.6) with the regularized force f replaced by the Coulomb kernel k. From now on, we will leave out the S-dependence entirely in the notation unless it is explicitly necessary.

We also define the regularized resp. regularized mean field force field F resp. $\overline{F}_t : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ and (heuristically speaking) the associated gradients G resp. $\overline{G}_t : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ component-wise as follows: given a solution u for the regularized Vlasov-Poisson system, for $Z \in \mathbb{R}^{3N}$, $k \in [n]$ and $i \in [N_k]$,

$$\begin{aligned} F_k^i(Z) &= \frac{1}{N} \sum_{l=1}^{n} \sum_{j=1}^{N_l} c_k c_l \cdot f(Z_k^{1,i} - Z_l^{1,j}), \qquad & \overline{F}_{k,t}^i(Z) &= \sum_{l=1}^{n} \alpha_l \cdot c_k c_l \,(f *_q u_{l,t})(Z_k^{1,i}), \\ G_k^i(Z) &= \frac{1}{N} \sum_{l=1}^{n} \sum_{j=1}^{N_l} c_k c_l \cdot g(Z_k^{1,i} - Z_l^{1,j}), \qquad & \overline{G}_{k,t}^i(Z) &= \sum_{l=1}^{n} \alpha_l \cdot c_k c_l \,(g *_q u_{l,t})(Z_k^{1,i}), \end{aligned} \tag{3.7}$$

where g is defined in lemma 4.73. Just like for f, g, u, we suppress the S-dependency of $F, \overline{F}_t, G$ and $\overline{G}_t$. Let us briefly mention some elementary bounds on these forces:

Lemma 3.3. *Under our general hypothesis A, given $T > 0$ there are constants $C_{1,2,3,4} > 0$ such that for $0 \le t \le T$, $\|F\|_\infty \le C_1 \cdot S^{2\delta}$, $\|\overline{F}_t\|_\infty \le C_2$, $\|G\|_\infty \le C_3 \cdot S^{3\delta}$ and $\|\overline{G}_t\| \le C_4 \cdot (1 + \ln(S))$.*

Proof. The claims for F, G are clear from the definitions, see (3.2) and (4.24). Note that f satisfies a S_δ^2-condition and g satisfies a S_δ^3-condition. Moreover, $f *_q u_{l,t} = f * \rho_{l,t}^u$ for all $l \in [n]$, $t \in [0, T]$ and likewise for g, and hence, recalling that by hypothesis A we get bounds on $\rho_{l,t}^u$ uniformly in $t \in [0, T]$, the statements for $\overline{F}_t$, $\overline{G}_u$ also directly follow from lemma 4.51. □

In fact, under the depicted scenario, we will be able to quantify the rate of convergence, namely that the probability for a bounded Lipschitz distance which is bigger than $S^{-\delta}$ for $\delta \in (0, \frac{1}{3})$ goes to 0 with some negative power of S. More precisely, the following is true, motivated and to be compared with theorem 4.4. from [28]:

Theorem 3.4 (Particle approximation of the Vlasov-Poisson system)**.** *Let u_0 satisfy our general hypothesis A and assume that the r-th moments of all $u_{k,0}$ exist, i.e.*

$$\int_{\mathbb{R}^3} |x|^r \cdot u_{k,0}(x) \,\mathrm{d}x < \infty \qquad \forall k \in [n].$$

Moreover, let $\delta \in (0, \frac{1}{3})$, $\gamma \in (0, \min\left\{\frac{1}{6}, \delta\right\})$ and $T > 0$. Let ν_t be the probability measure given by the n-tuple of densities v_t where v_t is the solution to unmodified the Vlasov-Poisson system with initial state u_0, and $\mu_{\mathrm{emp},t}$ the empirical distribution associated with (3.3) *and with initial state Z such that $(Z_k^i)_{(k,i)\in[n\times\mathbb{N}]}$ are independent and for $k \in [n]$, $(Z_k^i)_{i\in\mathbb{N}}$ are identically distributed with corresponding probability density $u_{k,0}$. Then there are constants $c, C > 0$ such that for $S \in \mathbb{N}$ large enough,*

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{0\le t\le T} d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},t}, \nu_t) > S^{-\gamma+1-3\delta}\right\}\right] \le C \cdot \left(e^{-cS^{1-6\gamma}} + S^{1-\frac{r}{2}}\right).$$

Note that r can be chosen arbitrarily large if u_0 has compact support, so for *nice* initial conditions, the rate of convergence is determined by γ, δ and r, and for $\delta \nearrow \frac{1}{3}$, the bound after which convergence starts goes to infinity. By contrast to [28], we restrict to the bounded Lipschitz distance because all the arguments which generalize the result to the p-th Wasserstein distance can directly be taken from [28], and it is convenient not to introduce another hurdle to get into the topic. At the end of the introductory and motivational part of this section, let us mention that in fact, some people think that one might be able to prove theorems concerning existence of global solutions to (3.3) by proving propagation of chaos with a better (stricter) cut-off than the one we use in(3.2) because one can then "transfer" the existence of solutions from the Vlasov-Poisson system to (3.3).

On the way to theorem 3.4, we will concentrate mostly on the building blocks of the proof and leave the technical, but not too insightful aggregation, which consists of tracing and adjusting the constants, mostly to [28]. Let us make one final remark that might further enhance understanding of what is going on here: in section 3.1, we used that solutions to the Vlasov equation with bounded Lipschitz forces which are close initially stay close deterministically. However, at the beginning of this section, we have given a hopefully convincing argument that this is not true for the Coulomb interaction case, albeit things supposedly go wrong only with small probability for large S, and that we get into trouble as soon as particles get too close. In particular, we need to ensure that particles from different species are also independent because otherwise it might happen that the probability densities for types k and l are similar and *all the particles of type k are picked next to a corresponding particle of type l*, which would imply large interaction forces, which make the mean field description likely to be faulty. We will soon see where in the proof this additional assumption enters.

From now on, the parameters n, N_k and thus α_k resp. c_k as defined in section 1.1 resp. in the microscopic model (3.3) will be fixed, as will $S \in \mathbb{N}$, $\delta \in (0, \frac{1}{3})$, at least preliminarily. We also only consider the case $S \ge 3$, which implies that also $\ln(S) \ge 1$ and thus makes some estimates look a bit more pleasant. Constants will typically be denoted by $C_1, C_2, \dots,$, but might change their value from time to time.

Our goal consists of proving that the time evolution of the empirical distribution $\mu_{\mathrm{emp},t}$ induced by the flow Ψ of (3.3) where f is given by (3.2) is close to the curve of (n-tuples of) measures μ_t induced by a solution of the Vlasov-Poisson system equation u with the regularized Coulomb force, where the initial distribution of point particles for $\mu_{\mathrm{emp},k,0}$ does of course take place in an independent manner with the law $\mu_{k,0}$. Let us give the central idea of the proof of theorem 3.4, namely to split for $\omega \in \Omega$

$$d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},t}, \mu_t) \le d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},t}, \overline{\mu}^{\omega}_{\mathrm{emp},t}) + d_{\mathrm{BL}}(\overline{\mu}^{\omega}_{\mathrm{emp},t}, \mu_t) + d_{\mathrm{BL}}(\mu_t, \nu_t). \tag{3.8}$$

Only $\overline{\mu}_{\mathrm{emp},t}$ has not been explained yet; it denotes the empirical measure of the time evolution according to the flows $\varphi_{k,t}$, i.e. associated to the flow

$$\Phi_t(Z) := \left(\prod_{k=1}^{n} \Phi^1_{k,t}(Z_k), \prod_{k=1}^{n} \Phi^2_{k,t}(Z_k) \right) := \left(\prod_{k=1}^{n} \prod_{i=1}^{N_k} \varphi^1_{k,t}(Z^i_k), \prod_{k=1}^{n} \prod_{k=1}^{N_k} \varphi^2_{k,t}(Z^i_k) \right), \tag{3.9}$$

where the $\varphi_{k,t} := \varphi^u_{k,t}$ are from (3.5). Let us start with discussing the second summand in (3.8). One would hope that we can apply the methods of section 2.2 because by definition of $\overline{\mu}_{\mathrm{emp}}$, both $\overline{\mu}_{\mathrm{emp}}$ and μ evolve according to the same flow: for $k \in [n]$, $\overline{\mu}_{\mathrm{emp},k,t} = \overline{\mu}_{\mathrm{emp},k,0} \circ (\varphi_{k,t})^{-1}$ and $\mu_{k,t} = \mu_{k,0} \circ (\varphi_{k,t})^{-1})$, so for any $h \in \mathrm{BL}(\mathbb{R}^6)$

$$\begin{aligned} \left| \int_{\mathbb{R}^6} h \left(\mathrm{d}\overline{\mu}_{\mathrm{emp},k,t} - \mathrm{d}\mu_{k,t} \right) \right| &= \left| \int_{\mathbb{R}^6} h \left(\mathrm{d}\big(\mu_{\mathrm{emp},k,0} \circ \varphi^{-1}_{k,t}\big) - \mathrm{d}\big(\mu_{k,0} \circ \varphi^{-1}_{k,t}\big) \right) \right| \\ &= \left| \int_{\mathbb{R}^6} (h \circ \varphi_{k,t}) \left(\mathrm{d}\overline{\mu}_{\mathrm{emp},k,0} - \mathrm{d}\mu_{k,0} \right) \right| \\ &\le \left\| h \circ \varphi_{k,t} \right\|_{\mathrm{BL}} \cdot d_{\mathrm{BL}}(\overline{\mu}_{\mathrm{emp},k,0}, \overline{\mu}_{\mathrm{emp},0}). \end{aligned} \tag{3.10}$$

Consequently, we only need a bound on $\|\varphi_{k,t}\|_{\mathrm{L}}$ by remark 4.37 (d). Fortunately, the mean field interaction coming from the regularized Coulomb interaction is indeed Lipschitz continuous because from our general assumption A, we know that the spatial densities are uniformly in S bounded, and hence by corollary 4.55, $\|f * \rho_{k,t}\|_{\mathrm{L}}$ and thus the $\overline{F}_{k,t} := c_k \cdot \sum_{l=1}^{n} \alpha_l c_l \cdot (f *_q \rho_{l,t})$ are Lipschitz continuous with $\|F_{k,t}\|_{\mathrm{L}} \le C \cdot (1 + \ln(S))$ uniformly in t on any fixed $[0,T]$. In lemma 2.10, we saw that for time-independent forces, we can bound $\|\varphi_{k,t}\|_{\mathrm{L}}$ by an exponential bound with exponential growth proportional to the Lipschitz constant of the force, which in case of generalization to time-dependent forces would amount approximately to $\ln(S)$. Unfortunately, it turns out that this growth is too fast to obtain good estimates as $S \to \infty$, so we need to improve. The crucial idea is to punish deviations in space, which are rather easy to control, harder than deviations in momentum. Thus, we introduce the auxiliary flows $\phi_{k,t} := \big(\phi^1_{k,t}, \phi^2_{k,t}\big)$, where

$$\phi^1_{k,t} := \sqrt{\ln(S)} \cdot \varphi^1_{k,t}, \qquad \phi^2_{k,t} := \varphi^2_{k,t}.$$

Then clearly, $\|\varphi_{k,t}\|_{\mathrm{L}} \le \|\phi_{k,t}\|_{\mathrm{L}}$, i.e. we can try to control $\|\phi_{k,t}\|_{\mathrm{L}}$ instead. Indeed, this works out very well: For $x, y \in \mathbb{R}^6$, $k \in [n]$ and $t \in (0, T)$,

$$\begin{aligned}
\partial_t^+ |\phi_{k,t}(x) - \phi_{k,t}(y)| &\leq \left|\partial_t^+ \left(\phi_{k,t}^1(x) - \phi_{k,t}^1(y)\right)\right| + \left|\partial_t^+ \left(\phi_{k,t}^2(x) - \phi_{k,t}^2(y)\right)\right| \\
&= \sqrt{\ln(S)} \cdot \left|\partial_t^+ \left(\varphi_{k,t}^1(x) - \varphi_{k,t}^1(y)\right)\right| + \left|\partial_t^+ \left(\varphi_{k,t}^2(x) - \varphi_{k,t}^2(y)\right)\right| \\
&\leq \sqrt{\ln(S)} \cdot \tfrac{1}{m} \left|\varphi_{k,t}^2(x) - \varphi_{k,t}^2(y)\right| + \left|\overline{F}_t\left(\varphi_{k,t}^1(x)\right) - \overline{F}_t\left(\varphi_{k,t}^1(y)\right)\right| \\
&\leq \sqrt{\ln(S)} \cdot \tfrac{1}{m} \cdot \left|\varphi_{k,t}^2(x) - \varphi_{k,t}^2(y)\right| + \left\|\overline{F}_t\right\|_{\mathrm{L}} \cdot \left|\varphi_{k,t}^1(x) - \varphi_{k,t}^1(y)\right| \\
&\leq \sqrt{\ln(S)} \cdot \tfrac{1}{m} \cdot \left|\phi_{k,t}(x) - \phi_{k,t}(y)\right| + C \ln(S) \cdot \sqrt{\ln(S)}^{-\frac{1}{2}} \cdot \left|\phi_{k,t}(x) - \phi_{k,t}(y)\right| \\
&\leq \sqrt{\ln(S)} \cdot \left(\tfrac{1}{m} + C\right) \cdot |\phi_{k,t}(x) - \phi_{k,t}(y)| \,.
\end{aligned}$$

By Grønwall's lemma, we obtain, using $\varphi_{k,0} = \mathrm{id}_{\mathbb{R}^6}$ and thus $|\phi_{k,0}(x) - \phi_{k,0}(y)| \leq \sqrt{\ln(S)} \cdot |x - y|$,

$$\left|\phi_{k,t}(x) - \phi_{k,t}(y)\right| \leq \sqrt{\ln(S)} \cdot e^{\left(\frac{1}{m} + C\right)\sqrt{\ln(S)}\, t} \cdot |x - y| \,,$$

showing that $\|\phi_{k,t}\|_{\mathrm{L}} \leq \sqrt{\ln(S)} \cdot e^{\left(\frac{1}{m}+C\right)\sqrt{\ln(S)}\, t}$ and consequently the same bound holds for $\|\varphi_{k,t}\|_{\mathrm{L}}$. Thus, for $h \in \mathrm{BL}(\mathbb{R}^6)$ with $\|h\|_{\mathrm{BL}} \leq 1$, we have by remark 4.64 (d) that

$$\|h \circ \varphi_{k,t}\|_{\mathrm{BL}} \leq \left(1 + \sqrt{\ln(S)}\right) \cdot e^{\left(\frac{1}{m}+C\right)\sqrt{\ln(S)}\, t}.$$

Inserting this in (3.10) and assuming $S \geq 3$ shows that

$$d_{\mathrm{BL}}(\overline{\mu}_{\mathrm{emp},t}, \mu_t) \leq C\sqrt{\ln(S)} \cdot e^{\left(\frac{1}{m}+C\right) \cdot \sqrt{\ln(S)}\, t} \cdot d_{\mathrm{BL}}(\overline{\mu}_{\mathrm{emp},0}, \mu_0).$$

We have therefore shown the following result:

Lemma 3.5. *Under our general notation, it holds that*

$$d_{\mathrm{BL}}(\overline{\mu}_{\mathrm{emp},t}, \mu_t) \leq C\sqrt{\ln(S)} \cdot e^{\left(\frac{1}{m}+C\right) \cdot \sqrt{\ln(S)}\, t} \cdot d_{\mathrm{BL}}(\overline{\mu}_{\mathrm{emp},0}, \mu_0).$$

Remark 3.6. This result is the first building block on the way to theorem 3.4. It is of entirely deterministic nature and corresponds to proposition 9.2 in [28]. Note that by this estimate, the bounded Lipschitz distance between the empirical measure evolving according to the regularized mean field force and the solution to the regularized Vlasov-Poisson system grows slower than any positive power of S.

Now, let us turn to the more involved control on the first term in (3.8). To get an estimate on $d_{\mathrm{BL}}(\mu_{\mathrm{emp},t}, \overline{\mu}_{\mathrm{emp},t})$, it suffices to show that typically, the trajectories $\Psi_t(Z)$ of the true system and the trajectories of a system evolving to the mean field flow Φ_t defined by (3.9) stay close. We will prove this by the following theorem:

Theorem 3.7. *Let $T > 0$, $\delta \in (0, \frac{1}{3})$ and assume that u_0 satisfies hypothesis A. Then for every $\gamma > 0$, there is some $C_\gamma > 0$ such that provided $(Z_k^i)_{(k,i)\in[n]\times\mathbb{N}}$ are independent and for $k \in [n]$, $(Z_k^i)_{i\in\mathbb{N}}$ are identically distributed by the law $u_{k,0}$ for all $k \in [n]$,*

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{0\leq t\leq T} \left|\Psi_t(Z(\omega)) - \Phi_t(Z(\omega))\right| \geq S^{-\delta}\right\}\right] \leq C_\gamma \cdot S^{-\gamma}.$$

Indeed, this directly yields an estimate for the bounded Lipschitz distance of the corresponding empirical measures:

Corollary 3.8. *Under the hypotheses of theorem 3.7,*

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{0\leq t\leq T} d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},t}, \overline{\mu}^{\omega}_{\mathrm{emp},t}) \geq S^{-\delta}\right\}\right] \leq C_\gamma \cdot S^{-\gamma}.$$

Proof. Recall that for $\mu, \nu \in (\mathcal{P}(\mathbb{R}^6))^n$, $d_{\mathrm{BL}}(\mu, \nu) = \max\{d_{\mathrm{BL}}(\mu_k, \nu_k) : k \in [n]\}$. Hence it suffices to prove that for every $N \in \mathbb{N}$ and $X = (X_1, \ldots, X_N)$, $Y = (Y_1, \ldots, Y_N) \in \mathbb{R}^{6N}$, defining $\mu := \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$ and $\nu := \frac{1}{N}\sum_{i=1}^N \delta_{Y_i}$, $d_{\mathrm{BL}}(\mu, \nu) \leq |X - Y|$ (as before, we use the maximum norm on $\mathbb{R}^d$ unless otherwise stated). Note that for every $i \in [N]$, $d_{\mathrm{BL}}(\delta_{X_i}, \delta_{Y_i}) = |X_i - Y_i|$. Indeed, for any $g \in \mathrm{BL}(\mathbb{R}^6)$ with $\|g\|_{\mathrm{BL}} = 1$,

$$\left|\int_{\mathbb{R}^d} g\ (\mathrm{d}\delta_{X_i} - d\delta_{Y_i})\right| = |g(X_i) - g(Y_i)| \leq \|g\|_{\mathrm{BL}} \cdot |X_i - Y_i| = |X_i - Y_i|\,.$$

Consequently, by the triangle inequality,

$$d_{\mathrm{BL}}(\mu, \nu) \leq \frac{1}{N}\sum_{i=1}^N d_{\mathrm{BL}}(\delta_{X_i}, \delta_{Y_i}) \leq \frac{1}{N}\sum_{i=1}^N |X - Y| = |X - Y|\,. \qquad \square$$

So, we have to estimate the deviation between trajectories which evolve according to the flow Ψ_t for (3.3) and trajectories which evolve according to the mean field flow Φ_t induced by the solution u to the Vlasov equation with cut-off Coulomb interactions (3.6). By hypothesis, $(Z^i_k)_{(k,i)\in[n]\times\mathbb{N}}$ are independent random variables such that for every $k \in [n]$, $(Z^i_k)_{i\in\mathbb{N}}$ are identically distributed with corresponding probability densities $u_{k,0}$ on phase space $\mathbb{R}^6$. Then for $k \in [n]$ and $i \in [N_k]$, the random variables defined by the *stochastic process* $Z^i_k(\omega, t) := \varphi_{k,t}(Z^i_k(\omega))$ are measurable (continuous) functions of Z^i_k. In particular, $(Z^i_k(\,\cdot\,, t))_{(k,i)\in[n]\times\mathbb{N}}$ are still independent (see [19, p. 71]), and the law corresponding to $Z^i_k(\omega, t)$ has the probability density given by the solution $u_{k,t}$ of the Vlasov equation: For $B \in \mathcal{B}(\mathbb{R}^6)$ and $i \in [N_k]$, by (3.5),

$$\begin{aligned}\mathbb{P}\big(\{\omega \in \Omega : Z^i_k(\omega, t) \in B\}\big) &= \mathbb{P}\big(\{\omega \in \Omega : Z^i_k(\omega) \in (\varphi_{k,t})^{-1}(B)\}\big) = \mu_{k,0}\left((\varphi_{k,t})^{-1}(B)\right) = \mu_{k,t}(B)\\ &= \int_B u_{k,t}\,\mathrm{d}x.\end{aligned}$$

Thus, we are in a setting to apply various forms of the law of large numbers, which in fact we will soon do.

Recall that we want to bound the probability that $|\Psi_t(Z(\omega) - \Phi_t(Z(\omega))|$ is large. As for obtaining the estimate on the second term in 3.8, it turns out that deviations in space are much easier to control than deviations in the momentum component. Hence, we again use the anisotropic scaling which punishes deviations in position much more than these in momentum here as well. Consequently, the quantity which we intend to get some control on is

$$\mathbb{P}\left(\left\{\omega \in \Omega : \sup_{0\leq t\leq T} \Delta(\omega, t)\right\} \geq S^{-\delta}\right),$$

where the stochastic process Δ is defined via

$$\begin{aligned}&\Delta : \Omega \times [0, T] \to \mathbb{R},\\ &\qquad (\omega, t) \mapsto \sqrt{\ln(S)} \cdot \left|\Psi^1_t(Z(\omega)) - \Phi^1_t(Z(\omega))\right| + \left|\Psi^2_t(Z(\omega)) - \Phi^2_t(Z(\omega))\right|.\end{aligned} \tag{3.11}$$

For fixed $\omega \in \Omega$, by corollary 4.19 and remark 4.2 (f), it follows that the map $\Delta_\omega : [0, T] \to \mathbb{R}$, $t \mapsto \Delta(\omega, t)$ is continuous and right-sided differentiable on $(0, T)$ with right-continuous right-sided derivative because $t \mapsto \Psi_t(Z(\omega))$, $t \mapsto \Phi_t(Z(\omega))$ are continuously differentiable. On the other hand, $\Delta_t : \Omega \to \mathbb{R},\ \omega \mapsto \Delta(\omega, t)$ is measurable for every $t \in [0, T]$ as a composition of measurable and continuous maps (recall from lemma 2.10 that Ψ_t, Φ_t are even Lipschitz continuous). Consequently, Δ is Carathéodory in the sense of definition 4.34.

Our first idea might be to directly find a Grønwall-type estimate for $\mathbb{P}\left(\sup_{0\le t\le T} \Delta(\omega, t) \ge S^{-\delta}\right)$ because for $t = 0$ it actually takes the value 0 ($\Psi_0 = \Phi_0^u = \mathrm{id}_{\mathbb{R}^{6N}}$). However, since expectation values are in general better-behaved ("smoother") than probabilities, we take an approach via the expectation value of another, associated stochastic process: We define

$$J : \Omega \times [0, T] \to \mathbb{R}, \qquad (\omega, t) \mapsto \min\left\{1, \sup_{0\le s\le t}\left[e^{\lambda\sqrt{\ln(S)}\,(T-s)} \cdot \left(S^{\delta} \cdot \Delta(\omega, s) + S^{\delta-\frac{1}{3}}\right)\right]\right\},$$

where $\lambda > 0$. Later, λ will be chosen appropriately to optimize our result. In order to keep the arguments in the sequel comprehensive, we introduce even more stochastic processes $\Omega \times [0, T] \to \mathbb{R}$, namely

$$\begin{aligned} I(\omega, t) &:= e^{\lambda\sqrt{\ln(S)}\,(T-t)} \cdot \left(S^{\delta} \cdot \Delta(\omega, t) + S^{\delta-\frac{1}{3}}\right), \\ R(\omega, t) &:= \sup_{0\le s\le t} I(\omega, s). \end{aligned}$$

With this notation, obviously $J = \min\{1, R\}$. The following observations are crucial:

Remark 3.9.

(1) Since $\Delta(\omega, t) \ge 0$ for all $(\omega, t) \in \Omega \times [0, T]$, we see that $I \ge 0$ and therefore $0 \le J \le 1$.

(2) $S \ge 1$, so $\ln(S) \ge 0$ and therefore $e^{\lambda\sqrt{\ln(S)}\,(T-s)} \ge 1$ for all $s \in [0, T]$ (note that $\lambda > 0$). Thus, $J(\omega, t) < 1$ implies that $\Delta(\omega, s) < S^{-\delta}$ for all $0 \le s \le t$.

(3) The map $J_\omega : [0, T] \to \mathbb{R},\ t \mapsto J(\omega, t)$ is obviously non-decreasing for every $\omega \in \Omega$.

(4) J is Carathéodory, and J_ω is right-continuously right-sided differentiable on $(0, T)$ for all $\omega \in \Omega$: We have already seen that Δ is Carathéodory and Δ_ω is continuous and right-sided differentiable with right-continuous right-sided derivative for every $\omega \in \Omega$. Combining this with remark 4.2 (f) (right-sided differentiation rules) and lemma 4.21 (right-sided derivative of the supremum), we obtain continuity and right-continuous right-sided differentiability for I_ω and R_ω. With corollary 4.17 (right-sided derivative of the minimum), we see that J_ω is continuous and right-sided differentiable on $(0, T)$ with right-continuous right-sided derivative. On the other hand, for $t \in [0, T]$, $\Delta_t, I_t : \Omega \to \mathbb{R}$ are measurable as compositions of continuous maps with the measurable map Z. In particular, I is Carathéodory. Using lemma 4.35, we see that R_t is measurable, and therefore J_t is measurable.

For fixed time $t \in [0, T]$, let us define $\mathcal{A}_t \subset \Omega$ as the subset of initial states which have the property that the trajectories of the microscopic and the mean field flow stay close for all times between 0 and t:

$$\mathcal{A}_t := \{\omega \in \Omega : |J(\omega, t)| < 1\}. \tag{3.12}$$

Note that $\mathcal{A}_t = J_t^{-1}((-\infty, 1)) \subset \Omega$ is measurable by (4). By (2) in our above remark, $\omega \in \mathcal{A}_t$ guarantees that $\Delta(\omega, s) < S^{-\delta}$ for all $s \in [0, t]$, so the description is indeed suitable. However, it turns out that being in $\mathcal{A}_t$ is not sufficient to give good bounds on the the growth of Δ_t and hence of J_t. Actually, in order to obtain sufficient control, we also need that the microscopic and mean field force and the respective *gradients* are close on mean field trajectories. We define the corresponding sets

$$\begin{aligned}\mathcal{B}_t &:= \left\{\omega \in \Omega : \left|F\big(\Phi_t^1(Z(\omega))\big) - \overline{F}_t\big(\Phi_t^1(Z(\omega))\big)\right| < S^{-\frac{1}{3}}\right\},\\ \mathcal{C}_t &:= \left\{\omega \in \Omega : \left|G\big(\Phi_t^1(Z(\omega))\big) - \overline{G}_t\big(\Phi_t^1(Z(\omega))\big)\right| < 1\right\},\end{aligned} \tag{3.13}$$

where $F, \overline{F}_t, G, \overline{G}_t : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ are defined in (3.7). That $\mathcal{B}_t, \mathcal{C}_t \subset \Omega$ are measurable for every $t \in [0,T]$ follows directly from measurability of Z and (Lipschitz-) continuity of the maps $F, \overline{F}_t, G, \overline{G}_t, \Phi_t : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ and $|\cdot| : \mathbb{R}^{3N} \to \mathbb{R}$ (for F, G this is clear, for $\overline{F}_t, \overline{G}_t$ it follows from lemma 2.9, and for Φ_t from applying lemma 2.10 component-wise).

We first show that as $S \to \infty$, both $\mathbb{P}[\mathcal{B}_t^c]$ and $\mathbb{P}[\mathcal{C}_t^c]$ decay faster than any negative power of S uniformly in $t \in [0,T]$, i.e. for any $\gamma > 0$, there is some $C_\gamma > 0$ independent of S such that

$$\max\left\{\mathbb{P}[\mathcal{B}_t^c], \mathbb{P}[\mathcal{C}_t^c]\right\} \le C \cdot S^{-\gamma} \qquad \forall t \in [0,T].$$

This implies that

$$\mathbb{P}[\mathcal{B}_t \cap \mathcal{C}_t] = 1 - \mathbb{P}[\mathcal{B}_t^c \cup \mathcal{C}_t^c] \ge 1 - (\mathbb{P}[\mathcal{B}_t^c] + \mathbb{P}[\mathcal{C}_t^c]) \ge 1 - 2C_\gamma \cdot S^{-\gamma}.$$

In other words, initial conditions in $\mathcal{B}_t \cap \mathcal{C}_t$ are *typical* provided that each particle type occurs in a sufficiently large number (and **all** particles are independently chosen initially). In order to prove this, we use a version of the law of large numbers (basically a *high order Markov inequality*, stated and proved in section 4.7) which fits exactly our purposes. The following theorem will find immediate application: By our discussion in section 4.3, for every $k \in [n]$ and $i \in [N_k]$, the probability density corresponding to $Z_k^{1,i}(\omega,t) = \varphi_{k,t}^1(Z_k^i) = \pi_1(\varphi_{k,t}(Z_k^i))$ is given by $\rho_{k,t}^u$. Hence, the Y_k^i in the following theorem should be regarded as $\varphi_{k,t}^1 \circ Z_k^i$.

Theorem 3.10. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and $(Y_k^i)_{(k,i)\in[n]\times\mathbb{N}}$ be independent random vectors $Y_k^i : \Omega \to \mathbb{R}^3$ such that for fixed $k \in [n]$, $(Y_k^i)_{i\in\mathbb{N}}$ are identically distributed with associated probability density $\rho_k \in L^\infty(\mathbb{R}^3)$. Moreover, for $\alpha \in \{2,3\}$ and $k, l \in [n]$, let $h_{k,l}^\alpha \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ satisfy a S_δ^α-condition with constant $|c_{k,l}| > 0$. Let $\delta \in (0, \frac{1}{3})$, $\beta \in \mathbb{R}$ and define*

$$\mathcal{D} := \left\{\omega \in \Omega : \max_{k\in[n]} \max_{i\in[N_k]} \left|\frac{1}{N}\sum_{l=1}^{n}\sum_{j=1}^{N_l}\Big(h_{k,l}^\alpha(Y_k^i(\omega) - Y_l^j(\omega)) - (h_{k,l}^\alpha * \rho_l)(Y_k^i(\omega))\Big)\right| \ge S^{-\beta}\right\}.$$

Provided that $\beta < \frac{1-\delta}{2}$ in case $\alpha = 2$ and $\beta < \frac{1-3\delta}{2}$ in case $\alpha = 3$, the probability of $\mathcal{D}$ goes to 0 with an arbitrary negative power of S, i.e. for all $\gamma > 0$, there is some $C_\gamma > 0$ only depending on γ, α, $\max_{k\in[n]} \|\rho_k\|_\infty$ and $\max_{k,l\in[n]} |c_{k,l}|$ such that

$$\mathbb{P}[\mathcal{D}] \le C_\gamma \cdot S^{-\gamma} \qquad \forall S \in \mathbb{N}. \tag{3.14}$$

Proof. For $k, l \in [n]$ and $i \in [N_k]$ let

$$\mathcal{D}_{k,l}^i := \left\{\omega \in \Omega : \left|\frac{1}{N}\sum_{j=1}^{N_l}\Big(h_{k,l}^\alpha(Y_k^i(\omega) - Y_l^j(\omega)) - (h_{k,l}^\alpha * \rho_l)(Y_k^i(\omega))\Big)\right| \ge n^{-1} \cdot S^{-\beta}\right\}.$$

Since $h^\alpha_{k,l}$ is by assumption Lipschitz-continuous, so is $h^\alpha_{k,l} * \rho_l$ for $k, l \in [n]$ by lemma 2.9, and thus these maps are in particular measurable. As measurability is preserved under taking finite linear combinations and countable suprema and therefore in particular finite maxima, measurability of $\mathcal{D}, \mathcal{D}^i_{k,l} \subset \Omega$ is shown.

If we regard $h^\alpha_{k,l}$ as a pair interaction, we interpret $\mathcal{D}^i_{k,l}$ as the subset of initial conditions for which the microscopic and the mean field force on the i-th particle of type k coming from particles of type l are **not** close. Let

$$\mathcal{D}_l := \bigcup_{k\in[n]} \bigcup_{i\in[N_k]} \mathcal{D}^i_{k,l}.$$

Then $\mathcal{D}_l$ contains all initial conditions for which the microscopic and the mean field force created by particles of type l differ *strongly* on at least one particle. Since $(Y^j_l)_{j\in\mathbb{N}}$ are identically distributed, we obtain

$$\mathbb{P}\left[\mathcal{D}_l\right] \leq \sum_{k=1}^{n} \sum_{i=1}^{N_k} \mathbb{P}\left[\mathcal{D}^i_{k,l}\right] \leq n \cdot \max_{k\in[n]} \sum_{i=1}^{N_k} \mathbb{P}\big[\mathcal{D}^i_{k,l}\big] \leq n \cdot \max_{k\in[n]} \left(N_k \cdot \mathbb{P}\big[\mathcal{D}^1_{k,l}\big]\right).$$

Next, we claim that $\mathcal{D} \subset \bigcup_{l\in[n]} \mathcal{D}_l$: Heuristically, the deviation between the true interaction and the mean field force can only be large if the share generated by at least one type of particles does so. Indeed, assume that $\omega \notin \bigcup_{l\in[n]} \mathcal{D}_l = \bigcup_{l\in[n]} \bigcup_{k\in[n]} \bigcup_{i\in[N_k]} \mathcal{D}^i_k$. Then for all $k, l \in [n]$ and $i \in [N_k]$, $\omega \notin \mathcal{D}^i_{k,l}$, i.e.

$$\left| \frac{1}{N} \sum_{j=1}^{N_l} \Big(h^\alpha_{k,l}(Y^i_k(\omega) - Y^j_l(\omega)) - (h^\alpha_{k,l} * \rho_l)(Y^i_k(\omega)) \Big) \right| < n^{-1} \cdot S^{-\beta}.$$

It follows that for all $k \in [n]$, $i \in [N_k]$,

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{l=1}^{n} \sum_{j=1}^{N_l} \Big(h^\alpha_{k,l}(Y^i_k(\omega) - Y^j_l(\omega)) - (h^\alpha_{k,l} * \rho_l)(Y^i_k(\omega)) \Big) \right| \\
&\leq n \cdot \max_{l\in[n]} \left| \frac{1}{N} \sum_{j=1}^{N_l} \Big(h^\alpha_{k,l}(Y^i_k(\omega) - Y^j_l(\omega)) - (h^\alpha_{k,l} * \rho_l)(Y^i_k(\omega)) \Big) \right| \\
&< n \cdot n^{-1} \cdot S^{-\beta} = S^{-\beta},
\end{aligned}$$

i.e. $\omega \notin \mathcal{D}$. To estimate $\mathbb{P}\big[\mathcal{D}^1_{k,l}\big]$, we check that the assumptions of our high order Markov inequality (theorem 4.57) are satisfied. For fixed $k, l \in [n]$ and $i \in [N_k]$, let us define

$$X^j_{k,l} : \Omega \to \mathbb{R}^3, \qquad \omega \mapsto h^\alpha_{k,l}(Y^1_k(\omega) - Y^j_l(\omega)) - (h^\alpha_{k,l} * \rho_l)(Y^1_k(\omega)).$$

We have already argued that the $X^j_{k,l}$ are measurable, and as measurable functions of the i.i.d variables $(Y^j_l)_{j\in\mathbb{N}}$, $(X^j_{k,l})_{j\in\mathbb{N}}$ are also i.i.d. By boundedness of $h_{k,l}$, also $h_{k,l} * \rho_k$ are bounded, and thus clearly $\|X^j_{k,l}\|_\infty < \infty$. Note that by construction,

$$D^1_{k,l} = \left\{ \omega \in \Omega : \left| \frac{1}{N_l} \sum_{j=1}^{N_l} X^j_{k,l} \right| \geq n^{-1} \cdot S^{-\beta} \right\}. \tag{3.15}$$

Since Y_k^1 and Y_l^j are independent with associated probability densities ρ_k resp. ρ_l, for every $m \in \mathbb{N}$,

$$\mathbb{E}\left[\left(X_{k,l}^j\right)^m\right] = \int_{\mathbb{R}^3\times\mathbb{R}^3} \left(h_{k,l}^\alpha(x-y) - (h_{k,l}^\alpha * \rho_l)(x)\right)^m \cdot \rho_k(x)\,\rho_l(y)\,\mathrm{d}x\,\mathrm{d}y.$$

In particular, for $m=1$, with Fubini's theorem,

$$\begin{aligned}
\mathbb{E}\left[X_{k,l}^j\right] &= \int_{\mathbb{R}^3\times\mathbb{R}^3} \left(h_{k,l}^\alpha(x-y) - (h_{k,l}^\alpha * \rho_l)(x)\right) \cdot \rho_k(x)\,\rho_l(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}^3} \left(\int_{\mathbb{R}^3} h_{k,l}^\alpha(x-y)\cdot\rho_l(y)\,\mathrm{d}y - \int_{\mathbb{R}^3} (h_{k,l}^\alpha * \rho_l)(x)\cdot\rho_l(y)\,\mathrm{d}y\right)\cdot\rho_k(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^3} \left((h_{k,l}^\alpha * \rho_l)(x) - (h_{k,l}^\alpha * \rho_l)(x)\right)\cdot\rho_k(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^3} 0\,\mathrm{d}x = 0.
\end{aligned}$$

Moreover, using lemma 4.51, there are constants $C, \tilde{C}_\alpha > 0$ depending only on $\max_{k,l\in[n]} c_{k,l}$, $\max_{k\in[n]} \|\rho_k\|_\infty$ such that for $m \geq 2$ and $\alpha \in \{2,3\}$,

$$\begin{aligned}
\left|\mathbb{E}\left[\left(X_{k,l}^j\right)^m\right]\right| &\leq \int_{\mathbb{R}^3}\left|\int_{\mathbb{R}^3}\left(h_{k,l}^\alpha(x-y) - (h_{k,l}^\alpha * \rho_l)(x)\right)^m\cdot\rho_l(y)\,\mathrm{d}y\right|\cdot\rho_k(x)\,\mathrm{d}x \\
&\leq \int_{\mathbb{R}^3}\left|\left(\left(h_{k,l}^\alpha - (h_{k,l}^\alpha * \rho_l)(x)\right)^m * \rho_l\right)(x)\right|\cdot\rho_k(x)\,\mathrm{d}x \\
&\leq \sup_{x\in\mathbb{R}^3}\left\|\left(h_{k,l}^\alpha - (h_{k,l}^\alpha * \rho_l)(x)\right)^m * \rho_l\right\|_\infty \\
&\leq C\cdot\tilde{C}_\alpha^m\, S^{(\alpha m-3)\delta}\cdot(1+\delta_{\alpha,3}\ln(S))^m \\
&= CS^{-3\delta}\cdot\left(\tilde{C}_\alpha\, S^{\alpha\delta}\cdot(1+\delta_{\alpha,3}\ln(S)\right)^m.
\end{aligned}$$

This shows that we can apply theorem 4.57 in order to estimate (3.15), and using $S \leq N_k$ for all $k \in [n]$, we obtain for $\alpha = 2$ and S large enough

$$\begin{aligned}
\mathbb{P}\left[\mathcal{D}_{k,l}^1\right] &\leq M^{2M+1}\tilde{C}_2^{2M}(S^{2\delta})^{2M}(nS^\beta)^{2M}N_l^{-2M}\left(N_l\,CS^{-3\delta}\right)^M \leq \overline{C}\cdot S^{M(4\delta+2\beta-3\delta)}N_l^{-M} \\
&\leq \overline{C}S^{M(4\delta+2\beta-3\delta-1+\frac{1}{M})}N_l^{-1} = \overline{C}\cdot S^{-M(1-\delta-2\beta-\frac{1}{M})}N_l^{-1}
\end{aligned}$$

because for $M \geq 1$, $N_l^{-M+1} \leq S^{-M+1}$ and $N_l\,S^{-3\delta} \geq S^{1-3\delta} \xrightarrow{S\to\infty} \infty$, i.e. for $\delta < \frac{1}{3}$ and S big enough, $N_l\,CS^{-3\delta} \geq 1$ and hence $\max\left\{1, (N_l\,CS^{-3\delta}\right\} = N_l\,CS^{-3\delta}$. Consequently, for $\alpha = 2$, we finally arrive at

$$\mathbb{P}\left[\mathcal{D}\right] \leq \sum_{l=1}^n \mathbb{P}\left[\mathcal{D}_l\right] \leq n^2\cdot\max_{k\in[n]}\left\{N_l\cdot\mathbb{P}\left[\mathcal{D}_{k,l}^1\right]\right\} \leq \overline{C}\cdot S^{-M(1-\delta-2\beta-\frac{1}{M})}.$$

Since for $\beta < \frac{1-\delta}{2}$, $\varepsilon := 1-\delta-2\beta > 0$ and hence we see that

$$\lim_{M\to\infty} M(1-\delta-2\beta-\tfrac{1}{M}) \geq \lim_{M\to\infty} M\cdot\tfrac{\varepsilon}{2} = +\infty,$$

we must only choose M large enough to guarantee $M(1-\delta-2\beta-\frac{1}{M}) \geq \gamma$ in order to prove (3.14) with $C_\gamma = \overline{C}$ for S large enough. Clearly, this proves the claim for all $S \in \mathbb{N}$ with a modified constant C_γ.

For $\alpha = 3$, we obtain from an analogous calculation that

$$\mathbb{P}\left[\mathcal{D}_{k,l}^1\right] \leq \overline{C}\cdot(1+\ln(S))^{2M}\cdot S^{-M(1-3\delta-2\beta-\frac{1}{M})}N_l^{-1}$$

and consequently

$$\mathbb{P}[\mathcal{D}] \le \sum_{l=1}^{n} \mathbb{P}[\mathcal{D}_l] \le n^2 \cdot \max_{k\in[n]} \left\{ N_l \cdot \mathbb{P}\left[\mathcal{D}^1_{k,l}\right] \right\} \le \overline{C} \cdot (1+\ln(S))^{2M} \cdot S^{-M(1-3\delta-2\beta-\frac{1}{M})}.$$

Since for $\beta < \frac{1-3\delta}{2}$, $1-3\delta-2\beta := 4\varepsilon > 0$, for M large enough, $1-3\delta-2\beta-\frac{1}{M} > 3\varepsilon$ and hence by lemma 4.76,

$$\begin{aligned}\lim_{M\to\infty} S^{-M(1-3\delta-2\beta-\frac{1}{M})} \cdot (1+\ln(S))^{2M} &\le \lim_{M\to\infty} S^{-3M\varepsilon} \cdot (1+\ln(S))^{2M} \\ &= \lim_{M\to\infty} S^{-M\varepsilon} \cdot \left(\frac{1+\ln(S)}{S^{\varepsilon}}\right)^{2M} \\ &\le S^{-M\varepsilon}\end{aligned}$$

for S large enough. Consequently, choosing $M > \frac{\gamma}{\varepsilon}$, the claim follows for S large enough, and hence again for all $S \in \mathbb{N}$ with a modified constant. □

Lemma 3.11. *For u_0, $(Z^i_k)_{(k,i)\in[n]\times\mathbb{N}}$ as in theorem 3.7, $T > 0$ and $\gamma > 0$ arbitrary, there is $C_\gamma > 0$ such that for all $0 \le t \le T$ and $S \in \mathbb{N}$,*

$$\mathbb{P}[\mathcal{B}_t] \le C_\gamma \cdot S^{-\gamma}, \qquad \mathbb{P}[\mathcal{C}_t] \le C_\gamma \cdot S^{-\gamma},$$

where $\mathcal{B}_t, \mathcal{C}_t$ were defined in (3.13)

Proof. Let $f \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ the regularized Coulomb force and $g \in \mathrm{BL}(\mathbb{R}^3;\mathbb{R}^3)$ the *local Lipschitz field* to f defined in lemma 4.73. Then f satisfies a S^2_δ-condition, and g satisfies a S^3_δ-condition by lemma 4.73. Observe that for $\omega \in \Omega$,

$$\begin{aligned}&\left|F\big(\Phi^1_t(Z(\omega))\big) - \overline{F}_t\big(\Phi^1_t(Z(\omega))\big)\right| = \max_{k\in[n]}\max_{i\in[N_k]} \left|F^i_{k,t}\big(\Phi^1_t(Z(\omega))\big) - \overline{F}^i_{k,t}\big(\Phi^1_t(Z(\omega))\big)\right| \\ &= \max_{k\in[n]}\max_{i\in[N_k]} \left| \frac{1}{N}\sum_{l=1}^{n}\sum_{j=1}^{N_l} c_k c_l\, f\big(Z^{1,i}_k(\omega,t) - Z^{1,j}_l(\omega,t)\big) - \sum_{l=1}^{n} \alpha_l \cdot c_k c_l\, (f *_q u_{l,t})(Z^{1,i}_k(\omega,t)) \right| \\ &= \max_{k\in[n]}\max_{i\in[N_k]} \left| \frac{1}{N}\sum_{l=1}^{n}\sum_{j=1}^{N_l} c_k c_l\, f\big(Z^{1,i}_k(\omega,t) - Z^{1,j}_l(\omega,t)\big) - \frac{1}{N}\sum_{l=1}^{n}\sum_{j=1}^{N_l} c_k c_l\, (f *_q u_{l,t})(Z^{1,i}_k(\omega,t)) \right| \\ &= \max_{k\in[n]}\max_{i\in[N_k]} \left| \frac{1}{N}\sum_{l=1}^{n}\sum_{j=1}^{N_k} \Big(c_k c_l\, f\big(Z^{1,i}_k(\omega,t) - Z^{1,j}_l(\omega,t)\big) - (c_k c_l\, f *_q u_{l,t})\,(Z^{1,i}_k(\omega,t)\big) \Big) \right|,\end{aligned}$$

where we used that $\alpha_l = \frac{N_l}{N} = \frac{1}{N} \cdot \sum_{j=1}^{N_l} 1$. Note that by our general hypothesis A, the spatial densities $\|\rho^u_{k,t}\|_\infty$ are bounded uniformly in $S \in \mathbb{N}$, $t \in [0,T]$ by a constant only depending on T and u_0. Consequently, since $\frac{1-\delta}{2} > \frac{1-\frac{1}{3}}{2} = \frac{1}{3}$, by theorem 3.10 with $\alpha = 2$, $\beta = \frac{1}{3}$ and $h^\alpha_{k,l} = c_k c_l\, f$,

$$\mathbb{P}[\mathcal{B}_t] = \mathbb{P}\left[\left\{\omega \in \Omega : \left|F\big(\Phi^1_t(Z(\omega))\big) - \overline{F}_t\big(\Phi^1_t(Z(\omega))\big)\right| \ge S^{-\frac{1}{3}}\right\}\right] \le C_\gamma \cdot S^{-\gamma},$$

where C_γ does depend only on u_0 and T. By almost the same calculation, replacing f by g, since $\frac{1-3\delta}{2} > 0$, with $\alpha = 3$ and $\beta = 0$ we obtain $\mathbb{P}[\mathcal{C}_t] \le C_\gamma \cdot S^{-\gamma}$. □

Now let us turn to the proof of the main theorem. Note that for every $\omega \in \Omega$,

$$J(\omega,0) = \min\left\{1, e^{\lambda\sqrt{\ln(S)}\,T} \cdot \left(S^\delta \cdot 0 + S^{\delta-\frac{1}{3}}\right)\right\} \xrightarrow{S\to\infty} 0 \tag{3.16}$$

is *deterministically* small because $\delta - \frac{1}{3} < 0$ and S is large (see also lemma 4.76). Consequently, $\mathbb{E}[J_0]$ is also bounded by the r.h.s. of (3.16). Our method is to control the right-sided time derivative of the map $[0,T] \to \mathbb{R},\ t \mapsto \mathbb{E}[J_t]$ in order to obtain that $\mathbb{E}[J_t]$ stays small. If this is the case, then since $J_t \geq 0$, we get that $\mathbb{P}[J_t \geq 1]$ is small, which in turn shows that $\mathbb{P}\left[\Delta_t \geq S^{-\delta}\right]$ is small for every $0 \leq t \leq T$, i.e. the true and mean field trajectories typically stay close.

Proof of theorem 3.7. Let us compute the right-sided derivative of $\mathbb{E}[J_t]$. We will split Ω into three parts determined by the sets $\mathcal{A}_t$, $\mathcal{B}_t$ and $\mathcal{C}_t$ such that we can get control either on the size of the sets or the rate of change of J_t on these sets:

$$\begin{aligned}\partial_t^+ \mathbb{E}[J_t] &= \partial_t^+ \left(\int_\Omega J(\omega,t)\,\mathrm{d}\mathbb{P}(\omega) \right) = \int_\Omega \partial_t^+ J(\omega,t)\,\mathrm{d}\mathbb{P}(\omega) = \mathbb{E}\left[\partial_t^+ J\right] \\ &= \mathbb{E}\left[\partial_t^+ J \,|\, \mathcal{A}_t^{\mathrm{c}}\right] + \mathbb{E}\left[\partial_t^+ J \,|\, (\mathcal{A}_t \setminus (\mathcal{B}_t \cap \mathcal{C}_t))\right] + \mathbb{E}\left[\partial_t^+ J \,|\, (\mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{C}_t)\right].\end{aligned}$$

However, we need to justify the interchange of integration and right-sided differentiation (we have already discussed that for every $\omega \in \Omega$, $J(\omega,\,\cdot\,)$ is right-continuously right-sided differentiable). Taking a closer look at the calculations in the following proof, we see that we can bound $\left|\partial_t^+ J(\omega,t)\right|$ uniformly in ω (and even uniformly in $t \in [0,T]$) by some constant (which grows with S, however, this is not a problem here): from (3.18), by Grønwall's lemma (theorem 4.3) we obtain a bound on Δ which does not depend on $\omega \in \Omega$. Again using (3.18), this yields a bound on $\partial_t^+ \Delta(\omega,t)$, and from (3.17), we finally obtain a uniform bound on $\partial_t^+ I(\omega,t)$ and therefore on $\partial_t^+ J(\omega,t)$. Clearly, the bound can be chosen to be uniform in, say, $t \in [0, T+1]$. Since $\mathbb{P}$ is a probability measure, we may apply theorem 4.10, and hence interchanging integration and right-sided differentiation is justified.

Now, we can start with the detailed estimates. Fix $t \in (0,T)$.

(a) For $\omega \in \mathcal{A}_t^{\mathrm{c}}$, it holds that $J(\omega,t) = 1$. We have already observed that $J(\omega,t)$ is non-decreasing in t, however, $J(\omega,t)$ is already maximal, showing that $\partial_t^+ J(\omega,t) = 0$ and consequently

$$\mathbb{E}\left[\partial_t^+ J \,|\, \mathcal{A}_t^{\mathrm{c}}\right] = 0.$$

(b) This time, let $\omega \in \mathcal{A}_t \setminus (\mathcal{B}_t \cap \mathcal{C}_t)$. By corollary 4.17 and lemma 4.21, together with the right-sided differentiation rules mentioned in remark 4.2 (f), we obtain $0 \leq \partial_t^+ J(\omega,t) \leq \partial_t^+ I(\omega,t)$ with

$$\partial_t^+ I(\omega,t) = e^{\lambda\sqrt{\ln(S)}\,(T-t)} \cdot \left(-\lambda\sqrt{\ln(S)} \cdot \left(S^\delta \cdot \Delta(\omega,t) + S^{\delta - \frac{1}{3}} \right) + S^\delta \cdot \partial_t^+ \Delta(\omega,t) \right). \tag{3.17}$$

Moreover,

$$\begin{aligned}\partial_t^+ \Delta(\omega,t) &\leq \sqrt{\ln(S)} \cdot \left|\Psi_t^2\big(Z(\omega)\big) - \Phi_t^2\big(Z(\omega)\big)\right| + \left|F\big(\Psi_t^1(Z(\omega))\big) - \overline{F}_t\big(\Phi_t^1(Z(\omega))\big)\right| \\ &\leq \sqrt{\ln(S)} \cdot \Delta(\omega,t) + \|F\|_\infty + \left\|\overline{F}_t\right\|_\infty.\end{aligned} \tag{3.18}$$

Since $\omega \in \mathcal{A}_t \setminus (\mathcal{B}_t \cap \mathcal{C})$, $\Delta(\omega,t) < S^{-\delta}$ (see remark 3.9 (2)), with lemma 3.3 we can conclude that

$$\partial_t^+ \Delta(\omega,t) \leq \sqrt{\ln(S)} \cdot S^{-\delta} + C_1 S^{2\delta} + C_2 \leq C_3 \cdot S^{2\delta}$$

for $S \geq 1$ and $C_3 > 0$ large enough because the map $\sqrt{\ln(S)} \cdot S^{-\delta}$ is bounded on $[1,\infty)$ (see lemma 4.76). Therefore,

$$\begin{aligned}\partial_t^+ I(\omega,t) &\leq e^{\lambda\sqrt{\ln(S)}\,(T-t)} \cdot \left(\lambda\sqrt{\ln(S)} \cdot \left(S^\delta \cdot S^{-\delta} + S^{\delta-\frac{1}{3}} \right) + S^\delta \cdot C_3 S^{2\delta} \right) \\ &\leq e^{\lambda\sqrt{\ln(S)}\,T} \cdot \left(C_4 \cdot S^{3\delta} \right),\end{aligned}$$

where we used that also the map $S \mapsto \lambda\sqrt{\ln(S)} \cdot (S^{-3\delta} + S^{-\frac{1}{3}-\delta})$ is bounded on $[1, \infty)$ (this is a direct consequence of lemma 4.76).

Now, let $\gamma > 0$ be arbitrary, choose $\varepsilon > 0$ and define $\tilde{\gamma} := \gamma + 3\delta + \varepsilon$. According to lemma 3.11, we can find $C_{\tilde{\gamma}} > 0$ such that $\mathbb{P}[\mathcal{B}_t^c \cup \mathcal{C}_t^c] \leq C_{\tilde{\gamma}} \cdot S^{-\tilde{\gamma}}$, and with the observation that $\mathcal{A}_t \setminus (\mathcal{B}_t \cap \mathcal{C}_t) \subset \Omega \setminus (\mathcal{B}_t \cap \mathcal{C}_t) = \mathcal{B}_t^c \cup \mathcal{C}_t^c$, it follows that

$$\mathbb{E}\left[\partial_t^+ J \,|\, (\mathcal{A}_t \setminus (\mathcal{B}_t \cap \mathcal{C}_t))\right] \leq e^{\lambda\sqrt{\ln(S)}\,T} \cdot C_4 S^{3\delta} \cdot C_{\tilde{\gamma}} S^{-\gamma-3\delta-\varepsilon} \leq C_5 \cdot S^{-\gamma}$$

because $S \mapsto e^{\lambda\sqrt{\ln(S)}\,T} \cdot S^{-\varepsilon}$ is bounded for $S \in [0, \infty)$, also by lemma 4.76.

(c) Finally, let $\omega \in \mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{C}_t$. We want to proceed as in (b), however, this time we need to get much better estimates for $\partial_t^+ \Delta(\omega, t)$ because we expect that initial conditions in $\mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{C}_t$ are typical. The part in (b) where our estimate was bad was the difference between the true force and the mean field force (last step of (3.18)). Let us try to do better here, using that $\omega \in \mathcal{B}_t \cap \mathcal{C}_t$. By the triangle inequality,

$$\begin{aligned} &\left|F(\Psi_t^1(Z(\omega))) - \overline{F}_t(\Phi_t^1(Z(\omega)\right| \\ &\leq \left|F(\Psi_t^1(Z(\omega))) - F(\Phi_t^1(Z(\omega)))\right| + \left|F(\Phi_t^1(Z(\omega))) - \overline{F}_t(\Phi_t^1(Z(\omega)))\right|. \end{aligned} \tag{3.19}$$

The second summand is bounded by $S^{-\frac{1}{3}}$ because $\omega \in \mathcal{B}_t$, see definition 3.13 and lemma 3.11. On the other hand, for the first summand, we compute for every $k \in [n]$, $i \in [N_k]$

$$\begin{aligned} &\left|F_{k,t}^i(\Psi_t^1(Z(\omega))) - F_{k,t}^i(\Phi_t^1(Z(\omega)))\right| \\ &\leq \sum_{l=1}^{n}\sum_{j=1}^{N_l} |c_{k,l}| \cdot \left|f\big(\Psi_{k,t}^{1,i}(Z(\omega)) - \Psi_{l,t}^{1,j}(Z(\omega))\big) - f\big(\Phi_{k,t}^{1,i}(Z(\omega)) - \Phi_{l,t}^{1,j}(Z(\omega))\big)\right|. \end{aligned}$$

Now, we observe that by definition of Δ (see 3.11),

$$\begin{aligned} &\left|\big(\Psi_{k,t}^{1,i}(Z(\omega)) - \Psi_{l,t}^{1,j}(Z(\omega))\big) - \big(\Phi_{k,t}^{1,i}(Z(\omega)) - \Phi_{l,t}^{1,j}(Z(\omega))\big)\right| \\ &\leq \left|\Psi_{k,t}^{1,i}(Z(\omega)) - \Phi_{k,t}^{1,i}(Z(\omega))\right| + \left|\Psi_{l,t}^{1,j}(Z(\omega)) - \Phi_{l,t}^{1,j}(Z(\omega))\right| \\ &\leq \frac{2}{\sqrt{\ln(S)}} \cdot \Delta(\omega, t). \end{aligned}$$

Since $\omega \in \mathcal{A}_t$, by remark 3.9 (2) we know that $\Delta(\omega, t) < S^{-\delta}$, and for S large enough ($S \geq e^{16}$), $\left|\big(\Psi_{k,t}^{1,i}(Z(\omega)) - \Psi_{l,t}^{1,j}(Z(\omega))\big) - \big(\Phi_{k,t}^{1,i}(Z(\omega)) - \Phi_{l,t}^{1,j}(Z(\omega))\big)\right| < \frac{1}{2}S^{-\delta}$. With lemma 4.73 and lemma 3.11 again, this time for $\mathcal{C}_t$,

$$\begin{aligned} &\left|F_{k,t}^i(\Psi_t^1(Z(\omega))) - F_{k,t}^i(\Phi_t^1(Z(\omega)))\right| \\ &\leq \frac{1}{N}\sum_{l=1}^{n}\sum_{j=1}^{N_l} |c_{k,l}| \cdot g\big(\Phi_{k,t}^{1,i}(Z(\omega)) - \Phi_{l,t}^{1,j}(Z(\omega))\big) \cdot \tfrac{2}{\sqrt{\ln(S)}} \cdot \Delta(\omega, t) \\ &\leq \tfrac{2}{\sqrt{\ln(S)}} \cdot \Delta(\omega, t) \cdot \big(1 + \left|(g * \rho_{k,t})(\Phi_{k,t}^{1,i}(Z(\omega)\right|\big) \\ &\leq \tfrac{2}{\sqrt{\ln(S)}} \cdot \Delta(\omega, t) \cdot (1 + \|g\|_\infty) \leq \frac{2}{\sqrt{\ln(S)}} \cdot \Delta(\omega, t) \cdot (C \cdot (1 + \ln(S))) \\ &\leq C_7 \cdot \Delta(\omega, t) \cdot \sqrt{\ln(S)} \end{aligned}$$

for $S \geq 3$, where we used $\omega \in \mathcal{C}_t$ in the second step and lemma 3.3 in the third step. Putting everything together, we arrive at

$$\begin{aligned}\partial_t^+ I(\omega,t) &\leq e^{\lambda\sqrt{\ln(S)}\,(T-t)} \cdot \Big[\big(-\lambda\sqrt{\ln(S)}\big)\cdot\Big(S^{\delta}\cdot\Delta(\omega,t)+S^{\delta-\frac{1}{3}}\Big) \\ &\qquad\qquad + S^{\delta}\cdot\Big(\sqrt{\ln(S)}\cdot\Delta(\omega,t)+S^{-\frac{1}{3}}+C_7\Delta(\omega,t)\sqrt{\ln(S)}\Big)\Big] \\ &= e^{\lambda\sqrt{\ln(S)}\,(T-t)}\cdot S^{\delta}\sqrt{\ln(S)}\cdot\Big[(-\lambda+1+C_7)\cdot\Delta\big(\omega,t\big)+\big(-\lambda+\tfrac{1}{\sqrt{\ln(S)}}\big)\cdot S^{-\frac{1}{3}}\Big].\end{aligned}$$

Choosing $\lambda \geq C_7 + 1$ and S big enough, wee see that $\partial_t^+ I(\omega,t) \leq 0$ and therefore

$$\mathbb{E}\left[\partial_t^+ J \,|\, \mathcal{A}_t \cap \mathcal{B}_t \cap \mathcal{C}_t\right] = 0.$$

Consequently, only (b) gives a positive contribution, and aggregating all results we get

$$\partial_t^+ \mathbb{E}\left[J_t\right] \leq C_\gamma \cdot S^{-\gamma}.$$

for all $t \in [0,T]$ and arbitrary $\gamma > 0$. In particular, by Grønwall's lemma 4.3,

$$\mathbb{E}\left[J_T\right] - \mathbb{E}\left[J_0\right] \leq TC_\gamma \cdot S^{-\gamma}.$$

Since $\Psi_0 = \Phi_0 = \mathrm{id}_{\mathbb{R}^{6N}}$, $\Delta(\omega,0) = 0$ for all $\omega \in \Omega$ and hence

$$I(\omega,0) = e^{\lambda\sqrt{\ln(S)}\,T}\cdot S^{\delta-\frac{1}{3}} = S^{\lambda T\cdot\frac{1}{\ln(S)}+\delta-\frac{1}{3}} \xrightarrow{n\to\infty} 0$$

by lemma 4.76. We could now directly apply Grønwall's lemma and get that $\mathbb{E}\left[J_t\right] \xrightarrow{S\to\infty} 0$ and therefore $\mathbb{P}\big(\Delta_t \geq S^{-\delta}\big) \leq \mathbb{P}\left[J_t \geq 1\right] \leq \mathbb{E}\left[J_t\right] \xrightarrow{S\to\infty} 0$, where we applied the common Markov inequality. However, note that the bound which we obtained for the growth of $\mathbb{E}\left[J_t\right]$ is much better than the one for $\mathbb{E}\left[J_0\right]$, so we try to get even more: Choose S large enough such that $\mathbb{E}\big(J_0\big) = e^{\lambda\sqrt{\ln(S)}T}\cdot S^{\delta-\frac{1}{3}} \leq \frac{1}{2}$; this is possible by lemma 4.76. By the Markov inequality, this time applied to the non-negative random variable $J_T - J_0$, it follows that

$$\mathbb{P}\left[\left\{\omega\in\Omega : J_T(\omega) - J_0(\omega) \geq \tfrac{1}{2}\right\}\right] \leq 2\mathbb{E}\left[J_T - J_0\right] \leq 2TC_\gamma \cdot S^{-\gamma}.$$

On the other hand, for $J_T(\omega) - J_0(\omega) < \frac{1}{2}$, with $J_0(\omega) < \frac{1}{2}$, we obtain $J_T(\omega) < 1$ and consequently

$$J_T(\omega) - J_0(\omega) \geq e^{\lambda\sqrt{\ln(S)}\,T}\cdot S^{\delta}\cdot \sup_{0\leq t\leq T}\Delta(\omega,t) \geq S^{\delta}\cdot\sup_{0\leq t\leq T}\Delta(\omega,t),$$

i.e. $\sup_{0\leq t\leq T}\Delta(\omega,t) \leq \frac{1}{2}S^{-\delta}$. Consequently, $\sup_{0\leq t\leq T}\Delta(\omega,t)$ can only be larger than $S^{-\delta}$ if $J_T(\omega) - J_0(\omega) \geq \frac{1}{2}$, and consequently for S large enough,

$$\mathbb{P}\left[\left\{\omega\in\Omega : \sup_{0\leq t\leq T}\Delta(\omega,t) \geq S^{-\delta}\right\}\right] \leq 2C\cdot S^{-\gamma}. \qquad \square$$

This, together with corollary 3.8, constitutes the second, hardest building block. There are still two other, smaller building blocks. One of them is only an application of theorem 4.48 with ε chosen as a suitable function of S, which together with the first building block gives corollary 9.4 in [28] and the proof of which can almost literally be taken from the paper because we can reduce everything to components. Moreover, as in proposition 9.1 in [28], we can estimate the difference between solutions of the Vlasov-Poisson system and corresponding solutions to the Vlasov-Poisson equation with cut-off, i.e. (3.4) with f replaced by the Coulomb kernel resp. without modification, for the same initial conditions. Note that is is there where the measure preserving property of the flow of a solution to the Vlasov-Poisson equation, i.e. the results of section 4.9, are crucial. The computation is in fact similar to the one executed for building block 1 and can be taken from [28] without noteworthy modification. We just state the result here, which of course again relies on the validity of our assumptions stated in section 2.3.

Theorem 3.12. *Let $T > 0$ and $u_0 = (u_{1,0}, \dots, u_{n,0})$ satisfy our general assumptions A from definition 3.2. Then there is some $C > 0$ such that for $S \geq 3$,*

$$d_{\mathrm{BL}}(\mu_t, \nu_t) \leq S^{-\delta} \cdot e^{C\sqrt{\ln(S)}\cdot t} \qquad \forall t \in [0, T].$$

Note that this is again an entirely deterministic result.

Consequently, we have collected all the building blocks for theorem 3.4. Actually, since for all the terms occurring in (3.8), we have shown convergence in probability as announced in the beginning of the section, it follows that solutions of (1.4) with regularized Coulomb interactions converge in probability to the solution of the Vlasov-Poisson system, measured in the bounded Lipschitz distance. As already mentioned in the beginning of this chapter, a more precise analysis for the rate of convergence can be obtained by following the proof of corollary 9.4 and the proof of theorem 4.4. thereafter in [28] almost literally because the results we got in the building blocks are virtually the same for n particles (provided the initial conditions for *all* particles are given by realizations of independent random variables) as they are for $n = 1$. In summary, propagation of chaos holds in probability, more precisely with the rate of convergence as stated in theorem 3.4. This is certainly an interesting result, and the techniques used for obtaining it are rather new and thus maybe not yet exhausted, so we might hope that in future, cutoffs and estimates can be improved further. This is also highly desirable because yet, the physical significance of the results is rather low because still the average particle distance in the system is such that they "feel" the cut-off, i.e. the regularized microscopic model deviates too heavily from the true microscopic model. Nevertheless, the results are a milestone on the hopefully successful way to extend the yet achieved results to cut-offs which converge to the true interaction at a much faster rate and therefore finally give a complete and convincing derivation of the Vlasov equation from the microscopic time evolution, therefore theoretically justifying the application of the Vlasov equation in the scenarios in which it is already used.

4 Appendix: Mathematical Resources

4.1 One-sided differentiability

In the course of the main text, we have frequently encountered situations with a need for bounds on various time-dependent quantities. However, in these cases, the quantity of interest is in general not differentiable, which a priori prevents us from applying useful results from single-variable calculus such as the well-known Grønwall's lemma in differential form (see e.g. [14, p. 708 f.]) or the mean value theorem of differentiation (cf. [3, p. 169]). Hence, we seek for a notion slightly more general than differentiability for which important results from the theory of differentiable functions of one variable are - possibly in a weaker form - still valid. It turns out that *one-sided differentiability* provides a convenient generalization of differentiability to tackle this task. Therefore, in this section, we want to derive some auxiliary results on one-sided differentiable functions. The main results will be a remarkable generalization of the above-mentioned Grønwall's lemma and a weak version of the mean value theorem of differentiation, leading to a result concerning the interchange of one-sided differentiation and integration.

In the following section, we will sometimes talk about differentiability of maps defined on non-open subsets of $\mathbb{R}$. In this case, we follow the convention from definition 4.59, which defines differentiability on an arbitrary subset of $\mathbb{R}$ by existence of a differentiable extension on an open subset of $\mathbb{R}$.

Definition 4.1. *We call a subset $I \subset \mathbb{R}$* ***right-open*** *if for every $t \in I$, there is some $\delta > 0$ such that $[t, t+\delta] \subset I$. Let $I \subset \mathbb{R}$ right-open and $f : I \to \mathbb{R}^d$ a map.*

(i) We say that f is ***right-continuous in $t \in I$*** *if for all $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}^+$ such that $t + h_n \in I$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} h_n = 0$,*

$$\lim_{n\to\infty} f(t + h_n) = f(t).$$

In short form, we will also write $\lim_{n\to\infty} f(t_n) = f(t) \ \ \forall (t_n)_{n\in\mathbb{N}} \searrow t$.

(ii) We call f ***right-sided differentiable in $t \in I$*** *if there is some $a \in \mathbb{R}^d$ such that for all $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}^+$ satisfying $t + h_n \in I$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} h_n = 0$,*

$$\lim_{n\to\infty} \frac{1}{h_n}\big(f(t + h_n) - f(t) - h_n \cdot a\big) = 0.$$

In this case, we call $f'^{,+}(t) := \partial_t^+ f(t) := \frac{\mathrm{d}}{\mathrm{d}t^+} f(t) := a$ the ***right-sided derivative of f in $t \in I$****.*

If f is right-continuous in all $t \in I$, we say that f is right-continuous. Analogously, we call f right-sided differentiable if this property holds for all $t \in I$.

In the sequel, unless specified otherwise, I will always denote a right-open subset of $\mathbb{R}$.

Remark 4.2.

(a) By the definition of a right-open set, we see that the conditions for right-continuity and right-sided differentiability are always non-empty.

(b) It is straightforward to check that $I \subset \mathbb{R}$ is a right-open interval if and only if there are $a, b \in \mathbb{R}$ with $a < b$ such that

$$I \in \{\mathbb{R}, (-\infty, b), [a, b), (a, b), [a, \infty), (a, \infty)\}.$$

(c) Obviously, every continuous function on I is right-continuous, and by our notion of differentiability on I (definition 4.59) and remark 4.60, every differentiable function on I is right-sided differentiable, with its right-sided derivative coinciding with the (classical) derivative. Moreover, it is readily checked that the right-sided derivative is unique provided it exists.

(d) Just as it is the case for continuous functions from I to $\mathbb{R}^d$, we can characterize right-continuity by an appropriate ε-δ-criterion: f is right-continuous on I if and only if

$$\forall \varepsilon > 0 \quad \forall t \in I \quad \exists \delta > 0 : \ s \in [t, t+\delta] \cap I \ \Rightarrow \ |f(s) - f(t)| < \varepsilon.$$

The proof can almost literally be taken from the corresponding proof for (classical) continuity in any elementary calculus textbook.

(e) Let us define a right-sided analogue for the little-o-notation: For $\delta > 0$ and $g : (0, \delta) \to \mathbb{R}$, we say $g \in o(h^+)$ if for all $(h_n)_{n \in \mathbb{N}} \subset (0, \delta)$ with $\lim_{n \to \infty} h_n = 0$, it holds that $\frac{g(h_n)}{h_n} \xrightarrow{n \to \infty} 0$. Note that this definition can be regarded as being independent of δ because convergence is only determined by the *backmost* elements of a sequence. Now, we observe that $f : I \to \mathbb{R}^d$ is right-sided differentiable in $t \in I$ iff there is some (then unique) $a \in \mathbb{R}^d$ such that for $\delta > 0$ small enough (namely such that $[t, t+\delta) \subset I$) and r defined by $r : (0, \delta) \to \mathbb{R}, \ h \mapsto f(t+h) - f(t) - h \cdot a$, we have $r \in o(h^+)$.

(f) The common limit theorems for sums and products of sequences directly imply that sums and products of right-continuous resp. right-sided differentiable functions $f, g : I \to \mathbb{R}$ are right-continuous resp. right-sided differentiable, and in the latter case, the corresponding differentiation rules hold. In particular, the right-sided derivative is a linear operator. However, note that in general, the analogue to the chain rule for compositions of right-sided differentiable functions fails to hold: one can easily convince oneself that it is necessary that the inner function is non-decreasing in order to ensure bare existence of the considered limit. On the other hand, for g right-sided differentiable at t and f differentiable at $g(t)$, it is clear from the proof of the classical result (see e.g. [2, p. 305 f.]) that $f \circ g$ is right-sided differentiable at t, with $(f \circ g)^{\prime,+}(t) = f'(g(t)) \cdot g^{\prime,+}(t)$.

(g) All the definitions can easily be modified with "left" replacing "right", and usually the results in the following remain valid when formulated accordingly, which can be seen either by adapting the definition/proof in the obvious way or applying our results to $f(-\,\cdot\,) : -I \to \mathbb{R}$, which is obviously right-continuous resp. right-sided differentiable iff f is left-continuous resp. left-sided differentiable.

Directly from the first definitions, we can prove the announced generalization of Grønwall's lemma:

Theorem 4.3 (Grønwall's lemma, generalized differential form)**.** *Let $t_0, T \in \mathbb{R}$, $T > t_0$, and $f : [t_0, T] \to \mathbb{R}$ a continuous map which is right-sided differentiable on (t_0, T). Moreover, assume that $g, h : [t_0, T] \to \mathbb{R}$ are continuous functions such that*

$$f'^{,+}(t) \le g(t) \cdot f(t) + h(t) \qquad \forall t \in (t_0, T).$$

Then for all $t \in [t_0, T]$,

$$f(t) \le \exp\left(\int_{t_0}^{t} g(s)\,\mathrm{d}s\right) \cdot \left[f(t_0) + \int_{t_0}^{t} \exp\left(-\int_{t_0}^{s} g(\tau)\,\mathrm{d}\tau\right) \cdot h(s)\,\mathrm{d}s\right]. \tag{4.1}$$

Corollary 4.4. *Under the hypotheses of theorem 4.3 with the additional assumption that $g, h \ge 0$,*

$$f(t) \le \exp\left(\int_{t_0}^{t} g(s)\,\mathrm{d}s\right) \cdot \left(f(t_0) + \int_{t_0}^{t} h(s)\,\mathrm{d}s\right) \qquad \forall t \in [t_0, T]. \tag{4.2}$$

Proof. Clear. $\square$

Proof of theorem 4.3. Let $u(t)$ denote the r.h.s. of (4.1) for $t \in [t_0, T]$, which is well-defined because all involved integrands are continuous and thus bounded on the compact interval $[t_0, T]$; in particular, they are integrable on any sub-interval of $[t_0, T]$. Using continuity of g and h and the (classical) fundamental theorem of calculus, one can readily check that u is differentiable and the (unique) solution to the initial value problem

$$u'(t) = g(t) \cdot f(t) + h(t), \qquad u(t_0) = f(t_0) \tag{4.3}$$

on $[t_0, T]$ in the sense of definition 4.61. It remains to prove that $f(t) \le u(t)$ for all $t \in (t_0, T)$, since there is nothing to prove for $t = t_0$, and the case $t = T$ follows from continuity of f and u (conservation of weak inequalities under taking limits, an argument that will abundantly and often hiddenly be used in the sequel). For $\varepsilon > 0$, let $h_\varepsilon : [t_0, T] \to \mathbb{R},\ t \mapsto h(t) + \varepsilon$, and u_ε denote the solution of (4.3) with $f(t_0)$ replaced by $f(t_0) + \varepsilon$ and h replaced by h_ε. By looking at u_ε, namely the r.h.s of (4.1) with the said substitutions, we immediately deduce from continuity of all involved functions that for fixed $t \in [t_0, T]$,

$$u(t) = \lim_{\varepsilon \searrow 0} u_\varepsilon(t).$$

Hence we need only prove that for all $\varepsilon > 0$ and $t \in (t_0, T)$, $f(t) \le u_\varepsilon(t)$. Assume for contradiction that for some $\varepsilon > 0$, there exists $t \in (t_0, T)$ such that $f(t) > u_\varepsilon(t)$. Define

$$\tau := \inf\left\{t \in (t_0, T) : f(t) > u_\varepsilon(t)\right\} \in [t_0, T).$$

Note that actually $\tau > t_0$ because $f(t_0) < f(t_0) + \varepsilon = u_\varepsilon(t_0)$ and f, u_ε are continuous. Now, the crucial observation is that continuity of f, u_ε also implies that $f(\tau) = u_\varepsilon(\tau)$: on the one hand, by definition of τ, $f(t) \le u_\varepsilon(t)$ for all $t < \tau$, so $f(\tau) \le u_\varepsilon(\tau)$. On the other hand, we can rule out that $f(\tau) > u_\varepsilon(\tau)$ because otherwise due to continuity of f, u_ε we could find some $\delta > 0$ such that $f(s) > u_\varepsilon(s)$ for all $s \in (\tau - \delta, \tau + \delta)$, contradicting the infimum property of τ. Consequently, $f(\tau) = u_\varepsilon(\tau)$, and thus by definition of the infimum there is a sequence $(t_n)_{n\in\mathbb{N}} \subset (\tau, T)$ such that $t_n \searrow \tau$ and $f(t_n) > u_\varepsilon(t_n)$ for all $n \in \mathbb{N}$. For $n \in \mathbb{N}$, define $a_n := t_n - \tau > 0$. Then clearly $t_n = \tau + a_n \in (\tau, T)$ for all $n \in \mathbb{N}$, and $a_n \searrow 0$ as $n \to \infty$. We use $(a_n)_{n\in\mathbb{N}}$ as a *tester* for the right-sided derivative of f at τ:

$$\begin{aligned} f'^{,+}(\tau) &= \lim_{n\to\infty} \frac{1}{a_n}\big(f(\tau+a_n)-f(\tau)\big) \geq \lim_{n\to\infty} \frac{1}{a_n}\big(u_\varepsilon(\tau+a_n)-u_\varepsilon(\tau)\big) = u'_\varepsilon(\tau) \\ &= g(\tau)\cdot u_\varepsilon(\tau) + (h(\tau)+\varepsilon) = g(\tau)\cdot f(\tau) + h(\tau) + \varepsilon \\ &> g(\tau)\cdot f(\tau) + h(\tau), \end{aligned}$$

yielding the desired contradiction. □

Note that the *initial condition* $u_\varepsilon(t_0) = f_0 + \varepsilon$ allowed us to dispense with the requirement of right-sided differentiability in t_0 with the corresponding bound on the right-sided derivative there. The method which we used in the proof of Grønwall's lemma suggests that we can even show a more general result, which will be very useful for estimating the growth of functions with other bounds on their right-sided derivatives:

Theorem 4.5. *Let $t_0 \in \mathbb{R}$, $a, b > 0$, $p_0 \in \mathbb{R}$ and $F \in \mathcal{C}([p_0-b, p_0+b]\times[t_0, t_0+a])$. Define $\alpha := \min\left\{a, b\,\|F\|_\infty^{-1}\right\} > 0$. Consider the IVP*

$$u'(t) = F(u(t),t), \qquad u(t_0) = p_0. \tag{$*$}$$

By Peano's existence theorem (see e.g. [20, p. 10]), there exists a solution of ($*$) (in the sense of definition 4.61) defined on $[t_0, t_0+\alpha]$. Even more is true: By [20, p. 25], ($*$) has a maximal solution u^0 on $[t_0, t_0+\alpha]$ in the sense that every other solution $u(t)$ of ($*$) defined on $[t_0, \delta]$ for some $\delta > 0$ satisfies $u(t) \leq u^0(t)$ for all $t \in [t_0, t_0 + \min\{\alpha, \delta\}]$. Now, assume that $v : [t_0, t_0+\alpha] \to \mathbb{R}$ is a continuous map which is right-sided differentiable on $(t_0, t_0+\alpha)$ satisfying $v(t_0) = p_0$ and

$$v'^{,+}(t) \leq F(v(t), t) \qquad \forall t \in (t_0, t_0+\alpha).$$

Under these hypotheses, it holds that

$$v(t) \leq u^0(t) \qquad \forall t \in [t_0, t_0+\alpha].$$

Proof. W.l.o.g. let $F \not\equiv 0$ (we could handle this easily with the previous Grønwall's lemma). By continuity of v, u^0, it suffices to prove the claim for every $0 < \alpha' < \alpha$. Consequently, let $0 < \alpha' < \alpha$ and $N \in \mathbb{N}$ such that

$$N \geq \frac{1+\alpha'}{b - \alpha'\,\|F\|_\infty},$$

which is well-defined because $\alpha \leq b\,\|F\|_\infty^{-1}$ and therefore $b \geq \alpha\,\|F\|_\infty > \alpha'\,\|F\|_\infty$. Then for any $n \geq N$, the IVP

$$u_n'(t) = F(u_n(t), t) + \frac{1}{n}, \qquad u_n(t_0) = p_0 + \frac{1}{n} \tag{$**$}$$

has a solution u_n on $[t_0, t_0+\alpha']$: Let $\tilde{t}_0 := t_0$, $\tilde{p}_0 := p_0 + \frac{1}{n}$, $\tilde{a} := a$, $\tilde{b} := b - \frac{1}{n}$. In our notation, we omit the n-dependence of the parameters for the sake of readability. We also define $F_n := F + \frac{1}{n}$ on $[t_0, t_0+\tilde{a}] \times [\tilde{p}_0 - \tilde{b}, \tilde{p}_0 + \tilde{b}]$ (for the corresponding n, of course). Observe that $\tilde{b} > 0$ because

$$\frac{1}{n} \leq \frac{1}{N} \leq \frac{b - \alpha'\,\|F\|_\infty}{1+\alpha'} = \frac{b}{1+\alpha'} - \frac{\alpha'\,\|F\|_\infty}{1+\alpha'} < b,$$

and that $[\tilde{t}_0, \tilde{t}_0+\tilde{a}] \times [\tilde{p}_0 - \tilde{b}, \tilde{p}_0 + \tilde{b}] = [t_0, t_0+a] \times [p_0 - b + \frac{2}{n}, p_0 + b] \subset [t_0, t_0+\alpha] \times [p_0 - b, p_0 + b]$. Consequently, F_n is well-defined and continuous with $\|F_n\|_\infty \leq \|F\|_\infty + \frac{1}{n}$. Using $\alpha' < \alpha \leq b\|F\|_\infty^{-1}$

and therefore $b^{-1}\left\|F\right\|_\infty < (\alpha')^{-1}$, we obtain

$$\tilde{b}\cdot\|F_n\|_\infty^{-1} = \frac{b-\frac{1}{n}}{\|F\|_\infty + \frac{1}{n}} \geq \frac{b-\frac{b-\alpha'\|F\|_\infty}{1+\alpha'}}{\|F\|_\infty + \frac{b-\alpha'\|F\|_\infty}{1+\alpha'}} = \frac{b+\alpha' b - b + \alpha'\left\|F\right\|_\infty}{\|F\|_\infty + \alpha'\left\|F\right\|_\infty + b - \alpha'\left\|F\right\|_\infty} = \alpha'.$$

This finally shows that $\alpha' \leq \min\left\{\tilde{a}, \tilde{b}\cdot\|F_n\|_\infty^{-1}\right\}$. Since $(**)$ is nothing but the IVP

$$u_n'(t) = F_n(u_n(t), t), \qquad u_n(\tilde{t}_0) = \tilde{p}_0,$$

Peano's existence theorem does indeed provide us a solution u_n on $[t_0, t_0+\alpha']$ for $n \geq N$.

By theorem I.2.4 in [20, p. 4] (which is itself basically a consequence of the Arzelà-Ascoli theorem, to be found on the very same page), there is a subsequence $(n_k)_{k\in\mathbb{N}}$ of $(n)_{n\geq N}$ and a solution $\overline{u}$ of $(*)$ such that

$$\sup\left\{\,|u_{n_k}(t) - \overline{u}(t)| : t \in [t_0, t_0+\alpha']\right\} \xrightarrow{k\to\infty} 0.$$

In particular, for all $t \in [t_0, t_0+\alpha']$, $u_{n_k}(t) \xrightarrow{k\to\infty} \overline{u}(t)$.

Recall that we want to prove $v(t) \leq u^0(t)$ for all $t \in (t_0, \alpha')$. Since $\overline{u}(t) \leq u^0(t)$, it suffices to show that $v(t) \leq \overline{u}(t)$ for all $t \in (t_0, t_0+\alpha')$, which in turn is true provided $v(t) \leq u_n(t)$ for all $n \geq N$ and all $t \in (t_0, t_0+\alpha')$. Assume for contradiction that there is some $n \geq N$ and some $t \in (t_0, t_0+\alpha')$ such that $v(t) > u_n(t)$. Let

$$\tau := \inf\left\{t \in (t_0, t_0+\alpha') : v(t) > u_n(t)\right\} \in [t_0, t_0+\alpha').$$

By proceeding as in the proof of theorem 4.3, we may conclude that $\tau > t_0$ and

$$v'^{,+}(\tau) \geq u_n'^{,+}(\tau) = F(u_n(\tau), \tau) + \frac{1}{n} = F(v(\tau), \tau) + \frac{1}{n} > F(v(\tau), \tau),$$

which yields the desired contradiction. □

Remark 4.6.

(a) Going through the proof of the previous theorem with $v := u^0$, one obtains that $u^0 \leq \overline{u}$, i.e. in fact we *constructed* the maximal solution $u^0 = \overline{u}$ as pointwise (even uniform) limit of the subsequence u_{n_k} in the proof. Moreover, we could apply the theorem to see that actually the whole sequence $(u_n)_{n\geq N}$ is non-increasing and therefore converges pointwise (even uniformly) to $u^0 = \overline{u}$.

(b) Grønwall's lemma is indeed a special case of theorem 4.5, with $F(z,t) := g(t)\cdot z + h(t)$ defined on $[-C, C]\times[t_0, T]$ for $C > 0$ large enough, since the unique (F is Lipschitz continuous in z) and therefore maximal solution $u(t)$ to the IVP

$$u'(t) = g(t)\cdot u(t) + h(t) = F(u(t), t), \qquad u(t_0) = u_0$$

on $[t_0, T]$ is given by the r.h.s. of (4.1).

(c) A less general version of this theorem, which requires right-sided differentiability of v with the corresponding bound on $v'^{,+0}$ also in t_0 and uniqueness of the solution for $(*)$, can be found in [38, p. 425 f.]. We are going to adapt the nomenclature and sometimes call theorem 4.5 *comparison theorem* from now on.

Let us finally mention one small adaption of theorem 4.5:

Corollary 4.7. *Under the hypotheses of theorem 4.5, let us slightly change the assumptions in the sense that we do not require u^0 to be a maximal solution anymore, and instead assume that v is right-sided differentiable also in t_0, with*

$$v^{\prime,+}(t) < F(v(t),t) \qquad \forall t \in [t_0, t_0+\alpha).$$

Then the conclusions of theorem 4.5 remain true.

Proof. Again, we assume for contradiction that for some solution u of the IVP $(*)$ and some $t \in (t_0, t_0+\alpha)$, $v(t) > u(t)$. Defining τ accordingly, this time we only get $\tau \in [t_0, t_0+\alpha)$, and we arrive at the contradiction

$$v^{\prime,+}(\tau) \geq u^{\prime,+}(\tau) = F(u(\tau),\tau) = F(v(\tau),\tau). \qquad \square$$

The following theorem provides an analogue for the mean value theorem of differentiation for right-sided differentiable functions. One can actually prove it from scratch (see e.g. [27]), however, with our general version of Grønwall's lemma at hand, the proof can be given in a much shorter way:

Theorem 4.8. *Let $g : [a,b] \to \mathbb{R}$ be continuous and right-sided differentiable on (a,b), then there are $\xi, \xi' \in (a,b)$ such that*

$$g^{\prime,+}(\xi) \leq \frac{g(b)-g(a)}{b-a} \leq g^{\prime,+}(\xi').$$

In particular, for all $s,t \in [a,b]$ with $s > t$,

$$\inf\left\{g^{\prime,+}(\tau) : \tau \in (t,s)\right\} \leq \frac{g(s)-g(t)}{s-t} \leq \sup\left\{g^{\prime,+}(\tau) : \tau \in (t,s)\right\}.$$

Proof. We first claim that if g satisfies $g(a) = g(b) = 0$, then there is some $\xi \in (a,b)$ with $g^{\prime,+}(\xi) \geq 0$. Indeed, assume for contradiction that $g^{\prime,+}(t) < 0$ for all $t \in (a,b)$. Applying corollary 4.4 with $g \equiv h \equiv 0$, we obtain $g(t) \leq g(a) = 0$ for all $t \in [a,b]$. However, $g \equiv 0$ on $[a,b]$ yields $g^{\prime,+}(t) = 0$ for all $t \in (a,b)$ and hence cannot be the case. Thus, we find some $\xi \in (a,b)$ with $g(\xi) < 0$. Using corollary 4.4 again, this time on the interval $[\xi, b]$, now shows that $g(b) \leq g(\xi) < 0$, yielding the desired contradiction. Note that by considering $-g$ instead, we see that there is also some $\xi' \in (a,b)$ with $g^{\prime,+}(\xi') \geq 0$.

Inspired by the proof of the mean value theorem for differentiable functions, we now define

$$\eta : [a,b] \to \mathbb{R}, \qquad t \mapsto g(t) - g(a) - \frac{g(b)-g(a)}{b-a} \cdot (t-a).$$

Then obviously $\eta(a) = \eta(b) = 0$, and η is continuous and right-sided differentiable on (a,b) with $\eta^{\prime,+}(t) = g^{\prime,+}(t) - \frac{g(b)-g(a)}{b-a}$ (see the discussion in remark 4.2 (f)). Hence, from the initial claim we obtain existence of some $\xi \in (a,b)$ such that $\eta^{\prime,+}(\xi) \leq 0$, which shows the inequality $g^{\prime,+}(\xi) \leq \frac{g(b)-g(a)}{b-a}$. Likewise for the other inequality. $\square$

From this version of the mean value theorem (or, alternatively, from theorem 4.5), we can prove the (intuitively clear) statement that maps with smaller right-sided derivative *do not outrun* maps with larger right-sided derivative:

Corollary 4.9. *Let $f, g : [a,b] \to \mathbb{R}$ be continuous and right-sided differentiable on (a,b) with $f(a) \leq g(a)$ and $f'^{,+}(t) \leq g'^{,+}(t)$ for all $t \in (a,b)$. Then $f(t) \leq g(t)$ for all $t \in [a,b]$.*

Proof. Define $h(t) := f(t) - g(t)$ on $[a,b]$, then h is continuous, right-sided differentiable on (a,b), and

$$h'^{,+}(t) = f'^{,+}(t) - g'^{,+}(t) \leq 0 \qquad \forall t \in (a,b).$$

By theorem 4.8, for all $s,t \in [a,b]$ with $s > t$,

$$h(s) - h(t) \leq (s-t) \cdot \sup\left\{h'^{,+}(\tau) : \tau \in (t,s)\right\} \leq 0.$$

Hence, h is non-increasing, and with $h(a) = f(a) - g(a) \leq 0$, we obtain $h \leq 0$, which yields $f \leq g$. $\square$

Another important application of theorem 4.8 is that it allows us to prove interchangeability of right-sided differentiation and integration under quite general assumptions:

Theorem 4.10. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $I \subset \mathbb{R}$ right-open and $f : \Omega \times I \to \mathbb{R}$ such that*

(i) $f(\,\cdot\,, t) \in \mathcal{L}^1(\Omega, \mathrm{d}\mu)$ for every $t \in I$,

(ii) $f(\omega, \,\cdot\,) : I \to \mathbb{R}$ is right-sided differentiable for all $\omega \in \Omega$, and

(iii) there is some $g \in \mathcal{L}^1(\Omega, \mathrm{d}\mu)$ such that for all $\omega \in \Omega$, there exists $\delta_\omega > 0$ with

$$\sup\left\{\left|\partial_s^+ f(\omega, s)\right| : s \in [t, t+\delta_\omega] \cap I\right\} \leq g(\omega).$$

Then the map $I \to \mathbb{R},\ t \mapsto \int_\Omega f(\omega,t)\,\mathrm{d}\mu(\omega)$ is right-sided differentiable with

$$\partial_t^+ \left(\int_\Omega f(\omega,t)\,\mathrm{d}\mu(\omega)\right) = \int_\Omega \partial_t^+ f(\omega,t)\,\mathrm{d}\mu(\omega) \qquad \forall t \in I.$$

Proof. Let $t \in I$ and fix $\omega \in \Omega$. W.l.o.g. we may assume that $[t, t+\delta_\omega] \subset I$ (I is right-open!). Let $(h_n)_{n\in\mathbb{N}} \subset (0, \delta_\omega]$ with $h_n \searrow 0$. By theorem 4.8, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\left|\frac{1}{h_n}\left(f(\omega, t+h_n) - f(\omega,t)\right)\right| &\leq \sup\left\{\left|\partial_s^+ f(\omega,s)\right| : s \in [t, t+h_n]\right\} \\
&\leq \sup\left\{\left|\partial_s^+ f(\omega,s)\right| : s \in [t, t+\delta_\omega] \cap I\right\} \\
&\leq g(\omega).
\end{aligned}$$

It follows that g is an integrable majorant for the sequence of measurable functions $(g_n)_{n\in\mathbb{N}}$, where $g_n := \frac{1}{h_n}\left(f(\,\cdot\,, t+h_n) - f(\,\cdot\,,t)\right)$. By linearity of the integral and the dominated convergence theorem (cf. [29, p. 19]), it follows that

$$\begin{aligned}
\partial_t^+ \left(\int_\Omega f(\omega,t)\,\mathrm{d}\mu(\omega)\right) &= \lim_{n\to\infty} \int_\Omega \frac{1}{h_n}\big(f(\omega, t+h_n) - f(\omega,t)\big)\,\mathrm{d}\mu(\omega) \\
&= \int_\Omega \lim_{n\to\infty} \frac{1}{h_n}\big(f(\omega, t+h_n) - f(\omega,t)\big)\,\mathrm{d}\mu(\omega) \\
&= \int_\Omega \partial_t^+ f(\omega,t)\,\mathrm{d}\mu(\omega).
\end{aligned}$$ $\square$

Finally in this section, we prove that a version of the fundamental theorem of calculus still holds for one-sided differentiable funtioos. However, in advance we must ensure that integrating a right-continuous function makes sense:

Lemma 4.11. *Let $f : \mathbb{R} \to \mathbb{R}$ be right-continuous. Then f is (Borel-) measurable.*

Proof. From measure theory, we know that it suffices to find a sequence of simple functions $(f_n)_{n\in\mathbb{N}}$ which converges to f pointwise (see e.g. [8, p. 106]). Let us define

$$f_n := \sum_{i=-n^2-1}^{n^2+1} f\left(\tfrac{i+1}{n}\right) \cdot \mathbb{1}_{\left[\frac{i}{n}, \frac{i+1}{n}\right)}, \qquad n \in \mathbb{N}.$$

Then $(f_n)_{n\in\mathbb{N}}$ is obviously a sequence of simple functions. Note that for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, $x \in \left[\frac{\lfloor nx\rfloor}{n}, \frac{\lfloor nx\rfloor+1}{n}\right)$ because of the obvious inequality $\lfloor y\rfloor \le y < \lfloor y\rfloor + 1$ for $y \in \mathbb{R}$, i.e. $x \in \left[\frac{i}{n}, \frac{i+1}{n}\right)$ for $i = \lfloor nx\rfloor$. Hence for $n \ge |x|$, $|\lfloor nx\rfloor| \le n^2 + 1$ (distinguish the cases $x \ge 0$ and $x < 0$), and thus

$$|f_n(x) - f(x)| = \left|f\left(\tfrac{\lfloor nx\rfloor+1}{n}\right) - f(x)\right| \xrightarrow{n\to\infty} 0$$

by right-continuity because $\frac{\lfloor nx\rfloor+1}{n} \searrow x$ as $n \to \infty$. ☐

For a right-continuous map f defined on an interval, it is easy to extend f to a right-continuous and hence measurable map $\overline{f}$ defined on $\mathbb{R}$, and hence f is measurable as restriction of a measurable function to a measurable set. Consequently, the following makes sense:

Lemma 4.12 (Fundamental theorem of calculus analogue for right-sided differentiable maps)**.** *Let $a, b \in \mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ be bounded and right-continuous on $[a, b)$. Define*

$$F : [a, b] \to \mathbb{R}, \quad t \mapsto \int_a^t f(s)\,\mathrm{d}s := \int_{[a,t]} f(s)\,\mathrm{d}s.$$

Then F is continuous and right-sided differentiable on $[a, b)$ with right-sided derivative $F^{\prime,+} = f$.

Proof. The expression on the r.h.s is well-defined because f is bounded and right-continuous and therefore measurable by the previous discussion; in particular, it is integrable on bounded intervals contained in $[a, b]$. Let us first prove continuity of F: for $t \in [a, b]$ and $(h_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ such that $(h_n)_{n\in\mathbb{N}} \to 0$ and $t + h_n \in [a, b]$ for all $n \in \mathbb{N}$, define $f_n := f \cdot \mathbb{1}_{[t,t+h_n]}$, then $f_n \xrightarrow{n\to\infty} 0$ pointwise almost everywhere and $|f_n| \le |f| \cdot \mathbb{1}_{[a,b]}$ for all $n \in \mathbb{N}$, therefore (f is integrable!) showing that $|f| \cdot \mathbb{1}_{[a,b]}$ is an integrable majorant. It follows by dominated convergence that

$$\begin{aligned}\lim_{n\to\infty} F(t + h_n) &= \lim_{n\to\infty} \int_a^{t+h_n} f(s)\,\mathrm{d}s = \lim_{n\to\infty} \left(\int_a^t f(s)\,\mathrm{d}s + \int_{\mathbb{R}} f_n(s)\,\mathrm{d}s\right) \\ &= F(t) + \int_{\mathbb{R}} 0\,\mathrm{d}s = F(t).\end{aligned}$$

Note that we used that integrals over singletons are zero because $\{t\}$ is a Lebesgue null set for every $t \in \mathbb{R}$. Now, let $t \in (a, b)$ and $\varepsilon > 0$. Choose $\delta > 0$ such that for all $s \in [t, t + \delta]$, $|f(s) - f(t)| < \varepsilon$; this is possible by right-continuity of f (see remark 4.2 (iii)). Then for all $0 < h < \delta$,

$$\left|\frac{1}{h}\big(F(t+h) - F(t)\big) - h \cdot f(t)\right| = \left|\frac{1}{h}\left(\int_t^{t+h} (f(s) - f(t))\,\mathrm{d}s\right)\right| \le \frac{1}{h}\int_t^{t+h} |f(s) - f(t)|\,\mathrm{d}s \le \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this yields the claim. ☐

Remark 4.13. Right-continuity alone does not imply boundedness and/or integrability, as can be seen from considering the function

$$f : [-1, 1] \to \mathbb{R}, \qquad t \mapsto \sum_{n=1}^{\infty} n \cdot \mathbb{1}_{[-\frac{1}{n}, -\frac{1}{n+1})}(t),$$

which is obviously right-continuous on $(-1, 1)$, but not bounded/integrable on any neighbourhood of $0 \in \mathbb{R}$.

Lemma 4.14. *Let $a, b \in \mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ be continuous and right-sided differentiable on (a, b) with right-continuous and bounded right-sided derivative. Then for $t \in [a, b]$,*

$$f(t) = f(a) + \int_a^t f'^{,+}(s) \, \mathrm{d}s. \tag{4.4}$$

Proof. Let $g : [a, b] \to \mathbb{R}$ denote the r.h.s. of (4.4). Then f and g are continuous (for g this follows from lemma 4.12) and obviously agree for $t = a$. Again by lemma 4.12, g is also right-sided differentiable on (a, b) with $g'^{,+} = f'^{,+}$, i.e. for $h := f - g$, we have $h'^{,+} \equiv 0$ on (a, b). By applying Grønwall's theorem or the weak mean value theorem to both h and $-h$, we find that $h \equiv 0$ and consequently $f \equiv g$ on $[a, b]$. □

4.2 Some one-sided differentiable maps

In this section, we want to prove that for a large class of curves γ, the maps $t \mapsto |\gamma(t)|$ and $t \mapsto \sup_{a \leq s \leq t} \gamma(s)$ are right-sided differentiable. Moreover, we calculate bounds on the right-sided derivatives in terms of the right-sided derivative of γ. This makes the results from the previous section applicable to many proofs in the main text.

Lemma 4.15. *Let $I \subset \mathbb{R}$ be right-open and $f, g : I \to \mathbb{R}$ continuous and right-sided differentiable with right-continuous right-sided derivatives. Define*

$$h : I \to \mathbb{R}, \qquad t \mapsto \max\{f(t), g(t)\}.$$

Then h is also continuous and right-sided differentiable with right-continuous right-sided derivative, and

$$h'^{,+}(t) = \begin{cases} f'^{,+}(t), & f(t) > g(t) \\ g'^{,+}(t), & f(t) < g(t) \\ \max\{f'^{,+}(t), g'^{,+}(t)\}, & f(t) = g(t) \end{cases} \qquad \forall t \in I.$$

Proof. Let $t \in I$. We need to distinguish 3 cases:

(i) If $f(t) \neq g(t)$ we may assume w.l.o.g. that $f(t) < g(t)$. Then by right-openness of I and continuity of f and g, there is some $\delta > 0$ such that $f(s) < g(s)$ for all $s \in [t, t + \delta] \subset I$. Hence, for all these s, $h(s) = g(s)$ and therefore (*locality* of the right-sided derivative) clearly $h'^{,+}(t) = g'^{,+}(t)$. Note that this reasoning also shows that $h'^{,+}(s) = g'^{,+}(s)$ for all $s \in [t, t+\delta)$, so we also obtain right-continuity of $h'^{,+}$ in t.

(ii) Next, consider the case $f(t) = g(t)$ and $f'^{,+}(t) \neq g'^{,+}(t)$. W.l.o.g. let $f'^{,+}(t) < g'^{,+}(t)$. By right-continuity of the right-sided derivatives, there is some $\delta > 0$ such that $f'^{,+}(s) < g'^{,+}(s)$

for all $s \in [t, t+\delta] \subset I$. From corollary 4.9, we deduce that $f(s) \leq g(s)$ for all $s \in [t, t+\delta]$, i.e. $h(s) = g(s)$ for these s. Again, it follows that $h'^{,+}(t) = g'^{,+}(t)$ and $h'^{,+}(s) = g'^{,+}(s)$ for $s \in [t, t+\delta)$, i.e. $h'^{,+}$ is right-continuous in t.

(iii) Finally, suppose that $f(t) = g(t)$ and $f'^{,+}(t) = g'^{,+}(t)$. Then for $s > t$,

$$\begin{aligned} f(s) &= f(t) + f'^{,+}(t) \cdot (s-t) + r_1(s-t), \\ g(s) &= g(t) + g'^{,+}(t) \cdot (s-t) + r_2(s-t) = f(t) + f'^{,+}(t) \cdot (s-t) + r_2(s-t), \end{aligned}$$

where $r_1, r_2 \in o(h^+)$ (see remark 4.2 (e)). This shows that for every $s > t$,

$$h(s) = f(t) + f'^{,+}(t) \cdot (s-t) + \tilde{r}(s-t),$$

where $\tilde{r}(s-t) := \max\{r_1(s-t), r_2(s-t)\}$. It follows that $\tilde{r} \in o(h^+)$, so by uniqueness of the right-sided derivative we obtain that $h'^{,+}(t) = f'^{,+}(t) = g'^{,+}(t)$. But we have already seen that $h'^{,+}(s) \in \{f'^{,+}(s), g'^{,+}(s)\}$ for all $s \in I$ (in any of the three distinguished cases), so it follows from right-continuity of $f'^{,+}$ and $g'^{,+}$ that $h'^{,+}$ is right-continuous in t. □

Lemma 4.16. *Let $I \subset \mathbb{R}$ be right-open and $f_1, \ldots, f_n : I \to \mathbb{R}$ be right-sided differentiable with right-continuous right-sided derivatives. Then*

$$h : I \to \mathbb{R}, \qquad t \mapsto \max\{f_i(t) : i \in [n]\}$$

is right-sided differentiable with right-continuous right-sided derivative, and

$$h'^{,+}(t) = \max\left\{f_i'^{,+}(t) : i \in [n],\ f_i(t) = h(t)\right\} \qquad \forall t \in I.$$

Proof. For $n = 1$, there is nothing to prove, and for $n = 2$, the assertion is precisely the statement of lemma 4.15, written in an aggregate form. Hence we only need to execute the induction step $n \to n+1$. Let us assume that we have shown the assumption for $n \geq 2$. Then since $\max\{f_i : i \in [n]\}$ and f_{n+1} are right-sided differentiable with right-continuous right-sided derivatives, we can again apply lemma 4.15 and obtain

$$\begin{aligned} h'^{,+}(t) &= \partial_t^+ \max\left\{\max\{f_i(t) : i \in [n]\}, f_{n+1}(t)\right\} \\ &= \begin{cases} \partial_t^+ \max\{f_i(t) : i \in [n]\}, & \max\{f_i(t) : i \in [n]\} > f_{n+1}(t) \\ f_{n+1}'^{,+}(t), & \max\{f_i(t) : i \in [n]\} < f_{n+1}(t) \\ \max\left\{\partial_t^+ \max\{f_i(t) : i \in [n]\}, f_{n+1}'^{,+}(t)\right\}, & \max\{f_i(t) : i \in [n]\} = f_{n+1}(t) \end{cases} \\ &= \max\left\{f_i'^{,+}(t) : f_i(t) = h(t)\right\}. \end{aligned}$$

□

Corollary 4.17. *Under the hypotheses of lemma 4.16, define*

$$h : I \to \mathbb{R}, \qquad t \mapsto \min\{f_i(t) : i \in [n]\}.$$

Then h is right-sided differentiable with right-continuous right-sided derivative, and

$$h'^{,+}(t) = \min\left\{f_i'^{,+}(t) : i \in [n],\ f_i(t) = h(t)\right\} \qquad \forall t \in I.$$

Proof. By lemma 4.16, using the identity $\inf A = -\sup(-A)$ for $A \subset \mathbb{R}$, we obtain that for $t \in I$,

$$\begin{aligned} h'^{,+}(t) &= \partial_t^+ \min\{f_i(t) : i \in [n]\} = \partial_t^+\left(-\max\{-f_i(t) : i \in [n]\}\right) \\ &= -\max\left\{-f_i'^{,+}(t) : i \in [n],\ -f_i(t) = \max\left\{-f_i'^{,+}(t) : i \in [n]\right\}\right\} \\ &= \min\left\{f_i'^{,+}(t) : i \in [n],\ f_i(t) = h(t)\right\}. \end{aligned}$$

□

We want to apply all that we have seen yet to the maximum norm on $\mathbb{R}^d$:

Lemma 4.18. *Let $I \subset \mathbb{R}$ be right-open and $f_1, \ldots, f_d : I \to \mathbb{R}$ be right-sided differentiable functions with right-continuous right-sided derivatives. Define $f := (f_1, \ldots, f_d) : I \to \mathbb{R}^d$. Then*

$$h : I \to \mathbb{R}, \qquad t \mapsto |f(t)| := \max\{|f_i(t)| : i \in [d]\}$$

is right-sided differentiable with right-continuous right-sided derivative, and

$$\left|h'^{,+}(t)\right| \leq \max\left\{\left|f_i'^{,+}(t)\right| : |f_i(t)| = h(t)\right\}.$$

Proof. Apply lemma 4.16 to $f_1, \ldots, f_{2d}$, where $f_{d+i} := -f_i$ for $i \in [d]$. □

We will typically need this lemma in the following form:

Corollary 4.19. *Let $I \subset \mathbb{R}$ be open and $\gamma : (a,b) \to \mathbb{R}^d$ a $\mathcal{C}^1$-curve. Then*

$$\partial_t^+ |\gamma(t)| \leq |\dot{\gamma}(t)| \qquad \forall t \in (a,b).$$

Proof. Clear. □

Remark 4.20. The corollary can also be found in [38, p. 424], without requiring as much preliminary work for the proof as we did here. However, in the course of the main text we also need the more general version of lemma 4.15 for right-sided differentiable curves with right-continuous right-sided derivatives, namely for the stochastic process J_ω appearing in section 3.2, so it makes sense to use the approach we pursued here.

Finally, let us check right-sided differentiability for one further class of functions:

Lemma 4.21. *Let $I := [a,b] \subset \mathbb{R}$ a closed interval and $g : I \to \mathbb{R}$ continuous and right-sided differentiable on $[a,b)$ with right-continuous right-sided derivative. Let*

$$h : [a,b] \to \mathbb{R}, \quad t \mapsto \sup\{g(s) : a \leq s \leq t\}.$$

Then h is continuous. Moreover, it is right-sided differentiable on $[a,b)$ with

$$h'^{,+}(t) = \begin{cases} 0, & g(t) < h(t) \\ \max\{0, g'^{,+}(t)\}, & g(t) = h(t) \end{cases}.$$

In particular,

$$0 \leq h'^{,+}(t) \leq \max\left\{0, g'^{,+}(t)\right\} \qquad \forall t \in [a,b).$$

Proof. Note that h is well-defined (finite) and the supremum is actually a maximum since g is continuous and $[a,t] \subset \mathbb{R}$ is compact and non-empty for all $t \in [a,b]$. We also observe that h is non-decreasing with $h(t) \geq g(t)$ for all $t \in [a,b]$.

First, we show left-continuity of h on $(a,b]$. Afterwards, we prove that h is right-sided differentiable on $[a,b)$; since this implies that h is right-continuous on $[a,b)$, we also obtain continuity. For the left-continuity of h, let $t \in (a,b]$. We distinguish between two cases:

(1) If $g(t) = h(t)$, given $\varepsilon > 0$ choose $\delta > 0$ such that for all $s \in I$ with $|t-s| \leq \delta$, $|g(t) - g(s)| \leq \varepsilon$. Then for all $s \in [t-\delta, t] \cap I$,

$$h(s) \geq g(s) \geq g(t) - \varepsilon = h(t) - \varepsilon.$$

Since h is non-decreasing, also $h(s) \leq h(t)$ holds, so altogether $|h(t) - h(s)| \leq \varepsilon$ for all $s \in [t-\delta, t] \cap I$.

(2) In case $g(t) < h(t)$, we can find some $\delta > 0$ such that for all $s \in I$ with $|t-s| \leq \delta$, $|g(t) - g(s)| \leq \frac{1}{2}(h(t) - g(t))$. Consequently,

$$\sup \left\{ g(s) : t - \delta \leq s \leq t \right\} \leq g(t) + \tfrac{1}{2}(h(t) - g(t)) = h(t) - \tfrac{1}{2}(h(t) - g(t)) < h(t).$$

Therefore,

$$h(t) = \max \left\{ h(t-\delta), \sup \left\{ g(s) : t - \delta \leq s \leq t \right\} \right\} = h(t-\delta).$$

But h is non-decreasing, so $h(s) = h(t-\delta)$ for all $s \in [t-\delta, t] \cap I$; in particular, h is left-continuous in t.

Now, let us show that h is right-sided differentiable on the right-open interval $[a, b)$. This time, we distinguish between four cases:

(i) For the case $g(t) < h(t)$, by right-openness of $[a, b)$ and continuity of g there is some $\delta > 0$ such that $g(\tau) < h(t)$ for all $\tau \in [t, t+\delta] \subset [a, b)$. Consequently, $\sup \{g(s) : t \leq s \leq \tau\} \leq h(t)$ and therefore

$$h(\tau) = \max \left\{ h(t), \sup \left\{ g(s) : t \leq s \leq \tau \right\} \right\} = h(t) \qquad \forall \tau \in [t, t+\delta].$$

This also implies that $h'^{,+} \equiv 0$ on $[t, t+\delta)$.

(ii) If $g(t) = h(t)$ and $g'^{,+}(t) < 0$, by right-continuity of $g'^{,+}$, there is some $\delta > 0$ such that for all $s \in [t, t+\delta] \subset [a, b)$, $g'^{,+}(s) \leq 0$. Hence, by corollary 4.9, $g(s) \leq g(t)$ for all $s \in [t, t+\delta]$, showing that $\sup \{g(s) : t \leq s \leq t+\delta]\} \leq g(t) = h(t)$ and thus

$$h(\tau) = \max \left\{ h(t), \sup \left\{ g(s) : t \leq s \leq \tau \right\} \right\} = h(t) \qquad \forall \tau \in [t, t+\delta].$$

It follows that $h'^{,+} \equiv 0$ on $[t, t+\delta)$.

(iii) For $g(t) = h(t)$ and $g'^{,+}(t) > 0$, by right-continuity of $g'^{,+}$ there are some $\eta, \delta > 0$ such that $g'^{,+} > \eta$ on $[t, t+\delta] \subset [a, b)$. With theorem 4.8, it follows that g is strictly increasing on $[t, t+\delta]$, showing that $\sup \{g(s) : t \leq s \leq \tau\} = g(\tau)$ and thus $h(\tau) = g(\tau)$ for all $\tau \in [t, t+\delta] \subset [a, b)$. This proves that $h'^{,+} \equiv g'^{,+}$ on $[t, t+\delta)$.

(iv) Finally, assume that $g(t) = h(t)$ and $g'^{,+}(t) = 0$. Then for $s \in I$ with $s > t$, we have

$$g(s) = g(t) + g'^{,+}(t) \cdot (s-t) + r(s-t) = g(t) + r(s-t),$$

where $r \in o(h^+)$. Let $\varepsilon > 0$ be arbitrary and choose $\delta > 0$ such that $[t, t+\delta] \subset [a, b)$ and $\left| \frac{r(s-t)}{s-t} \right| < \varepsilon$ for all $s \in (t, t+\delta]$. Then for $\tau \in (t, t+\delta]$,

$$\sup \left\{ g(s) : t < s \leq \tau \right\} \leq \sup \left\{ g(t) + (s-t) \cdot \left| \tfrac{r(s-t)}{s-t} \right| : t < s \leq \tau \right\} \leq h(t) + \varepsilon(\tau - t).$$

Consequently

$$h(t) \leq h(\tau) = \max\{h(t), \sup\{g(s) : t < s \leq \tau\}\} \leq h(t) + \varepsilon(\tau - t) \qquad \forall \tau \in [t, t+\delta].$$

It follows that for every $a \in (0, \delta]$,

$$\left| \frac{1}{a} \big(h(t+a) - h(a)\big) \right| \leq \frac{1}{a} \cdot (\varepsilon a) = \varepsilon,$$

which proves that $h'^{,+}(t) = 0 = g'^{,+}(t)$ since $\varepsilon > 0$ was arbitrary and the right-sided derivative is unique. But we have seen in all four cases that $h'^{,+}(\tau) \in \{0, g'^{,+}(\tau)\}$ for all $\tau \in [a, b)$, which by right-continuity of $g'^{,+}$ proves that $g'^{,+}$ is right-continuous in t in this case as well. □

4.3 Aspects in measure theory

In this section, we collect and derive concepts and results related to measure theory which put methods from the main text into a more general or abstract setting and thus turn out to be useful there because it gives some structure to the involved thoughts.

Definition 4.22 (Pushforward measure)**.** *Let $(\Omega_1, \mathcal{A}_1, \mu)$ a measure space, $(\Omega_2, \mathcal{A}_2)$ a measurable space and $f : \Omega_1 \to \Omega_2$ a measurable map. We define the **pushforward** or **image measure** of μ under f as*

$$f\#\mu := \mu \circ f^{-1} : \mathcal{A}_2 \to [0, \infty), \qquad A \mapsto \mu(f^{-1}(A)).$$

Lemma 4.23. *Under the hypotheses of lemma 4.22, $f\#\mu$ is indeed a measure on $(\Omega_2, \mathcal{A}_2)$. Moreover, for every $g : \Omega_2 \to \mathbb{R}$ measurable, $g \in \mathcal{L}^1(\Omega_2; \mathrm{d}(f\#\mu))$ if and only if $g \circ f \in \mathcal{L}^1(\Omega_1; \mathrm{d}\mu)$, and in this case,*

$$\int_{\Omega_2} g \,\mathrm{d}(f\#\mu) = \int_{\Omega_2} g \,\mathrm{d}(\mu \circ f^{-1}) = \int_{\Omega_1} (g \circ f) \,\mathrm{d}\mu. \tag{4.5}$$

Proof. See [8, p. 190 f.]. □

Definition 4.24. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and $X : \Omega \to \mathbb{R}^d$ a random variable. We say that X has **probability density u** (w.r.t. Lebesgue measure) if and only if the pushforward measure $X\#\mathbb{P}$ has probability density u w.r.t. Lebesgue measure, i.e. if $u \in L^1(\mathbb{R}^d)$ and for any $B \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mathbb{P}_X(B) := (X\#\mathbb{P})(B) = \int_B u(x) \,\mathrm{d}x.$$

From measure theory, it is well-known that for μ a measure on $\mathbb{R}^d$ which has a density u w.r.t. Lebesgue measure and $g : \mathbb{R}^d \to \mathbb{R}$ measurable, $g \in \mathcal{L}^1(\mathbb{R}^d; \mathrm{d}\mu)$ if and only if $g \cdot u \in L^1(\mathbb{R}^d)$, and in this case,

$$\int_{\mathbb{R}^d} g \,\mathrm{d}\mu = \int_{\mathbb{R}^d} g \cdot u \,\mathrm{d}x.$$

The proof is a standard argument, called *algebraic induction*: first observe that by definition this is true for indicator functions, then generalize to simple functions by linearity, and finally use monotone convergence and splitting into positive and negative part to show the claim for integrable functions. We will often use this result without further mentioning in the sequel.

Our first important observation is a tight relationship between the expectation value of functions of translates of a random variable and convolutions. In case the reader is not yet familiar with convolutions or the notation, she should consider having a close look at definition 4.49.

Lemma 4.25. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X : \Omega \to \mathbb{R}^d$ a random variable and $h : \mathbb{R}^d \to \mathbb{R}$ measurable. For $y \in \mathbb{R}^d$, define*

$$h_y : \mathbb{R}^d \to \mathbb{R}, \qquad x \mapsto h(y - x).$$

Then the random variable $h_y \circ X$ has an expectation value if and only if $h_y \in \mathcal{L}^1(\Omega; \mathrm{d}(X\#\mathbb{P}))$, and in this case,

$$\mathbb{E}\left[h_y \circ X\right] = (h * (X\#\mathbb{P}))(y). \tag{4.6}$$

In particular, if X has probability density u, then

$$\mathbb{E}\left[h_y \circ X\right] = (h * u)(y). \tag{4.7}$$

Proof. Note that $h_y \circ X$ is indeed measurable because we can write it as composition of the two measurable maps h, X and a (continuous and therefore measurable) euclidean motion. With lemma 4.23, we conclude that the l.h.s. of (4.6) is defined if and only if $h_y \in \mathcal{L}^1(\Omega; \mathrm{d}(X\#\mathbb{P}))$, and under these circumstances,

$$\mathbb{E}\left[h_y \circ X\right] = \int_\Omega h_y \circ X \,\mathrm{d}\mathbb{P} = \int_{\mathbb{R}^d} h_y \,\mathrm{d}(X\#\mathbb{P}) = \int_{\mathbb{R}^d} h(y - x) \,\mathrm{d}(X\#\mathbb{P})(x) = (h * (X\#\mathbb{P}))(y).$$

This clearly implies (4.7), too. □

Next, we want to consider *marginals*, which we need to formalize the operation $*_q$ occurring in the main text.

Definition 4.26. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and assume that there are measurable spaces $(\Omega_i, \mathcal{A}_i)$, $i \in \{1, 2\}$ such that $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$, where*

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma\left(\{B_1 \times B_2 : B_1 \in \mathcal{A}_1, B_2 \in \mathcal{A}_2\}\right)$$

*is the σ-algebra generated by measurable rectangles, called **product-$\boldsymbol{\sigma}$-algebra**. Then*

$$\pi_1(\mu) : \Omega_1 \to [0, \infty], \quad B_1 \mapsto \mu(B_1 \times \Omega_2), \qquad \pi_2(\mu) : \Omega_2 \to [0, \infty], \quad B_2 \mapsto \mu(\Omega_1 \times B_2)$$

*are easily checked to be measures on Ω_1 resp. Ω_2. We call them **first** resp. **second marginals of $\boldsymbol{\mu}$**.*

Lemma 4.27. *Under the hypotheses of definition 4.26, let $f \in \mathcal{L}^1(\Omega_1; \mathrm{d}(\pi_1(\mu)))$ and define*

$$\tilde{f} : \Omega = \Omega_1 \times \Omega_2 \to \mathbb{R}, \qquad (\omega_1, \omega_2) \mapsto f(\omega_1).$$

Then $\tilde{f} \in \mathcal{L}^1(\Omega; \mathrm{d}\mu)$, and

$$\int_\Omega \tilde{f} \,\mathrm{d}\mu = \int_{\Omega_1} f \,\mathrm{d}(\pi_1(\mu)).$$

Proof. First note that $\tilde{f}$ is $\mathcal{A}$-measurable because for any $B \subset \mathbb{R}$ measurable,

$$\tilde{f}^{-1}(B) = f^{-1}(B) \times \Omega_2 \in \mathcal{A}_1 \times \mathcal{A}_2 \subset \mathcal{A}_1 \otimes \mathcal{A}_2.$$

For $B \in \mathcal{A}_1$ and $f = \mathbb{1}_B$ an indicator function on Ω_1, we have

$$\tilde{f}(\omega_1, \omega_2) = \mathbb{1}_B(\omega_1) = \mathbb{1}_{B\times\Omega_2}(\omega_1, \omega_2) \qquad \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 = \Omega$$

and hence

$$\int_\Omega \tilde{f} \,\mathrm{d}\mu = \mu(B \times \Omega_2) = \big(\pi_1(\mu)\big)(B) = \int_{\Omega_1} \mathbb{1}_B(\omega_1) \,\mathrm{d}(\pi_1(\mu))(\omega_1) = \int_{\Omega_1} f \,\mathrm{d}(\pi_1(\mu)).$$

By algebraic induction, the claim follows. $\square$

Lemma 4.28. *Let $d = d_1 + d_2$ and μ be a finite measure on $\mathbb{R}^d = \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}$ which has Lebesgue density $u \in L^1(\mathbb{R}^d)$. Then $\pi_1(\mu)$ has Lebesgue density $\rho \in L^1(\mathbb{R}^{d_1})$, where*

$$\rho(q) := \int_{\mathbb{R}^{d_2}} u(q,p) \,\mathrm{d}p \qquad \textit{for a.e. } q \in \mathbb{R}^{d_1}. \tag{4.8}$$

Proof. By Fubini's theorem (cf. [8, p. 185]), $\rho \in L^1(\mathbb{R}^{d_1})$ is finite a.e., and for $B \in \mathcal{B}(\mathbb{R}^{d_1})$,

$$\begin{aligned}\big(\pi_1(\mu)\big)(B) = \mu(B \times \mathbb{R}^{d_2}) &= \int_{\mathbb{R}^d} \mathbb{1}_{B\times\mathbb{R}^{d_2}} \cdot u \,\mathrm{d}x = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} \mathbb{1}_{B\times\mathbb{R}^{d_2}}(q,p) \cdot u(q,p) \,\mathrm{d}p \right) \mathrm{d}q \\ &= \int_{\mathbb{R}^{d_1}} \mathbb{1}_B(q) \cdot \left(\int_{\mathbb{R}^{d_2}} u(q,p) \,\mathrm{d}p \right) \mathrm{d}q = \int_B \rho(q) \,\mathrm{d}q. \end{aligned}$$ $\square$

We will mainly need the previous results in the context of random variables:

Corollary 4.29. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ a measure space and $X : \Omega \to \mathbb{R}^d$ a random variable with probability density u. Let $d = d_1 + d_2$ as above and define the random variable*

$$\pi_1 \circ X : \Omega \to \mathbb{R}^{d_1}, \qquad \omega \mapsto \pi_1(X(\omega)),$$

where $\pi_1 : \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$, $(q,p) \mapsto q$. Then $\pi_1(X)$ is distributed by the law $\pi_1(X\#\mathbb{P})$. In particular, it has probability density ρ w.r.t. Lebesgue measure, where ρ is defined by (4.8)*.*

Proof. For $B \in \mathcal{B}(\mathbb{R}^{d_1})$,

$$\begin{aligned}((\pi_1 \circ X)\#\mathbb{P})(B) &= (\mathbb{P} \circ (\pi_1 \circ X)^{-1})(B) = \big((\mathbb{P} \circ X^{-1}) \circ \pi_1^{-1}\big)(B) = (X\#\mathbb{P})(B \times \mathbb{R}^{d_2}) \\ &= (\pi_1(X\#\mathbb{P}))(B).\end{aligned}$$

The second claim now immediately follows from lemma 4.28. $\square$

To conclude the discussion of marginals, let us introduce a shorthand notation which we heavily make use of in chapters 1–3:

Remark 4.30. On phase space $\mathbb{R}^6 \cong \mathbb{R}^3 \times \mathbb{R}^3$, for μ a finite measure on $\mathbb{R}^6$ and $f : \mathbb{R}^3 \to \mathbb{R}^3$ measurable, we will often write $f *_q \mu := f * \pi_1(\mu)$, i.e. if $f * \pi_1(\mu) : \mathbb{R}^3 \to \mathbb{R}^3$ exists, then

$$(f *_q \mu)(q) = \int_{\mathbb{R}^3} f(q - \tilde{q}) \,\mathrm{d}(\pi_1(\mu))(\tilde{q}) = \int_{\mathbb{R}^6} \tilde{f}(q - \tilde{q}, \tilde{p}) \,\mathrm{d}\mu(\tilde{q}, \tilde{p}),$$

where we used the notation and statement of lemma 4.27. In particular, if μ has Lebesgue density $u \in L^1(\mathbb{R}^6)$, by lemma 4.28,

$$(f *_q \mu)(q) = \int_{\mathbb{R}^3} f(q - \tilde{q}) \cdot \rho(\tilde{q})\, \mathrm{d}\tilde{q} = (f * \rho)(q).$$

Our next measure theoretic topic is a short introduction to measure preserving maps, which we will depend on when discussing our generalization of Liouville's theorem in section 4.9.

Definition 4.31. *Let $(\Omega, \mathcal{A}, \mu)$ a measure space. A measurable map $\Phi : \Omega \to \Omega$ is called **measure preserving** (w.r.t. μ) if μ agrees with its pushforward under Φ, i.e. if*

$$(\Phi\#\mu)(B) = \mu(B) \qquad \forall B \in \mathcal{A}.$$

Remark 4.32. In case Φ is an invertible map with Φ^{-1} measurable, too, then Φ is measure preserving if and only if Φ^{-1} is measure preserving: Assume that Φ is measure preserving, then for every $B \in \mathcal{A}$,

$$(\Phi^{-1}\#\mu)(B) = \mu\big((\Phi^{-1})^{-1}(B)\big) = \mu(\Phi(B)) = (\Phi\#\mu)(\Phi(B)) = (\mu \circ \Phi^{-1})(\Phi(B)) = \mu(B).$$

By interchanging the roles of Φ and Φ^{-1}, the other direction follows, too.

Lemma 4.33. *A measurable map Φ on a σ-finite measure space $(\Omega, \mathcal{A}, \mu)$ is measure preserving if and only if for every $f \in \mathcal{L}^1(\Omega; \mathrm{d}\mu)$, $f \circ \Phi \in \mathcal{L}^1(\Omega; \mathrm{d}\mu)$ with*

$$\int_\Omega (f \circ \Phi)\, d\mu = \int_\Omega f\, d\mu. \tag{4.9}$$

Proof. The crucial observation is the identity $\mathbb{1}_C \circ \Phi = \mathbb{1}_{\Phi^{-1}(C)}$ for all $C \subset \Omega$, which is readily checked.

„$\Leftarrow$“: For $B \in \mathcal{A}$ with $\mu(B) < \infty$, clearly $\mathbb{1}_B \in \mathcal{L}^1(\Omega; \mathrm{d}\mu)$, and by (4.9) and our observation,

$$(\Phi\#\mu)(B) = \mu(\Phi^{-1}(B)) = \int_\Omega \mathbb{1}_{\Phi^{-1}(B)}\, \mathrm{d}\mu = \int_\Omega (\mathbb{1}_B \circ \Phi)\, \mathrm{d}\mu = \int_\Omega \mathbb{1}_B\, \mathrm{d}\mu = \mu(B). \tag{$*$}$$

In case $\mu(B) = \infty$, let $(S_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ with $\mu(S_n) < \infty$ for all $n \in \mathbb{N}$ and $S_n \nearrow \Omega$ (existence of such a sequence is precisely provided by the assumption of σ-finiteness), then also $B \cap S_n \nearrow B$, and from continuity of a measure from below, we conclude

$$(\Phi\#\mu)(B) = \lim_{n\to\infty} (\Phi\#\mu)(B \cap S_n) = \lim_{n\to\infty} \mu(B \cap S_n) = \mu(B).$$

„$\Rightarrow$“: Note that for Φ measure preserving w.r.t. μ, by the computation $(*)$, (4.9) holds for indicator functions. Then apply algebraic induction. $\square$

The final part of this section introduces a definition which aggregates some important properties of the stochastic processes Δ, J, I introduced in chapter 3.2.

Definition 4.34. *Let $(\Omega, \mathcal{A})$ a measurable space and X, Y topological spaces. We equip X, Y with their Borel-σ-algebras, which makes them measurable spaces as well. A map $c : \Omega \times X \to Y$ is called **Carathéodory function** if*

(i) $c_\omega : X \to Y, \quad x \mapsto c(\omega, x)$ *is continuous for all* $\omega \in \Omega$, *and*

(ii) $c_x : \Omega \to Y, \quad \omega \mapsto c(\omega, x)$ *is measurable for all* $x \in X$.

Lemma 4.35. *Under the hypotheses of definition 4.34, assume additionally that* X *is separable and* $Y = \mathbb{R}$. *Then the map*

$$f : \Omega \to \overline{R}, \qquad \omega \mapsto \sup_{x \in X} c(\omega, x)$$

is measurable. In particular, this holds if $X \subset \mathbb{R}$ *is an interval (and* X *is equipped with the subspace topology induced by the standard topology on* $\mathbb{R}$*, of course).*

Proof. Let $C \subset X$ a countable, dense subset. We first claim that

$$s_\omega := \sup_{x \in X} c(\omega, x) = \sup_{x \in C} c(\omega, x) \qquad \forall \omega \in \Omega.$$

Indeed, fix $\omega \in \Omega$. Since $X \supset C$, "$\geq$" is clear. On the other hand, if $s_\omega < \infty$, let $\varepsilon > 0$ be arbitrary, then by definition of the supremum there is some $y \in X$ such that $c(\omega, y) > s_\omega - \varepsilon$. By continuity of the map $c(\omega, \cdot)$, $((c(\omega, \cdot))^{-1}((s_\omega - \varepsilon, s_\omega + \varepsilon)) \subset X$ is open, and it is non-empty because it contains y. Since $C \subset X$ is dense, there is some $z \in ((c(\omega, \cdot))^{-1}(s_\omega - \varepsilon, s_\omega + \varepsilon)) \cap C$. Consequently, $\sup_{x \in C} c(\omega, x) \geq c(\omega, z) > s_\omega - \varepsilon$. But $\varepsilon > 0$ was arbitrary, so we also obtain $s_\omega \leq \sup_{x \in C} c(\omega, x)$. For $s_\omega = \infty$, one can argue in a similar fashion.

Since $C \subset X$ is countable, we can write $C = \bigcup_{n \in \mathbb{N}} \{x_n\}$ for an appropriate sequence $(x_n)_{n \in \mathbb{N}} \subset X$ and for $n \in \mathbb{N}$ define the map $c_n : \Omega \to \mathbb{R}$, $\omega \mapsto c(\omega, x_n)$. Then by hypothesis, c_n is measurable for all $n \in \mathbb{N}$, and

$$f(\omega) = \sup_{x \in X} c(\omega, x) = \sup_{x \in C} c(\omega, x) = \sup_{n \in \mathbb{N}} c(\omega, x_n) = \sup_{n \in \mathbb{N}} c_n(\omega).$$

By [15, p. 17], f is measurable as the pointwise supremum of measurable functions $X \to \overline{\mathbb{R}}$. $\square$

4.4 Bounded Lipschitz topics

Let $d, d' \in \mathbb{N}$ and denote by $\mathcal{M}(\mathbb{R}^d)$ be the space of finite, signed measures on $\mathbb{R}^d$. This means that any $\sigma \in \mathcal{M}(\mathbb{R}^d)$ can be written via $\sigma = \mu - \nu$ where $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$ are finite (classical/non-negative) measures. We are going to need the following definitions:

Definition 4.36. *We call a map* $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ ***Lipschitz continuous*** *if*

$$\|f\|_{\mathrm{L}} := \sup_{\substack{x, y \in \mathbb{R}^d \\ x \neq y}} \frac{|f(x) - f(y)|}{|x - y|} < \infty. \qquad (*)$$

The class of ***bounded Lipschitz functions*** *is then defined by*

$$\mathrm{BL}(\mathbb{R}^d; \mathbb{R}^{d'}) := \left\{ f : \mathbb{R}^d \to \mathbb{R}^{d'} : \|f\|_\infty < \infty, \ \|f\|_{\mathrm{L}} < \infty \right\}.$$

As usual, we write $\mathrm{BL}(\mathbb{R}^d)$ *in case* $d' = 1$.

Remark 4.37.

(a) Since all norms on $\mathbb{R}^d$ and $\mathbb{R}^{d'}$ are equivalent, the definition of $\mathrm{BL}(\mathbb{R}^d;\mathbb{R}^{d'})$ does not depend on the choice of norms on $\mathbb{R}^d$ resp. $\mathbb{R}^{d'}$. Unless mentioned otherwise, we will use the maximum norm both on $\mathbb{R}^d$ and $\mathbb{R}^{d'}$ in the sequel because it blends particularly well with treating product spaces.

(b) In the literature, there are different conventions regarding the definition of Lipschitz continuity; what we call Lipschitz continuity is sometimes also considered as *global Lipschitz continuity*, by contrast to *local Lipschitz continuity* where $(*)$ is computed for fixed x and the bound on the r.h.s. of $(*)$ may depend on $x \in \mathbb{R}^d$.

(c) Note that by construction, $|f(x) - f(y)| \leq \|f\|_{\mathrm{L}} \cdot |x-y|$ for all $x, y \in \mathbb{R}^d$. From this, we see that every Lipschitz continuous map f is uniformly continuous ("$\delta = \varepsilon \cdot \|f\|_{\mathrm{L}}^{-1}$") and in particular continuous. Consequently, $\mathrm{BL}(\mathbb{R}^d;\mathbb{R}^{d'}) \subset \mathcal{C}_b(\mathbb{R}^d;\mathbb{R}^{d'})$.

(d) For $d'' \in \mathbb{N}$, $f \in \mathrm{BL}(\mathbb{R}^d;\mathbb{R}^{d'})$ and $g \in \mathrm{BL}(\mathbb{R}^{d'};\mathbb{R}^{d''})$, one has $g \circ f \in \mathrm{BL}(\mathbb{R}^d;\mathbb{R}^{d''})$, with

$$\|g \circ f\|_{\mathrm{BL}} \leq \max\{\|g\|_\infty, \|f\|_{\mathrm{L}} \cdot \|g\|_{\mathrm{L}}\}.$$

Indeed, it is clear that $\|g \circ f\|_\infty \leq \|g\|_\infty$, and for $x, y \in \mathbb{R}^d$,

$$|(g \circ f)(x) - (g \circ f)(y)| = |g(f(x)) - g(f(y))| \leq \|g\|_{\mathrm{L}} \cdot |f(x) - f(y)| \leq \|g\|_{\mathrm{L}} \cdot \|f\|_{\mathrm{L}} \cdot |x-y|,$$

which proves that $\|g \circ f\|_{\mathrm{L}} \leq \|f\|_{\mathrm{L}} \cdot \|g\|_{\mathrm{L}}$.

(e) Be aware that unlike one might expect from the notation, actually $\|\cdot\|_{\mathrm{L}}$ is only a *seminorm* on $\{f \in \mathrm{Abb}(\mathbb{R}^d;\mathbb{R}^{d'}) : \|f\|_{\mathrm{L}} < \infty\}$, since for every $c \in \mathbb{R}^{d'}$, the map $g_c(x) := c$ obviously satisfies $\|g_c\|_{\mathrm{L}} = 0$.

(f) However, one can easily convince oneself that $\mathrm{BL}(\mathbb{R}^d;\mathbb{R}^{d'})$ becomes a complete normed space w.r.t. the norm $\|\cdot\|_{\mathrm{BL}} := \max\{\|\cdot\|_\infty, \|\cdot\|_{\mathrm{L}}\}$ and therefore with all the equivalent norms $\|\cdot\|_{\mathrm{BL}'} := |(\|\cdot\|_\infty, \|\cdot\|_{\mathrm{L}})|_\sim$ where $|\cdot|_\sim$ is any norm on $\mathbb{R}^2$: that $\|\cdot\|_{\mathrm{BL}}$ satisfies all requirements for a norm is obvious. Since $\|\cdot\|_{\mathrm{BL}}$ is obviously *stronger* than $\|\cdot\|_\infty$, it is well-known that any Cauchy sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{BL}(\mathbb{R}^d;\mathbb{R}^{d'}) \subset \mathcal{C}_b(\mathbb{R}^d;\mathbb{R}^{d'})$ w.r.t. $\|\cdot\|_{\mathrm{BL}}$ converges uniformly to some $f \in \mathcal{C}_b(\mathbb{R}^d;\mathbb{R}^{d'})$. The only thing that remains to prove is that f is Lipschitz-continuous with $\|f - f_n\|_{\mathrm{L}} \xrightarrow{n\to\infty} 0$. Indeed, for every $\varepsilon > 0$, we can choose $N \in \mathbb{N}$ such that $\|f_m - f_n\|_{\mathrm{L}} \leq \varepsilon$ for all $m, n \geq N$. By uniform convergence and continuity of $|\cdot|$, we obtain that for all $x, y \in \mathbb{R}^d$ with $x \neq y$ and all $n \geq N$,

$$\frac{\big|(f-f_n)(x) - (f-f_n)(y)\big|}{|x-y|} = \lim_{m\to\infty} \frac{\big|(f_m-f_n)(x) - (f_m-f_n)(y)\big|}{|x-y|} \leq \lim_{m\to\infty} \|f_m - f_n\|_{\mathrm{L}} \leq \varepsilon.$$

This proves that

$$\|f - f_n\|_{\mathrm{L}} = \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \frac{\big|(f-f_n)(x) - (f-f_n)(y)\big|}{|x-y|} \leq \varepsilon.$$

Hence, we have indeed shown that $\|f - f_n\|_{\mathrm{L}} \xrightarrow{n\to\infty} 0$. Moreover, for, say, $\varepsilon = 1$ and $n \in \mathbb{N}$ large enough such that $\|f - f_n\|_{\mathrm{L}} < 1$, we deduce from the triangle inequality that $\|f\|_{\mathrm{L}} \leq \|f - f_n\|_{\mathrm{L}} + \|f_n\|_{\mathrm{L}} \leq 1 + \|f_n\|_{\mathrm{L}}$, i.e. $\|f\|_{\mathrm{L}} < \infty$, as claimed. □

Definition 4.38. *For $\sigma \in \mathcal{M}(\mathbb{R}^d)$, we define the* ***bounded Lipschitz norm***

$$\|\sigma\|_{\mathrm{BL}} := \sup_{\substack{f \in \mathrm{BL}(\mathbb{R}^d) \\ \|f\|_{\mathrm{BL}}=1}} \int_{\mathbb{R}^d} f \,\mathrm{d}\sigma. \tag{4.10}$$

Remark 4.39.

(a) Note that $\|\cdot\|_{\mathrm{BL}}$ is well-defined: denoting by $|\sigma|$ the total variation measure of σ (i.e. for $\sigma = \mu - \nu$ where $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$, $|\sigma| = \mu + \nu$) and using that $|\sigma|$ is a finite, non-negative measure, for every $f \in \mathrm{BL}(\mathbb{R}^d)$ with $\|f\|_{\mathrm{BL}} = 1$ we have

$$\int_{\mathbb{R}^d} |f| \,\mathrm{d}|\sigma| \le \int_{\mathbb{R}^d} \|f\|_\infty \,\mathrm{d}|\sigma| = |\sigma|(\mathbb{R}^d),$$

which shows that $|\|\sigma\|_{\mathrm{BL}}| \le |\sigma|(\mathbb{R}^d) < \infty$. Since it is readily checked that $f \in \mathrm{BL}(\mathbb{R}^d)$ if and only if $-f \in \mathrm{BL}(\mathbb{R}^d)$, with $\|f\|_{\mathrm{BL}} = \|-f\|_{\mathrm{BL}}$, it follows that

$$\|\sigma\|_{\mathrm{BL}} = \sup_{\substack{f \in \mathrm{BL}(\mathbb{R}^d) \\ \|f\|_{\mathrm{BL}}=1}} \left| \int_{\mathbb{R}^d} f \,\mathrm{d}\sigma \right|, \tag{4.11}$$

in particular, $\|\cdot\|_{\mathrm{BL}} \ge 0$. Moreover, it is clear that we could equivalently take the supremum over all $f \in \mathrm{BL}(\mathbb{R}^d)$ with $\|f\|_{\mathrm{BL}} \le 1$ in (4.10).

(b) As the name suggests, $\|\cdot\|_{\mathrm{BL}}$ actually defines a norm on $\mathcal{M}(\mathbb{R}^d)$: We have already shown that $\|\cdot\|_{\mathrm{BL}} : \mathcal{M}(\mathbb{R}^d) \to [0, \infty)$. Absolute homogenicity and the triangle inequality are obvious using (4.11). The only non-trivial task is to prove that $\|\sigma\|_{\mathrm{BL}} = 0$ implies $\sigma = 0$. Let us briefly sketch this: For $C \subset \mathbb{R}^d$ a closed set and $n \in \mathbb{N}$, define

$$g_n : \mathbb{R}^d \to \mathbb{R}, \qquad x \mapsto \tfrac{1}{n} \cdot \max\{0, 1 - n \cdot \operatorname{dist}(x, C)\}.$$

One can then check that $\|g_n\|_\infty = \frac{1}{n}$ and $\|g_n\|_{\mathrm{L}} = 1$ for all $n \in \mathbb{N}$; for the latter equality, use that $|x - y| \ge |\operatorname{dist}(x, C) - \operatorname{dist}(y, C)|$, which is itself a consequence of the triangle inequality for the metric d. This shows that $f_n \in \mathrm{BL}(\mathbb{R}^d)$ with $\|f_n\|_{\mathrm{BL}} = 1$ for all $n \in \mathbb{N}$. Moreover, since $C \subset \mathbb{R}^d$ is closed, $n \cdot g_n \xrightarrow{n\to\infty} \mathbb{1}_C$ pointwise on $\mathbb{R}^d$. Now, let us assume that $\|\sigma\|_{\mathrm{BL}} = 0$, then by (4.11),

$$\int_{\mathbb{R}^d} n \cdot g_n \,\mathrm{d}\sigma = n \int_{\mathbb{R}^d} g_n \,\mathrm{d}\sigma = 0 \qquad \forall n \in \mathbb{N}.$$

Since $\mathbb{1}_{\mathbb{R}^d}$ is an integrable majorant for $(n \cdot g_n)_{n\in\mathbb{N}}$ (recall that $|\sigma|$ is finite), dominated convergence yields

$$0 = \lim_{n\to\infty} \int_{\mathbb{R}^d} n \cdot g_n \,\mathrm{d}\sigma = \int_{\mathbb{R}^d} \lim_{n\to\infty} (n \cdot g_n) \,\mathrm{d}\sigma = \int_{\mathbb{R}^d} \mathbb{1}_C \,\mathrm{d}\sigma = \sigma(C).$$

But the closed sets are a $\cap$-stable generator of $\mathcal{B}(\mathbb{R}^d)$, and the zero measure and σ coincide on closed sets, so we may conclude (see [12, p. 39]) that $\sigma = 0$ (actually, we may only use this directly for $\sigma \in \mathcal{M}^+(\mathbb{R}^d)$, however, using the *Hahn decomposition* (which precisely corresponds to writing $\sigma = \mu - \nu$ for $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$), we can easily generalize this statement to arbitrary signed measures).

(c) Note that for $f \in \mathrm{BL}(\mathbb{R}^d)$, $g := \pm\|f\|_{\mathrm{BL}}^{-1} \cdot f$ satisfies $g \in \mathrm{BL}(\mathbb{R}^d)$ with $\|g\|_{\mathrm{BL}} = 1$, and hence for every $\sigma \in \mathcal{M}(\mathbb{R}^d)$,

$$\left|\int_{\mathbb{R}^d} g \,\mathrm{d}\sigma\right| \leq \sup_{\substack{\tilde{f} \in \mathrm{BL}(\mathbb{R}^d) \\ \|\tilde{f}\|_{\mathrm{BL}}=1}} \int_{\mathbb{R}^d} \tilde{f} \,\mathrm{d}\sigma = \|\sigma\|_{\mathrm{BL}}.$$

Consequently, by absolute homogenicity,

$$\left|\int_{\mathbb{R}^d} f \,\mathrm{d}\sigma\right| = \|f\|_{\mathrm{BL}} \cdot \left|\int_{\mathbb{R}^d} g \,\mathrm{d}\sigma\right| \leq \|f\|_{\mathrm{BL}} \cdot \|\sigma\|_{\mathrm{BL}} \qquad \forall f \in \mathrm{BL}(\mathbb{R}^d). \tag{4.12}$$

(d) We can generalize $\|\cdot\|_{\mathrm{BL}}$ in a straightforward way to product spaces using the product metric: for $n \in \mathbb{N}$ and $\sigma \in (\mathcal{M}(\mathbb{R}^d))^n$,

$$\|\sigma\|_{\mathrm{BL}} := \max\left\{\|\sigma_i\|_{\mathrm{BL}} : i \in [n]\right\}.$$

(e) Let $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space and $(X_n)_{n\in\mathbb{N}}, X : \Omega \to \mathbb{R}^d$ be random variables distributed by the laws $(\mu_n)_{n\in\mathbb{N}}$ resp. μ, i.e. $X_n\#\mathbb{P} = \mu_n$, $X\#\mathbb{P} = \mu$. Let $f \in \mathrm{BL}(\mathbb{R}^d)$, then by the observation we just made, for $\|\mu_n - \mu\|_{\mathrm{BL}} \to 0$, we obtain

$$\mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right] + \mathbb{E}\left[f(X_n) - f(X)\right] = \mathbb{E}\left[f(X)\right] + \int_{\mathbb{R}^d} f \,\mathrm{d}(\mu_n - \mu) \xrightarrow{n\to\infty} \mathbb{E}\left[f(X)\right]$$

since

$$\left|\int_{\mathbb{R}^d} f \,\mathrm{d}(\mu_n - \mu)\right| \leq \|f\|_{\mathrm{BL}} \cdot \|\mu_n - \mu\|_{\mathrm{BL}} \xrightarrow{n\to\infty} 0.$$

Consequently, $\|\cdot\|_{\mathrm{BL}}$ induces a metric on (signed) finite measures which is highly relevant from the physical point of view because convergence of (probability) measures in $\|\cdot\|_{\mathrm{BL}}$ characterizes convergence of expectation values of classical observables, e.g. on phase space. From the mathematical point of view, the Portemanteau theorem (see e.g. [26, p. 254]) states that for $\|\mu_n - \mu\|_{\mathrm{BL}} \xrightarrow{n\to\infty} 0$, $\mu_n \xrightarrow{w} \mu$, i.e μ_n converges to μ weakly in the sense of probability measures.

(f) Unfortunately, $(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathrm{BL}})$ is *not* complete: It is well-known (see e.g. [12, p. 119]) that $(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathrm{tot}})$ is complete, where $\|\cdot\|_{\mathrm{tot}}$ denotes the *total variation norm*, i.e. for $\sigma = \mu - \nu$ with $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$,

$$\|\sigma\|_{\mathrm{tot}} := |\sigma|(\mathbb{R}^d) = \mu(\mathbb{R}^d) + \nu(\mathbb{R}^d).$$

Moreover, $\|\cdot\|_{\mathrm{BL}}$ and $\|\cdot\|_{\mathrm{tot}}$ are not equivalent: We have already seen that $\|\cdot\|_{\mathrm{BL}} \leq \|\cdot\|_{\mathrm{tot}}$ (this is hidden in remark 4.39 (a)). On the other hand, consider the sequence of Dirac measures $(\delta_{\frac{1}{n}})_{n\in\mathbb{N}} \subset \mathcal{M}(\mathbb{R}^d)$, then clearly

$$\left\|\delta_{\frac{1}{n}} - \delta_0\right\|_{\mathrm{tot}} = \delta_{\frac{1}{n}}(\mathbb{R}^d) + \delta_0(\mathbb{R}^d) = 2 \qquad \forall n \in \mathbb{N}.$$

Moreover, for $f \in \mathrm{BL}(\mathbb{R}^d)$ with $\|f\|_{\mathrm{BL}} = 1$, $\|f\|_{\mathrm{L}} \leq 1$, and hence we obtain that for all $n \in \mathbb{N}$, $|f(\frac{1}{n}) - f(0)| \leq |\frac{1}{n} - 0| = \frac{1}{n}$. Consequently, $\|\delta_{\frac{1}{n}} - \delta_0\|_{\mathrm{BL}} \leq \frac{1}{n} \xrightarrow{n\to\infty} 0$. This already implies that $\left(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathrm{BL}}\right)$ is not complete: Assume for contradiction it was, then since $\mathrm{id} : \left(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathrm{tot}}\right) \to \left(\mathcal{M}(\mathbb{R}^d), \|\cdot\|_{\mathrm{BL}}\right)$ is bounded and surjective (in fact, bijective), it would be an easy consequence of the *open mapping theorem*, also known as *Banach-Schauder theorem* (see e.g. [11, p. 83]) that id is invertible with bounded inverse, which leads to the desired contradiction because this would mean that $\|\cdot\|_{\mathrm{tot}}$, $\|\cdot\|_{\mathrm{BL}}$ were equivalent.

Definition 4.40. *On the subspace of probability measures $\mathcal{P}(\mathbb{R}^d) \subset \mathcal{M}(\mathbb{R}^d)$ on $\mathbb{R}^d$, $\|\cdot\|_{\mathrm{BL}}$ induces a metric d_{BL} via*

$$d_{\mathrm{BL}} : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to [0, \infty), \qquad (\mu, \nu) \mapsto d_{\mathrm{BL}}(\mu, \nu) := \|\mu - \nu\|_{\mathrm{BL}}.$$

*We call d_{BL} the **bounded Lipschitz distance** or **bounded Lipschitz metric**.*

In the light of items (e) and (f) in the preceding remarks, the following result, which is highly important for our purposes, comes as quite a big surprise:

Theorem 4.41. *The topological space $\mathcal{P}(\mathbb{R}^d)$ of probability measures with the topology stemming from weak convergence is metrizable by $\|\cdot\|_{\mathrm{BL}}$, i.e. for $\mu \in \mathcal{P}(\mathbb{R}^d)$, $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$,*

$$\mu_n \xrightarrow{w} \mu \qquad \Leftrightarrow \qquad d_{\mathrm{BL}}(\mu_n, \mu) \xrightarrow{n\to\infty} 0.$$

Moreover, $(\mathcal{P}(\mathbb{R}^d), d_{\mathrm{BL}})$ is complete.

Proof. Observe first that $\mathbb{R}^d$ is a complete, separable, metric space (w.r.t. the metric induced by $|\cdot|$, which we always use). Then, for the first statement, see [9, p. 193, thm. 8.3.2], noting that by [9, p. 13, corollary 6.3.5], the Borel-σ-algebra and the Baire-σ-algebra on $\mathbb{R}^d$ coincide and hence $\mathcal{M}^+_\sigma(\mathbb{R}^d) = \mathcal{M}^+(\mathbb{R}^d)$ in the notation of the book. For the second part, see [9, p. 232–233, thm. 8.10.43], using that by [9, p. 70, thm. 7.1.7], every Borel measure on $\mathbb{R}^d$ is Radon, i.e. $\mathcal{P}_r(\mathbb{R}^d) = \mathcal{P}(\mathbb{R}^d)$. $\square$

Remark 4.42. The only reason why the counterexample from (f) in remark 4.39 does not work when we replace $\mathcal{M}(\mathbb{R}^d)$ by $\mathcal{P}(\mathbb{R}^d)$ is that $\mathcal{P}(\mathbb{R}^d)$ is not a vector space anymore, and hence the open mapping theorem does not apply. It might be enlightening to find a Cauchy sequence w.r.t. $\|\cdot\|_{\mathrm{BL}}$ in $\mathcal{M}(\mathbb{R}^d)$ which does not converge; unfortunately, the author of this thesis was not able to find any. Another good reference for theorem 4.41 is [39, p. 73].

We have just seen that we can *test* weak convergence of (probability) measures by bounded Lipschitz functions. This suggests that one might use $\mathrm{BL}(\mathbb{R}^d)$ also as *test space* for other *weak properties.*

Definition 4.43. *For $I \subset \mathbb{R}$, consider a curve $\mu : I \to \mathcal{M}(\mathbb{R}^d)$, $t \mapsto \mu(t) =: \mu_t$. We say that μ is **weakly continuous** if for all $f \in \mathrm{BL}(\mathbb{R}^d)$, the map*

$$I \to \mathbb{R}, \qquad t \mapsto \int_{\mathbb{R}^d} f \,\mathrm{d}\mu_t$$

is continuous.

One would wish that the bounded Lipschitz norm of a weakly continuous curve is a continuous map, however, it is not obvious whether this is true. At least, one can show the following:

Lemma 4.44. *Let $\mu : I \to \mathcal{M}(\mathbb{R}^d)$ be weakly continuous. Then the map $I \to \mathbb{R}$, $t \mapsto \|\mu_t\|_{\mathrm{BL}}$ is lower semi-continuous. In particular, for $\mu, \nu \in \mathcal{C}^*([0,T]; (\mathcal{P}(\mathbb{R}^6))^n)$, $t \mapsto d_{\mathrm{BL}}(\mu_t, \nu_t)$ is measurable and bounded.*

Proof. Let $\mu : I \to \mathcal{M}(\mathbb{R}^d)$ a weakly continuous curve, $t \in I$ and $\varepsilon > 0$. By definition of $\|\cdot\|_{\mathrm{BL}}$, there is some $g \in \mathrm{BL}(\mathbb{R}^d)$ with $\|g\|_{\mathrm{BL}} = 1$ such that

$$\|\mu_t\|_{\mathrm{BL}} \leq \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_t + \frac{\varepsilon}{2}.$$

By weak continuity of μ, there is some $\delta > 0$ such that for all $s \in I$ with $|s - t| < \delta$,

$$\left| \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_t - \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_s \right| \leq \frac{\varepsilon}{2}.$$

Consequently, for these s,

$$\|\mu_s\|_{\mathrm{BL}} \geq \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_s \geq \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_t - \frac{\varepsilon}{2} \geq \|\mu_t\|_{\mathrm{BL}} - \varepsilon.$$

This already shows that $t \mapsto \|\mu_t\|_{\mathrm{BL}}$ is *lower semi-continuous.* However, every lower semi-continuous map is measurable (one can readily show that the preimage of sets of the form (a, ∞) is open for every $a \in \mathbb{R}$). Moreover, the computation

$$d_{\mathrm{BL}}(\mu_t, \nu_t) = \|\mu_t - \nu_t\|_{\mathrm{BL}} \leq \|\mu_t - \nu_t\|_{\mathrm{tot}} \leq \|\mu_t\|_{\mathrm{tot}} + \|\nu_t\|_{\mathrm{tot}} = 2 \qquad \forall t \in [0, T]$$

shows that $[0, T] \to \mathbb{R},\ t \mapsto d_{\mathrm{BL}}(\mu_t, \nu_t)$ is also bounded. The claim now follows since the maximum of several measurable, bounded maps is again measurable and bounded. $\square$

The following lemma will be crucial for our existence and uniqueness proof in section 2.2:

Lemma 4.45. *Let d_{BL} be the bounded Lipschitz metric on $\mathcal{M}(\mathbb{R}^d)$ and $A \subset \mathcal{M}(\mathbb{R}^d)$ a closed subset w.r.t. this metric, i.e. $(A, d_{\mathrm{BL}}|_A)$ is itself a complete metric space. Moreover, let $f : I \to \mathbb{R}^+$ a continuous map. Then*

$$\mathcal{C}^*(I; A) := \{\mu : I \to A : \mu \text{ weakly continuous}\}$$

is complete w.r.t. the metric

$$\overline{d}_{\mathrm{BL}}(\mu, \nu) := \sup\big\{f(t) \cdot d_{\mathrm{BL}}(\mu(t), \nu(t)) : t \in I\big\}.$$

Proof. Let $(\mu_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{C}^*(I; A)$, i.e. $\overline{d}_{\mathrm{BL}}(\mu_m, \mu_n) \xrightarrow{m,n\to\infty} 0$. Then for every fixed $t \in I$, $f(t) \cdot d_{\mathrm{BL}}(\mu_m(t), \mu_n(t)) \xrightarrow{m,n\to\infty} 0$. Since $f(t) \neq 0$, we conclude that $(\mu_n(t))_{n\in\mathbb{N}}$ is a Cauchy sequence in A. By closedness of A w.r.t. d_{BL}, there is some $\mu(t) \in A$ such that $\mu_n(t) \xrightarrow{n\to\infty} \mu(t)$. Consequently, we define the expected limit curve $\mu : I \to \mathbb{R},\ t \mapsto \mu(t)$. All we need to prove is that μ is weakly continuous and $\overline{d}_{\mathrm{BL}}(\mu, \mu_n) \xrightarrow{n\to\infty} 0$.

Let $\varepsilon > 0$ be arbitrary and choose $N \in \mathbb{N}$ such that for all $m, n \geq N$, $\overline{d}_{\mathrm{BL}}(\mu_m, \mu_n) \leq \varepsilon$. It follows by continuity of the metric d_{BL} that for all $t \in I$,

$$f(t) \cdot d_{\mathrm{BL}}(\mu(t), \mu_n(t)) = f(t) \cdot \lim_{m\to\infty} d_{\mathrm{BL}}(\mu_m(t), \mu_n(t)) = \lim_{m\to\infty} f(t) \cdot d_{\mathrm{BL}}(\mu_m(t), \mu_n(t)) \leq \varepsilon,$$

which shows that $\overline{d}_{\mathrm{BL}}(\mu, \mu_n) \leq \varepsilon$. This already proves that $\mu_n \xrightarrow{n\to\infty} \mu$ w.r.t. $\overline{d}_{\mathrm{BL}}$.

Now, let $g \in \mathrm{BL}(\mathbb{R}^d)$ and $t \in I$. By an easy distinction of cases, one can show that there is some $\delta_1 > 0$ such that

$$[t - \delta_1, t + \delta_1] \cap I \in \big\{\{t\}, [t - \delta_1, t], [t, t + \delta_1], [t - \delta_1, t + \delta_1]\big\}.$$

In any case, $m := \inf\{f(s) : s \in [t-\delta_1, t+\delta_1] \cap I\} > 0$ since the continuous function f attains its minimum on any non-empty, compact interval contained in I. Let us choose $N \in \mathbb{N}$ such that for all $n \geq N$, $\overline{d}_{\mathrm{BL}}(\mu, \mu_n) \leq \frac{m\varepsilon}{3}$. In particular, for $s \in [t-\delta_1, t+\delta_1] \cap I$, $f(s) \cdot d_{\mathrm{BL}}(\mu(s), \mu_n(s)) \leq \frac{m\varepsilon}{3}$. It follows that

$$d_{\mathrm{BL}}(\mu(s), \mu_n(s)) \leq \frac{1}{f(s)} \cdot \overline{d}_{\mathrm{BL}}(\mu, \mu_n) \leq \frac{1}{f(s)} \cdot \frac{m\varepsilon}{3} \leq \frac{m\varepsilon}{3m} = \frac{\varepsilon}{3}.$$

By weak continuity of μ_n, we can find some $\delta_2 > 0$ such that for all $s \in I$ with $|s-t| \leq \delta_2$,

$$\left| \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_n(s) - \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_n(t) \right| \leq \frac{\varepsilon}{3}.$$

Let $\delta := \min\{\delta_1, \delta_2\} > 0$. Then for all $s \in I$ with $|s-t| < \delta$,

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d} g \,\mathrm{d}\mu(t) - \int_{\mathbb{R}^d} g \,\mathrm{d}\mu(s) \right| \\
&\leq \left| \int_{\mathbb{R}^d} g \,\mathrm{d}\mu(t) - \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_n(t) \right| + \left| \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_n(t) - \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_n(s) \right| + \left| \int_{\mathbb{R}^d} g \,\mathrm{d}\mu_n(s) - \int_{\mathbb{R}^d} g \,\mathrm{d}\mu(s) \right| \\
&\leq d_{\mathrm{BL}}(\mu(t), \mu_n(t)) + \frac{\varepsilon}{3} + d_{\mathrm{BL}}(\mu(s), \mu_n(s)) \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

This proves weak continuity of μ. □

4.5 The empirical probability measure

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $X, (X_n)_{n\in\mathbb{N}} : \Omega \to \mathbb{R}^d$ be i.i.d. random variables distributed by the law $\mu = \mathbb{P} \circ X^{-1}$. Let $F : \mathbb{R}^d \to [0,1]$ be the distribution function of X, i.e. for $x \in \mathbb{R}^d$,

$$F_X(x) := \mathbb{P}[X \leq x] := \mathbb{P}\big[\{\omega \in \Omega : X(\omega) \leq x\}\big] = \mu((-\infty, x]),$$

where for $a, b \in \mathbb{R}^d$ we write $a \leq b$ iff $a_j \leq b_j$ for all $j \in [d]$. Then for any $N \in \mathbb{N}$, a fixed $\omega \in \Omega$ determines a unique element $(X_1(\omega), \ldots, X_N(\omega)) \in (\mathbb{R}^d)^N$, which we call **sample** of length N. Our interpretation is that w.r.t. the distribution of the random variable X, N vectors in $\mathbb{R}^d$ are randomly chosen. For $\omega \in \Omega$, the **empirical distribution function of the sample**, $F^{\omega}_{\mathrm{emp},N}$, is defined via

$$F^{\omega}_{\mathrm{emp},N} : \mathbb{R}^d \to [0,1], \quad x \mapsto \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{(-\infty,x]}(X_i(\omega)).$$

Then $F^{\omega}_{\mathrm{emp},N}$ is the distribution function belonging to the random variable whose law is given by

$$\mu^{\omega}_{\mathrm{emp},N} : \mathbb{R}^d \to \mathbb{R}, \qquad \omega \mapsto \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i(\omega)},$$

which we call **empirical (probability) measure**. We observe that for $f : \mathbb{R}^d \to \mathbb{R}$ bounded and measurable and $\omega \in \Omega$,

$$\int_{\mathbb{R}^d} f \,\mathrm{d}\mu^{\omega}_{\mathrm{emp},N} = \frac{1}{N} \sum_{i=1}^{n} f(X_i(\omega)),$$

and since $f \circ X$, $(f \circ X_i)_{i \in \mathbb{N}} : \Omega \to \mathbb{R}$ are also i.i.d. bounded random variables, we obtain from the strong law of large numbers (see e.g. [19, 295]) that

$$\mathbb{P}\left[\left\{\omega \in \Omega : \int_{\mathbb{R}^6} f \, \mathrm{d}\mu^{\omega}_{\mathrm{emp},N} \xrightarrow{N\to\infty} \mathbb{E}\left[f \circ X\right] = \int_{\mathbb{R}^d} f \, \mathrm{d}\mu\right\}\right] = 1.$$

In particular, $\int_{\mathbb{R}^d} f \, \mathrm{d}\mu^{\omega}_{\mathrm{emp},N} \xrightarrow{N\to\infty} \int_{\mathbb{R}^d} f \, \mathrm{d}\mu$ almost surely for all $f \in \mathrm{BL}(\mathbb{R}^d)$. However, the subset of Ω where this convergence does not happen might in general depend on $f \in \mathrm{BL}(\mathbb{R}^d)$. Consequently, having in mind that a countable union of null sets is still a null set, the following, quite surprising statement can be interpreted in the sense that the set of bounded Lipschitz functions is *almost separable* (in fact, $\mathrm{BL}(C)$ where $C \subset \mathbb{R}^d$ is totally bounded is separable).

Theorem 4.46 (Varadarajan)**.** *The empirical measures $\mu_{\mathrm{emp},N}$ converge weakly to μ almost surely in the following sense:*

$$\mathbb{P}\left[\left\{\omega \in \Omega : \mu^{\omega}_{\mathrm{emp},N} \xrightarrow[N\to\infty]{w} \mu\right\} = 1\right].$$

Proof. See [13, p. 399]. □

We have already seen in theorem 4.41 that weak convergence on $\mathbb{R}^d$ is metrizable by the bounded Lipschitz distance, i.e. as $N \to \infty$,

$$\mu^{\omega}_{\mathrm{emp},N} \xrightarrow{w} \mu \quad \Leftrightarrow \quad d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},N}, \mu) \to 0.$$

Hence, theorem 4.46 states that the bounded Lipschitz distance between a probability measure and its empirical measure with sample length N converges to 0 almost surely as $N \to \infty$:

Corollary 4.47. *It holds that*

$$\mathbb{P}\left[\left\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},N}, \mu) \xrightarrow{N\to\infty} 0\right\}\right] = 1.$$

Recall from probability theory that almost sure convergence of random variables does imply convergence in probability. In the context of empirical measures, one might thus expect results which give bounds on the probability that the bounded Lipschitz distance between a probability measure and its empirical measure is bigger than a certain constant, i.e. which tell us the *rate of convergence of the empirical measures for typical initial conditions*. In fact, we are going to use the following statement, which, heuristically speaking, shows that the rate of convergence gets faster as the *tails* of the distribution get smaller, measured in terms of moments:

Theorem 4.48. *Let μ a probability measure on $\mathbb{R}^d$ and assume that for some $q > 2$, the q-th moment $M_q := \int_{\mathbb{R}^d} |x|^q \, \mathrm{d}\mu(x)$ is finite. Then there are some $c, C > 0$ such that for all $N \geq 1$ and $\varepsilon \in (0, 1)$,*

$$\mathbb{P}\big[\big\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},N}), \mu) \geq \varepsilon\big\}\big] \leq c \cdot \exp\left(-cN\varepsilon^{\max\{d,2\}}\right) + CN \cdot (N\varepsilon)^{\varepsilon - q}.$$

In particular, for every $\varepsilon \in (0, 1)$,

$$\mathbb{P}\left[\left\{\omega \in \Omega : d_{\mathrm{BL}}(\mu^{\omega}_{\mathrm{emp},N}, \mu) \geq \varepsilon\right\}\right] \xrightarrow{N\to\infty} 0.$$

Proof. If we replace d_{BL} by the first Wasserstein distance, then everything follows immediately from [17]. Hence, the only thing which needs clarification is that the bounded Lipschitz distance is bounded by the first Wasserstein distance W_1. However, this is a direct consequence of [9, p. 234], observing that in the notation of the book, $W_1(\mu,\nu) = W(\mu,\nu) = \|\mu-\nu\|_0^* \geq \|\mu-\nu\|_0 = d_{\mathrm{BL}}(\mu,\nu)$ since for $\|\cdot\|_0^*$, the supremum over all Lipschitz-continuous functions f with $\|f\|_{\mathrm{L}} \leq 1$ is taken, whereas in $\|\cdot\|_0$, one only allows f where both $\|f\|_\infty, \|f\|_{\mathrm{L}} \leq 1$. Often, this or similar results on the connection between the first Wasserstein distance, which is defined in terms of measures and marginals, and metrics on measures defined by comparing these in terms of integration against suitable test functions, are called *Kantorovic-Rubinshtein duality.* □

4.6 Convolution estimates

This section mainly serves for finding estimates on the supremum norm and the Lipschitz seminorm of convolutions of the Coulomb force and functions which satisfy S_δ^α-conditions (see definition 3.1) with bounded probability densities in $\mathbb{R}^3$.

Definition 4.49. *Let $\sigma \in \mathcal{M}(\mathbb{R}^d)$ be a signed measure on $\mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$ a measurable function such that for all $x \in \mathbb{R}^d$, $f_x := f(x-\cdot\,) \in \mathcal{L}^1(\mathbb{R}^d; \mathrm{d}\sigma)$. Then the map*

$$f * \sigma : \mathbb{R}^d \to \mathbb{R}, \qquad x \mapsto \int_{\mathbb{R}^d} f_x \,\mathrm{d}\sigma = \int_{\mathbb{R}^d} f(x-y)\,\mathrm{d}\sigma(y)$$

*is well-defined and called **convolution** of f and μ. For $u \in L^1(\mathbb{R}^d)$, let σ_u denote the signed measure which has Lebesgue-density u, i.e. for all $B \in \mathcal{B}(\mathbb{R}^d)$, $\sigma_u(B) := \int_B u(y)\,\mathrm{d}y$. Then we define the **convolution** of f and u via $f * u := f * \sigma_u$, i.e.*

$$(f*u)(x) = \int_{\mathbb{R}^d} f(x-y)\,\mathrm{d}\sigma_u(y) = \int_{\mathbb{R}^d} f(x-y)\cdot u(y)\,\mathrm{d}y.$$

Remark 4.50. Of course, for $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ with $|f_x| \in \mathcal{L}^1(\mathbb{R}^d; \mathrm{d}\sigma)$ for all $x \in \mathbb{R}^d$, we see that every component $(f_x)_i \in \mathcal{L}^1(\mathbb{R}^d; \mathrm{d}\sigma)$ for all $x \in \mathbb{R}^d$, and consequently we can define the convolution $f * \sigma$ component-wise, i.e. $(f*\sigma)_i := f_i * \sigma$ for all $i \in [d']$.

Lemma 4.51. *Let $h : \mathbb{R}^3 \to \mathbb{R}$ satisfy a S_δ^α-condition for $\alpha \in [2,3]$, and $\rho \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$. Define*

$$|\!|\!|\rho|\!|\!| := \max\{1, \|\rho\|_1 + \|\rho\|_\infty\}.$$

Then there exists some $C_\alpha > 0$ such that

$$\|h * \rho\|_\infty \leq C_\alpha\, |\!|\!|\rho|\!|\!| \cdot (1 + \delta_{\alpha,3}\ln(S)),$$

where δ denotes the Kronecker delta. Moreover, there is some $c > 0$ such that for any $\alpha \in [2,3]$ and $m \geq 2$,

$$\|h^m * \rho\|_\infty \leq 8\pi c^m\, |\!|\!|\rho|\!|\!| \cdot S^{(\alpha m-3)\delta}.$$

In particular, there are $\tilde{C}_\alpha, C > 0$ such that for $m \geq 2$ and $x \in \mathbb{R}^3$,

$$\big\|\big(h - (h*\rho)(x)\big)^m * \rho\big\|_\infty \leq C \cdot \tilde{C}_\alpha^m\, S^{(\alpha m-3)\delta} \cdot (1 + \delta_{\alpha,3}\ln(S))^m.$$

Proof. Recall from definition 3.1 that since h satisfies a S_δ^α-condition, it holds that there is some $c > 0$ such that

$$|h(q)| \leq c \cdot \min\left\{S^{\alpha\delta}, |q|^{-\alpha}\right\} \qquad \forall q \in \mathbb{R}^3.$$

In particular, $|h(q)| \leq cS^{\alpha\delta}$ for $|q| \leq S^{-\delta}$ and $|h(q)| \leq c\,|q|^{-\alpha}$ for $|q| \geq S^{-\delta}$. Let $y \in \mathbb{R}^3$ be arbitrary. For $\alpha \in [2,3)$ we compute

$$\begin{aligned}
|(h*\rho)(y)| &\leq \int_{B_1(y)} |h(y-q)| \cdot |\rho(q)| \, \mathrm{d}q + \int_{B_1^c(y)} |h(y-q)| \cdot |\rho(q)| \, \mathrm{d}q \\
&\leq \|\rho\|_\infty \cdot 4\pi c \int_0^1 r^{-\alpha+2} \, \mathrm{d}r + c\|\rho\|_1 \\
&\leq C_\alpha \, |||\rho||| \, ,
\end{aligned}$$

where $C_\alpha = c \cdot \max\left\{\frac{4\pi}{3-\alpha}, 1\right\} \geq c$. Note that as one might expect, $C_\alpha \nearrow \infty$ as $\alpha \nearrow 3$. On the other hand, for $\alpha = 3$, we have

$$\begin{aligned}
|(h*\rho)(y)| &\leq \int_{B_{S^{-\delta}}(y)} |h(y-q)| \cdot |\rho(q)| \, \mathrm{d}q + \int_{B_1(y)\setminus B_{S^{-\delta}}(y)} \dots \; + \int_{B_1^c(y)} \dots \\
&\leq \frac{4\pi}{3} \left(S^{-\delta}\right)^3 cS^{3\delta} \, \|\rho\|_\infty + \|\rho\|_\infty \cdot 4\pi c \int_{S^{-\delta}}^1 r^{-1} \, \mathrm{d}r + c\|\rho\|_1 \\
&\leq \frac{4\pi}{3} c \, \|\rho\|_\infty + 4\pi c \, \|\rho\|_\infty \cdot \delta \ln(S) + c\|\rho\|_1 \\
&\leq C_3 \, |||\rho||| \cdot (1 + \ln(S)) \, ,
\end{aligned}$$

where $C_3 = c \cdot \max\left\{\frac{4\pi}{3}, 4\pi\delta, 1\right\} \geq c$. Moreover, for $\alpha \in [2,3]$ and $m \geq 2$, we have that $\alpha m \geq 4$ and therefore

$$\begin{aligned}
|(h^m * \rho)(y)| &= \int_{B_{S^{-\delta}}(y)} |h(y-q)|^m \cdot |\rho(q)| \, \mathrm{d}q + \int_{B^c_{S^{-\delta}}(y)} \dots \\
&\leq 4\pi c^m \, \|\rho\|_\infty \cdot \left[\frac{1}{3}\left(S^{-\delta}\right)^3 \left(S^{\alpha\delta}\right)^m + \int_{S^{-\delta}}^\infty r^{-m\alpha+2} \, \mathrm{d}r\right] \\
&\leq 4\pi c^m \, \|\rho\|_\infty \cdot \left(\frac{1}{3} + \frac{1}{\alpha m - 3}\right) S^{(\alpha m - 3)\delta} \\
&\leq 8\pi c^m \, |||\rho||| \cdot S^{(\alpha m-3)\delta}.
\end{aligned}$$

Since $y \in \mathbb{R}^3$ was arbitrary, this already proves the desired bounds on $\|h^m * \rho\|_\infty$ for $m \geq 1$. By the binomial theorem, using that $|(1*\rho)(z)| = \|\rho\|_1 \leq |||\rho|||$ for all $z \in \mathbb{R}^3$, it follows that for all $x, y \in \mathbb{R}^3$, $m \geq 2$,

$$\begin{aligned}
&\left|\left(\left(h - (h*\rho)(x)\right)^m * \rho\right)(y)\right| \leq \sum_{j=0}^m \binom{m}{j} |(h*\rho)(x)|^{m-j} \cdot \left|(h^j * \rho)(y)\right| \\
&\leq C_\alpha^m \, |||\rho|||^m \, (1 + \delta_{\alpha,3} \ln(S))^m \cdot |||\rho||| + mC_\alpha^{m-1} \, |||\rho|||^{m-1} \, (1 + \delta_{\alpha,3} \ln(S))^{m-1} \cdot C_\alpha \, |||\rho||| \\
&\qquad\qquad\qquad + 8\pi \, |||\rho||| \cdot \sum_{j=2}^m \binom{m}{j} C_\alpha^{m-j} \, |||\rho|||^{m-j} \, (1 + \delta_{\alpha,3})^{m-j} \cdot c^j S^{(\alpha j - 3)\delta} \\
&\leq 8\pi C_\alpha^m \, |||\rho|||^{m+1} \cdot (1 + \delta_{\alpha,3} \ln(S))^m \cdot \sum_{j=0}^m \binom{m}{j} S^{(\alpha m - 3)\delta} \\
&\leq 8\pi \cdot 2^m C_\alpha^m \, |||\rho|||^{m+1} \cdot (1 + \delta_{\alpha,3} \ln(S))^m \cdot S^{(\alpha m - 3)\delta} \\
&= C \cdot \widetilde{C}_\alpha^m \, S^{(\alpha m-3)\delta} \cdot (1 + \delta_{\alpha,3} \ln(S))^m,
\end{aligned}$$

where $\tilde{C}_\alpha := 2C_\alpha \,|||\rho|||$ and $C := 8\pi \,|||\rho|||$. Note that we used $S \geq 1$, $\alpha \geq 2$ and $\delta \geq 0$. $\square$

In the following two lemmata, for once $|\cdot| := |\cdot|_2$ denotes the euclidean norm on $\mathbb{R}^d$ and $|\cdot|_\infty$ denotes the maximum norm on $\mathbb{R}^d$ (note that this does not make a difference for $d = 1$). Recall the elementary inequality

$$|x|_\infty \leq |x|_2 \leq \sqrt{d} \cdot |x|_\infty \qquad \forall x \in \mathbb{R}^d.$$

We denote

$$V : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}, \quad q \mapsto \frac{1}{|q|}, \qquad k : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^3, \quad q \mapsto -\frac{q}{|q|^3}$$

the Coulomb potential resp. the Coulomb force.

Lemma 4.52. *Let $\rho \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$. Then $V * \rho$ and $k * \rho$ are bounded for all $i \in [3]$.*

Proof. For every $y \in \mathbb{R}^3$,

$$|(V * \rho)(y)| \leq \int_{\mathbb{R}^3} \frac{|\rho(y)|}{|q-y|} \,\mathrm{d}y \leq \|\rho\|_\infty \cdot \int_{B_1(y)} \frac{1}{|q-y|} \,\mathrm{d}y + \int_{B_1^c(y)} |\rho(y)| \,\mathrm{d}y \leq 4\pi \|\rho\|_\infty \cdot \frac{1}{2} + \|\rho\|_1.$$

Likewise, for $i \in [3]$ and $y \in \mathbb{R}^3$,

$$|(k_i * \rho)(y)| \leq \int_{\mathbb{R}^3} \frac{|\rho(y)|}{|q-y|^2} \,\mathrm{d}y \leq 4\pi \|\rho\|_\infty + \|\rho\|_1.$$

$\square$

Remark 4.53. One can optimize the estimates by choosing a radius R dependent on $\|\rho\|_\infty, \|\rho\|_1$ instead of the radius 1 for the splitting, however, we will not need this here.

The following theorem shows that convolutions of bounded densities with the Coulomb potential are well-behaved in the sense that they allow for an interchange of integration (convolution) and differentiation. Moreover, it states that a convolution of the Coulomb force with a bounded density, i.e. the mean field Coulomb force coming from a reasonable solution to the Vlasov-Poisson system, is log-Lip-continuous in the sense of definition 4.63.

Theorem 4.54. *Let $\rho \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$. Then $V * \rho \in C^1(\mathbb{R}^3)$ with*

$$\nabla(V * \rho) = (\nabla V) * \rho = k * \rho,$$

and $(k * \rho)$ is log-Lip-continuous.

Proof. We basically follow the proof in [18, pp. 74-81]. For $h \in \mathbb{R} \setminus \{0\}$ and $i \in [3]$, let us define

$$\begin{aligned}
I(h) :&= \frac{1}{h}\big((V * \rho)(q + he_i) - (V * \rho)(q)\big) - (k_i * \rho)(q) \\
&= \frac{1}{h}\left[\int_{\mathbb{R}^3} \frac{\rho(y)}{|q + he_i - y|} \,\mathrm{d}y - \int_{\mathbb{R}^3} \frac{\rho(y)}{|q-y|} \,\mathrm{d}y\right] + \int_{\mathbb{R}^3} \frac{\rho(y)(q_i - y_i)}{|q-y|^3} \,\mathrm{d}y \\
&= \frac{1}{h}\int_{B_{2|h|}(q)} \frac{\rho(y)}{|q + he_i - y|} \,\mathrm{d}y - \frac{1}{h}\int_{B_{2|h|}(q)} \frac{\rho(y)}{|q-y|} \,\mathrm{d}y + \int_{B_{2|h|}(q)} \frac{\rho(y)(q_i - y_i)}{|q-y|^3} \,\mathrm{d}y \\
&\qquad + \int_{B^c_{2|h|}(q)} \rho(y) \cdot \frac{1}{h}\left(\frac{1}{|q + he_i - y|} - \frac{1}{|q-y|} + \frac{q_i - y_i}{|q-y|^3}\right) \mathrm{d}y \\
&=: I_1(h) + I_2(h) + I_3(h) + I_4(h).
\end{aligned}$$

First, we prove that $I(h) \in o(|h|)$. The following computations also show that $I_{1,2,3,4}(h)$ are well-defined, the deeper reason being $|\cdot|^{-1}, |\cdot|^{-2} \in L^1(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$. We estimate

$$|I_2(h)| \leq \frac{1}{|h|} \int_{B_{2|h|}(q)} \frac{|\rho(y)|}{|q-y|} \, \mathrm{d}y \leq \frac{4\pi \|\rho\|_\infty}{|h|} \cdot \frac{1}{2} \cdot (2\,|h|)^2 = 8\pi \|\rho\|_\infty \cdot |h| \,,$$

$$|I_1(h)| \leq \frac{1}{|h|} \int_{B_{3|h|}(q)} \frac{|\rho(y)|}{|q-y|} \, \mathrm{d}y \leq \frac{4\pi \|\rho\|_\infty}{|h|} \cdot \frac{1}{2} \cdot (3\,|h|)^2 = 18\pi \|\rho\|_\infty \cdot |h| \,,$$

where we used that $B_{2|h|}(q+he_i) \subset B_{3|h|}(q)$. Next, since $\frac{|q_i-y_i|}{|q-y|^3} \leq \frac{|q-y|}{|q-y|^3} = \frac{1}{|q-y|^2}$ for all $q, y \in \mathbb{R}^3$ with $q \neq y$,

$$|I_3(h)| \leq \int_{B_{2|h|}(q)} \frac{|\rho(y)|\,|q_i-y_i|}{|q-y|^3} \, \mathrm{d}y \leq 4\pi \|\rho\|_\infty \cdot 2\,|h| = 8\pi \|\rho\|_\infty \cdot |h| \,.$$

This already proves that $I_1, I_2, I_3 \in o(|h|)$. It remains to show the same for I_4. We aim to apply a version of the mean value theorem for differentiation. Note that for $y \in B^{\mathrm{c}}_{2|h|}(q)$, $|q-y| \geq 2\,|h|$ and hence by the reverse triangle inequality, for every $\theta \in [0,1]$,

$$|q + \theta h e_i - y| \geq |q-y| - |\theta h e_i| \geq 2\,|h| - |h| = |h| \,.$$

Hence, the line joining q and $q + he_i$ has a *safety distance* $|h|$ from the singularity. Applying corollary 4.75 with $f = V$ and consequently $\partial_i f = k_i$, $y = e_i$, we obtain that there is some $\theta \in (0,1)$ such that

$$\begin{aligned} \left| \frac{1}{h} \left(\frac{1}{|q+he_i-y|} - \frac{1}{|q-y|} \right) + \frac{q_i - y_i}{|q-y|^3} \right| &= \left| \frac{h}{2} \cdot \frac{3(q_i + \theta h e_i - y_i)^2 - |q+\theta h e_i - y|^2}{|q+\theta h e_i - y|^5} \right| \\ &\leq \frac{|h|}{2} \cdot \frac{4}{|q+\theta h e_i - y|^3}. \end{aligned}$$

For $|q-y| \geq 2\,|h|$, one has

$$|q+\theta h e_i - y| \geq |q-y| - |\theta h e_i| \geq |q-y| - |h| \geq |q-y| - \tfrac{1}{2}\,|q-y| = \tfrac{1}{2}\,|q-y| \qquad \forall \theta \in [0,1].$$

Consequently, for $|h| < \frac{1}{2}$,

$$\begin{aligned} |I_4(h)| &\leq \int_{B^{\mathrm{c}}_{2|h|}(y)} |\rho(y)| \cdot \frac{|h|}{2} \cdot \frac{4}{\left(\frac{1}{2}\,|q-y|\right)^3} \, \mathrm{d}y \\ &\leq 16\,|h| \cdot \left(\|\rho\|_\infty \cdot \int_{B_1(y) \setminus B_{2|h|}(y)} \frac{1}{|q-y|^3} \, \mathrm{d}q + \int_{B^{\mathrm{c}}_1(q)} |\rho(y)| \, \mathrm{d}y \right) \\ &\leq 16\,|h| \cdot \left(-\ln(2\,|h|) \cdot \|\rho\|_\infty + \|\rho\|_1 \right). \end{aligned}$$

Since $t \cdot \ln(t) \to 0$ as $h \searrow 0$ (L'Hospital's rule), we see that $|I_4(h)| \to 0$ as $h \to 0$. Altogether, this shows $I_4 \in o(h)$ and therefore existence of $\partial_i (V * \rho)$ with $\partial_i (V * \rho) = \partial_i V * \rho = k_i * \rho$.

Let us now prove that $k * \rho$ is log-Lip-continuous. Let $i \in [3]$ and $q, \tilde{q} \in \mathbb{R}^3$ with $q \neq \tilde{q}$, then for $\varepsilon := |q - \tilde{q}|$,

$$\begin{aligned} (k_i * \rho)(\tilde{q}) - (k_i * \rho)(q) &= \int_{\mathbb{R}^3} \frac{\rho(y)(q_i - y_i)}{|q-y|^3} \, \mathrm{d}y - \int_{\mathbb{R}^3} \frac{\rho(y)(\tilde{q}_i - y_i)}{|\tilde{q}-y|^3} \, \mathrm{d}y \\ = \int_{B_{2\varepsilon}(q)} \frac{\rho(y)(q_i - y_i)}{|q-y|^3} \, \mathrm{d}y &- \int_{B_{2\varepsilon}(q)} \frac{\rho(y)(\tilde{q}_i - y_i)}{|\tilde{q}-y|^3} \, \mathrm{d}y + \int_{B^{\mathrm{c}}_{2\varepsilon}(q)} \rho(y) \cdot \left(\frac{q_i - y_i}{|q-y|^3} - \frac{\tilde{q}_i - y}{|\tilde{q}-y|^3} \right) \mathrm{d}y \\ =: J_1 + J_2 + J_3. \end{aligned}$$

Again using that $\left|\frac{z_i}{|z|^3}\right| \leq |z|^{-2}$ for all $z \in \mathbb{R}^3 \setminus \{0\}$ and $B_{2\varepsilon}(q) \subset B_{3\varepsilon}(\tilde{q})$, we obtain

$$|J_1| \leq 4\pi \|\rho\|_\infty \cdot 2\varepsilon, \qquad |J_2| \leq 4\pi \|\rho\|_\infty \cdot 3\varepsilon.$$

Moreover, for $|q-y| \geq 2\varepsilon = 2|q-\tilde{q}|$, we have

$$|\tilde{q}-y| \geq |q-y| - |q-\tilde{q}| \geq |q-y| - \tfrac{1}{2}|q-y| = \tfrac{1}{2}|q-y|,$$

and therefore

$$\begin{aligned}
&\left|\frac{q_i-y_i}{|q-y|^3} - \frac{\tilde{q}_i-y_i}{|\tilde{q}-y|^3}\right| \leq \left|\frac{q_i-y_i}{|q-y|^3} - \frac{\tilde{q}_i-y_i}{|q-y|^3}\right| + \left|\frac{\tilde{q}_i-y_i}{|q-y|^3} - \frac{\tilde{q}_i-y_i}{|\tilde{q}-y|^3}\right| \\
&\leq \frac{1}{|q-y|^3} \cdot |q_i-\tilde{q}_i| + |\tilde{q}_i-y_i| \cdot \left|\frac{1}{|q-y|^3} - \frac{1}{|\tilde{q}-y|^3}\right| \\
&\leq \frac{1}{|q-y|^3} \cdot |q-\tilde{q}| + |\tilde{q}-y| \cdot \frac{|\tilde{q}-y|-|q-y|}{|q-y|\,|\tilde{q}-y|} \cdot \left(\frac{1}{|q-y|^2} + \frac{1}{|q-y|\,|\tilde{q}-y|} + \frac{1}{|\tilde{q}-y|^2}\right) \\
&\leq \frac{1}{|q-y|^3} \cdot |q-\tilde{q}| + \frac{|(\tilde{q}-y)-(q-y)|}{|q-y|} \cdot \left(\frac{1}{|q-y|^2} + \frac{1}{\frac{1}{2}|q-y|^2} + \frac{1}{\frac{1}{4}|q-y|^2}\right) \\
&= \frac{1}{|q-y|^3} \cdot (|q-\tilde{q}| + 7|q-\tilde{q}|) = \frac{1}{|q-y|^3} \cdot 8|q-\tilde{q}|,
\end{aligned}$$

were we used that for $a, b \in \mathbb{R} \setminus \{0\}$,

$$\frac{1}{a^3} - \frac{1}{b^3} = \frac{b^3-a^3}{a^3b^3} = \frac{(b-a)(b^2+ab+a^2)}{a^3b^3} = \frac{b-a}{ab} \cdot \frac{b^2+ab+a^2}{a^2b^2} = \frac{b-a}{ab} \cdot \left(\frac{1}{a^2} + \frac{1}{ab} + \frac{1}{b^2}\right).$$

Hence, for $\varepsilon < \frac{1}{2}$,

$$\begin{aligned}
|J_3| &\leq 8\|\rho\|_\infty |q-\tilde{q}| \cdot \int_{B_1(q)\setminus B_{2\varepsilon}(q)} \frac{1}{|q-y|^3}\,\mathrm{d}y + 8|q-\tilde{q}| \cdot \int_{B_1^c(q)} |\rho(y)|\,\mathrm{d}y \\
&\leq 8\big(4\pi\|\rho\|_\infty \cdot \big(-\ln(2\varepsilon)\big) + \|\rho\|_1\big) \cdot |q-\tilde{q}|.
\end{aligned}$$

Putting everything together and re-substituting $\varepsilon = |q-\tilde{q}|$, we finally arrive at

$$\begin{aligned}
\big|(k_i*\rho)(q) - (k_i*\rho)(\tilde{q})\big| &\leq 52\pi \cdot \max\{\|\rho\|_\infty, \|\rho\|_1\} \cdot |q-\tilde{q}| \cdot (1 + |\ln(|q-\tilde{q}|)|) \\
&= C|q-\tilde{q}| \cdot (1 + |\ln(|q-\tilde{q}|)|)
\end{aligned}$$

for all $|q-\tilde{q}| < \frac{1}{2}$. Since $k_i * \rho$ is also bounded (see lemma 4.52), $k_i*\rho$ is log-Lip-continuous for every $i \in [3]$ by remark 4.64. □

Corollary 4.55. *Let $\rho \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$ and f the regularized Coulomb force defined in* (3.2) *Then $f*\rho$ is bounded and Lipschitz-continuous, with*

$$\|f*\rho\|_\infty \leq C \cdot |||\rho|||, \qquad \|f*\rho\|_{\mathrm{L}} \leq C\,|||\rho||| \cdot (1+\ln(S))$$

for some $C > 0$.

Proof. The first assertion follows directly from lemma 4.51 because f satisfies a S_δ^2- condition. For the second claim, note that it suffices to prove

$$|(f*\rho)(q) - (f*\rho)(\tilde{q})| \leq C\,|||\rho||| \cdot (1+\ln(S)) \cdot |q-\tilde{q}| \qquad \forall q, \tilde{q} \in \mathbb{R}^3 : |q-\tilde{q}| < \tfrac{1}{2}S^{-\delta}$$

because for $|q-\tilde{q}| > \frac{1}{2}S^{-\delta}$, one can join q and $\tilde{q}$ by a straight line, split it into segments of length at most $\frac{1}{2}S^{-\delta}$, use the triangle inequality to apply $(*)$ on every segment and finally put everything together again. However, from lemma 4.73, we know that for $|q-\tilde{q}| < \frac{1}{2}S^{-\delta}$, one has $|f(q)-f(\tilde{q})| \leq g(q)\cdot|q-\tilde{q}|$ where g satisfies a S^3_δ- condition. Consequently, for $q,\tilde{q} \in \mathbb{R}^3$ with $|q-\tilde{q}| < \frac{1}{2}S^{-\delta}$, one has $|(q-y)-(\tilde{q}-y)| = |q-\tilde{q}| < \frac{1}{2}S^{-\delta}$ for all $y \in \mathbb{R}^3$ and thus

$$\begin{aligned}\left|(f*\rho)(q)-(f*\rho)(\tilde{q})\right| &\leq \int_{\mathbb{R}^3} |\rho(y)| \cdot |f(q-y)-f(\tilde{q}-y)| \,\mathrm{d}y \leq \int_{\mathbb{R}^3} |\rho(y)| \cdot g(q-y)\cdot|q-\tilde{q}| \,\mathrm{d}y \\ &= \left(\int_{\mathbb{R}^3} |\rho(q-y)| \cdot g(y)\,\mathrm{d}y\right)\cdot|q-\tilde{q}| = ((g*|\rho|)(q))\cdot|q-\tilde{q}| \\ &\leq C(1+\ln(S))\cdot|q-\tilde{q}|\,,\end{aligned}$$

where again we used lemma 4.51. □

Remark 4.56. Heuristically, the factor of $1+\ln(S)$ in the Lipschitz constant for the mean field coming from the regularized Coulomb force is in the limit $S\to\infty$ converted to the log-Lip-regularity of mean field induced by the true Coulomb interaction.

4.7 A high order Markov inequality

In this section, we prove an upper bound for the probability of deviations of the sample mean from the expectation value that decays with arbitrary power of the sample length:

Theorem 4.57. *Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. bounded random variables on a probability space $(\Omega,\mathcal{A},\mathbb{P})$ satisfying $\mathbb{E}[X_1]=0$ and*

$$\left|\mathbb{E}\left[X_1^m\right]\right| \leq C_1\cdot C_2^m \qquad \forall m\geq 2 \tag{4.13}$$

for some constants $C_{1,2}>0$. For $N\in\mathbb{N}$, we define the sample mean of length N,

$$X := \frac{1}{N}\sum_{i=1}^N X_i.$$

Then for every $\varepsilon>0$ and $M\in\mathbb{N}$,

$$\mathbb{P}\left[|X|\geq\varepsilon\right] \leq C_{M,N}\cdot\varepsilon^{-2M}N^{-2M}, \tag{4.14}$$

where $C_{M,N} := M^{2M+1}C_2^{2M}\cdot\max\left\{1,(NC_1)^M\right\}$.

Remark 4.58. Since we suppose that X_1 is bounded, the moments of arbitrary order m exist and are bounded by $\|X_1\|_\infty^m$. Thus, equation (4.13) gives us the chance to use better estimates for the moments, which we utilize in the main text. Note also that no matter how large C_1 is, $\mathbb{P}\left[|X|\geq\varepsilon\right]$ decays at least as fast as N^{-M} for $N\to\infty$.

Proof. The map $\mathbb{R}_0^+\to\mathbb{R},\ x\mapsto x^{2M}$ is strictly increasing for all $M\in\mathbb{N}$. Therefore, the Markov inequality (cf. [19, p. 138 f.]) yields for any $\varepsilon>0$

$$\mathbb{P}\left[|X|\geq\varepsilon\right] \leq \varepsilon^{-2M}\cdot\mathbb{E}\left[X^{2M}\right] = \varepsilon^{-2M}\cdot\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N X_i\right)^{2M}\right] = \varepsilon^{-2M}N^{-2M}\cdot\mathbb{E}\left[\left(\sum_{i=1}^N X_i\right)^{2M}\right].$$

By the multinomial theorem,

$$\left(\sum_{i=1}^{N} X_i\right)^{2M} = \sum_{\substack{a\in\mathbb{N}_0^N \\ |a|=2M}} \left[\binom{2M}{a} \cdot \prod_{i=1}^{N} X_i^{a_i}\right] =: \sum_{\substack{a\in\mathbb{N}_0^N \\ |a|=2M}} \left[\binom{2M}{a} \cdot X^a\right],$$

where for $a \in \mathbb{N}_0^N$, $\binom{2M}{a} := \frac{(2M)!}{\prod_{i=1}^N a_i!}$ (*multiindex-notation*).

Since $(X_i)_{i\in[N]}$ are independent, also $(X_i^{a_i})_{i\in[N]}$ are independent random variables for every $a \in \mathbb{N}_0^N$ (cf. [19, p. 71]), and hence we obtain that for all $a \in \mathbb{N}_0^N$,

$$\mathbb{E}\left[X^a\right] = \mathbb{E}\left[\prod_{i=1}^{N} X_i^{a_i}\right] = \prod_{i=1}^{N} \mathbb{E}\left[X_i^{a_i}\right] = \prod_{i=1}^{N} \mathbb{E}\left[X_1^{a_i}\right], \tag{4.15}$$

where in the last step we used that the X_i are identically distributed.

Let $G := \left\{a \in \mathbb{N}_0^N : |a| = 2M\right\}$. For $a \in G$ and $i \in [N]$, we define $s, s_i : G \to \mathbb{N}_0$ by

$$s_i(a) := \begin{cases} 1, & a_i = 1 \\ 0, & \text{else} \end{cases}, \qquad s(a) := \sum_{i=1}^{N} s_i(a).$$

Therefore, s *counts* the number of indices $i \in [N]$ where $a_i = 1$. Similarly, we set $t, t_i : G \to \mathbb{N}_0$,

$$t_i(a) := \begin{cases} 1, & a_i \geq 1 \\ 0, & a_i = 0 \end{cases}, \qquad t(a) := \sum_{i=1}^{N} t_i(a).$$

We see that t *counts* the number of indices where $a_i \geq 1$. Note that it is always true that $s_i(a) \leq t_i(a)$, so $s(a) \leq t(a)$ for all $a \in \mathbb{N}_0^N$. Let us split

$$G = s^{-1}(\mathbb{N}_0) = s^{-1}(\{0\}) \sqcup s^{-1}(\mathbb{N}) =: \overline{G} \sqcup G_0.$$

For $a \in G_0$, $s(a) > 0$, i.e. there is some $i_0 \in [N]$ such that $s_{i_0}(a) = 1$, which shows that $a_{i_0} = 1$ and therefore, using (4.15) with $\mathbb{E}\left[X_1\right] = 0$,

$$\mathbb{E}\left[X^a\right] = \mathbb{E}\left[X_1^{a_{i_0}}\right] \cdot \prod_{\substack{i=1 \\ i\neq i_0}}^{N} \mathbb{E}\left[X_1^{a_i}\right] = \mathbb{E}\left[X_1\right] \cdot \prod_{\substack{i=1 \\ i\neq i_0}}^{N} \mathbb{E}\left[X_1^{a_i}\right] = 0. \tag{4.16}$$

Next, note that $G_0 \supset \{a \in G : t(a) > M\}$: Assume for contradiction that $a \in G$, $t(a) > M$ and $a \notin G_0$. Then $s(a) = 0$, so $s_i(a) = 0$ for all i, which implies $a_i \in \mathbb{N}_0 \setminus \{1\}$ for all $i \in [N]$. It follows that

$$|a| = \sum_{i=1}^{N} a_i = \sum_{\substack{i=1 \\ a_i\neq 0}}^{N} a_i \geq \sum_{\substack{i=1 \\ a_i\neq 0}}^{N} 2 = 2t(a) > 2M,$$

which yields the desired contradiction. Using that $t(a) \geq 1$ for all $a \in \overline{G}$, we conclude

$$\overline{G} = G \setminus G_0 \subset G \setminus \{a \in G : t(a) > M\} = \bigcup_{l=1}^{M} \{a \in G : t(a) = l\} =: \bigcup_{l=1}^{M} A_l.$$

Now, for arbitrary $a \in G$, we estimate, using (4.13) and $\mathbb{E}\left[X^0\right] = \mathbb{E}[1] = 1$,

$$\left|\mathbb{E}\left[X^a\right]\right| = \prod_{i=1}^{N} \left|\mathbb{E}\left[X_1^{a_i}\right]\right| \leq \prod_{\substack{i=1 \\ a_i \neq 0}}^{N} C_1 \cdot C_2^{a_i} = C_1^{t(a)} \cdot C_2^{|a|} = C_1^{t(a)} \cdot C_2^{2M}.$$

Combining (4.15) and (4.16), we therefore arrive at

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{N} X_i\right)^{2M}\right] &\leq \left|\sum_{a \in G_0} \binom{2M}{a} \mathbb{E}\left[X^a\right]\right| + \left|\sum_{a \in \overline{G}} \binom{2M}{a} \mathbb{E}\left[X^a\right]\right| \leq \sum_{a \in \overline{G}} \binom{2M}{a} C_1^{t(a)} \cdot C_2^{2M} \\
&\leq \sum_{l=1}^{M} \sum_{a \in A_l} \binom{2M}{a} C_1^l \cdot C_2^{2M} = C_2^{2M} \cdot \sum_{l=1}^{M} C_1^l \cdot \sum_{b \in B_l} \left|\gamma_l^{-1}(b)\right| \cdot \binom{2M}{b} \\
&\leq C_2^{2M} \cdot \sum_{l=1}^{M} C_1^l N^l \cdot l^{2M} \leq M C_2^{2M} \cdot \max\left\{1, (NC_1)^M\right\} \cdot M^{2M}.
\end{aligned}$$

There are several nontrivial arguments which we used in this computation:

(i) For $l \in \mathbb{N}$, consider the map

$$\gamma_l : A_l = \left\{a \in \mathbb{N}_0^N : |a| = 2M,\ t(a) = l\right\} \to \left\{b \in \mathbb{N}_0^l : |b| = 2M,\ b_i \neq 0\ \forall i \in [l]\right\} := B_l,$$
$$(a_1, \ldots, a_N) \mapsto (a_{i_1}, \ldots, a_{i_l}),$$

where $\{i_1, \ldots, i_l\} = \{i \in [N] : a_i \neq 0\}$ and $i_1 < \ldots < i_l$. Let $a \in A_l$ and $a_{i_1} < \ldots < a_{i_l}$ such that $a_{i_k} \neq 0$ for all $k \in [l]$. Then

$$a! = \prod_{i=1}^{N} a_i! = \prod_{\substack{i=1 \\ a_i \neq 0}}^{N} a_i! = \prod_{k=1}^{l} a_{i_k}! = \prod_{k=1}^{l} ((\gamma_l(a))_k)! = (\gamma_l(a))!$$

and thus $\binom{2M}{a} = \binom{2M}{\gamma_l(a)}$.

(ii) For $l \in \mathbb{N}$ and $b \in B_l$, $\left|\gamma_l^{-1}(b)\right| \leq N^l$ because the map

$$\left\{(i_1, \ldots, i_l) \subset [N]^l : i_1 < \ldots < i_l\right\} \to \gamma_l^{-1}(b), \qquad (i_1, \ldots, i_l) \mapsto (a_1, \ldots, a_N)$$

where

$$a_j := \begin{cases} b_k, & j = i_k \text{ for some (and therefore a unique) } k \in [l] \\ 0, & \text{else} \end{cases}$$

is obviously well-defined and surjective (even bijective) and thus

$$N^l = \left|[N]^l\right| \geq \left|\left\{(i_1, \ldots, i_l) \subset [N]^l : i_1 < \ldots < i_l\right\}\right| \geq \left|\gamma_l^{-1}(b)\right|.$$

(iii) By the multinomial theorem,

$$l^{2M} = \left(\sum_{k=1}^{l} 1\right)^{2M} = \sum_{\substack{b \in \mathbb{N}_0^l \\ |b| = 2M}} \left(\binom{2M}{b} \cdot \prod_{i=1}^{l} 1^{b_i}\right) = \sum_{\substack{b \in \mathbb{N}_0^l \\ |b| = 2M}} \binom{2M}{b} \geq \sum_{b \in B_l} \binom{2M}{b}.$$

Putting everything together, we finally arrive at

$$\mathbb{P}\left[|X| \geq \varepsilon\right] \leq \varepsilon^{-2M} N^{-2M} \cdot M C_2^{2M} \cdot \max\left\{1, (NC_1)^M\right\} \cdot M^{2M} = C_{M,N} \cdot \varepsilon^{-2M} N^{-2M},$$

as claimed. □

4.8 Global existence of flows

In this section, we prove a useful result which grants global existence (and also uniqueness) of flows for a large class of ODEs. Let us first agree on some terminological issues, mainly concerning differentiability and therefore the property of being a solution at the boundary of an interval:

Definition 4.59. *Let $d, d' \in \mathbb{N}$. We call a map $f : \mathbb{R}^d \supset U \to \mathbb{R}^{d'}$ (continuously) differentiable if there is some $\tilde{U} \supset U$ open and a (continuously) differentiable $\tilde{f} : \tilde{U} \to \mathbb{R}^{d'}$ such that $\tilde{f}|_U = f$. In this case, we define $f' := (\tilde{f})'|_U$.*

Remark 4.60. Let $I \subset \mathbb{R}$ be right-open and $f : I \to \mathbb{R}^d$ (continuously) differentiable in the sense of the previous definition. Then f is (right-continuously) right-sided differentiable on I, with $f^{\prime,+} = (\tilde{f})'|_U$. The proof is straightforward, requiring only the definitions 4.1 and 4.59 and thus left to the reader as an easy exercise.

Definition 4.61. *Let $(x_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$ and $U \subset \mathbb{R}^d$ a neighbourhood of x_0, $I \subset \mathbb{R}$ an interval containing t_0. Let $F \in \mathcal{C}(U \times I; \mathbb{R}^d)$. We call a map $x : I \supset J \to U$ a* ***(local) solution*** *to the initial value problem*

$$\dot{x}(t) = F(x(t), t), \qquad x(t_0) = x_0 \tag{4.17}$$

if $t_0 \in J$, $x(t_0) = x_0$ and $x \in \mathcal{C}^1(J; U)$ in the sense of definition 4.59 with $\dot{x}(t) = F(x(t), t)$ for all $t \in J$. It is called ***global solution*** *if $J = I$.*

Definition 4.62. *Let $F \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$. A map $\Phi : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ is called the* ***global flow*** *associated with the ODE $\dot{x}(t) = F(x(t), t)$ if*

(i) for every $t_0 \in \mathbb{R}$, $\Phi(\cdot\,, t_0, t_0) = \mathrm{id}_{\mathbb{R}^d}$ and

(ii) for every $(x_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$, the map $\Phi_{x_0,t_0} : \mathbb{R} \to \mathbb{R}^d$, $t \mapsto \Phi(x_0, t, t_0)$ is continuously differentiable with

$$\frac{\mathrm{d}}{\mathrm{d}t}\, \Phi_{x_0,t_0}(t) = F(\Phi_{x_0,t_0}(t), t), \tag{4.18}$$

i.e. Φ_{x_0,t_0} is a global solution to the IVP (4.17).

Definition 4.63. *Let $d, d' \in \mathbb{N}$ and $G : \mathbb{R}^d \to \mathbb{R}^{d'}$ a map. We say that G is* ***log-Lip-continuous*** *if there is some $C > 0$ such that for all $x, y \in \mathbb{R}^d$ with $x \neq y$,*

$$\frac{|G(x) - G(y)|}{|x - y|} \leq C \cdot \left(1 + \left|\ln\left(|x - y|\right)\right|\right). \tag{4.19}$$

In this case, we sometimes say that C is a (uniform) log-Lip-bound for G.

Remark 4.64.

(a) As for Lipschitz-continuity, the definition of log-Lip-continuity does not depend on the choice of norms on $\mathbb{R}^d$ and $\mathbb{R}^{d'}$ (however, the property of being a log-Lip-bound for G certainly does!), and taking the maximum norm on $\mathbb{R}^{d'}$, we see that G is log-Lip-continuous if and only if every component $G_i : \mathbb{R}^d \to \mathbb{R}$, $i \in [d']$, is log-Lip-continuous.

(b) Every Lipschitz-continuous map is clearly log-Lip-continuous. However, it is immediate that in general, the reverse direction fails to hold.

(c) Obviously, $(*)$ is equivalent to the statement

$$|G(x) - G(y)| \leq C \cdot \theta(|x-y|) \qquad \forall x, y \in \mathbb{R}^d,$$

where θ is defined in lemma 4.77. Using that $\theta(t) \searrow 0$ as $t \searrow 0$ (which is proved in the very same lemma), we see that every log-Lip-continuous function is in particular continuous, therefore justifying the nomenclature.

(d) If G is bounded, an equivalent definition of log-Lip-continuity is given by existence of some $C' > 0$ such that for all $x, y \in \mathbb{R}^d$ with, say, $0 < |x-y| \leq \frac{1}{2}$,

$$|G(x) - G(y)| \leq C' \cdot \theta(|x-y|). \tag{$*$}$$

Indeed, log-Lip-continuity of G certainly implies $(*)$. On the other hand, assuming boundedness of G and the validity of $(*)$ for $0 < |x-y| < \frac{1}{2}$, then for all $x, y \in \mathbb{R}^d$ with $|x-y| \geq \frac{1}{2}$,

$$\begin{aligned}|G(x) - G(y)| &\leq 2\,\|G\|_\infty \leq 2\,\|G\|_\infty \cdot \frac{|x-y|}{\frac{1}{2}} = 4\,\|G\|_\infty \cdot |x-y| \\ &\leq 4\,\|G\|_\infty \cdot |x-y| \cdot \big(1 + \left|\ln\left(|x-y|\right)\right|\big) \\ &= 4\,\|G\|_\infty \cdot \theta(|x-y|).\end{aligned}$$

In the literature, one usually finds that for $F \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ with $F(\,\cdot\,, t)$ Lipschitz continuous for every t, local solutions to (4.61) exist and are unique, and if the Lipschitz constant can be chosen uniformly in t, the solutions can be shown to exist globally, see e.g. [34, p. 113] (note that Lipschitz continuity implies linear boundedness by a computation which in similar form will be given below for the log-Lip case). Thus, patching these solutions together in the way suggested by (4.18), one obtains existence and uniqueness of a global flow Φ. We want to show that this is also true when $F \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ is only log-Lip-continuous uniformly in t, which in the light of remark 4.64 (b) is clearly a generalization. The reason why we aim for this result is the following: it turns out that in general, the mean field Coulomb force generated by bounded spatial densities is only log-Lip-continuous in the spatial argument, see also theorem 4.54. Consequently, the following theorem, together with the corresponding generalization of Liouville's theorem in the next section, a posteriori provides the basis of all the heuristics in chapter 1 and prepares for a thorough analysis of the Vlasov-Poisson system as suggested in section 2.3.

Theorem 4.65. *Let $F \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ such that F is - uniformly in the second variable - log-Lip-continuous in the first variable, i.e. there is some $C > 0$ such that for all $x, y \in \mathbb{R}^d$ and $t \in \mathbb{R}$,*

$$|F(x,t) - F(y,t)| \leq C \cdot \theta(|x-y|).$$

Then there exists a unique, global flow Φ associated with the ODE

$$\dot{x}(t) = F(x(t), t). \tag{4.20}$$

Proof. We first prove local existence and uniqueness of solutions to the IVP (4.17) and then argue that in fact, the local solutions extend to global solutions. By local uniqueness, one can then aggregate these to a global flow with standard ODE arguments (actually, (4.18) already suggests how to do this).

(1) Let $(x_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$. By Peano's existence theorem, for $\delta > 0$ small enough there is at least one solution $x : [t_0 - \delta, t_0 + \delta] \to \mathbb{R}^d$ of the IVP $\dot{x}(t) = F(x(t), t),\ x(t_0) = x_0$. On the other hand, letting $\omega := C \cdot \theta$, our assumption of uniform log-Lip-continuity reads $|F(x,t) - F(y,t)| \leq \omega(|x-y|)$ for all $x, y \in \mathbb{R}^d$ and $t \in \mathbb{R}$. Now, lemma 4.77 shows that ω and thus $C \cdot \omega$ satisfies all the hypotheses of *Osgood's condition*, cf. [40, p. 146-147]. Hence, we also obtain that there is at most one local solution of the IVP, i.e. altogether we get local existence and uniqueness.

(2) Next, we want to prove that the local solutions from (1) are in fact global. In any introductory course on ordinary differential equations, one proves that provided there is local uniqueness of solutions, one can define maximal intervals of existence for a solution $x(t)$ where $x(t_0) = x_0$, which are open and denoted by $I_{\max}(x_0, t_0) := (\alpha(x_0, t_0), \beta(x_0, t_0))$ satisfy $\alpha, \beta : \mathbb{R}^d \times \mathbb{R} \to \overline{\mathbb{R}}$ with $\alpha(x_0, t_0) < t_0 - \delta < t_0 + \delta < \beta(x_0, t_0)$. Since F is continuous and defined on all of $\mathbb{R}^d \times \mathbb{R}$, one can employ a standard argument (basically local extension using Peano's existence theorem) which shows that provided $\beta(x_0, t_0) < \infty$, $|x(t)| \to \infty$ as $t \nearrow \beta(x_0, t_0)$. Therefore, if we could show that $|x(t)|$ actually stays bounded as $t \nearrow \beta(x_0, t_0)$ for every $\beta(x_0, t_0) \in \mathbb{R}$, then we would see that $\beta(x_0, t_0) = +\infty$. So, let us assume for contradiction that $\beta(x_0, t_0) < \infty$, then $|x(t)| \to \infty$ as $t \nearrow \beta(x_0, t_0)$. Choose $\tau \in [t_0, \beta(x_0, t_0))$ such that for all $t \in [\tau, \beta(x_0, t_0))$, $|x(t)| \geq 1$ and consequently $\theta(|x(t)|) \geq 1$. Let $\tilde{C} := \sup_{t \in [t_0, \beta(x_0, t_0)]} |F(0,t)| + C$, then $\tilde{C} < \infty$ by continuity of F (the set $0 \times [\tau, \beta(x_0, t_0) \subset \mathbb{R}^d \times \mathbb{R}$ is compact!). Consequently, for every $t \in [\tau, \beta(x_0, t_0))$, with corollary 4.19 and the triangle inequality, we obtain

$$\begin{aligned} \partial_t^+ |x(t)| &\leq |\dot{x}(t)| = |F(x(t), t)| \leq |F(x(t), t) - F(0, t)| + |F(0, t)| \\ &\leq C \cdot \theta(|x(t) - 0|) + |F(0, t)| \leq \tilde{C} \cdot \theta(|x(t)|). \end{aligned}$$

Now, on $[\tau, \beta(x_0, t_0)]$ consider the IVP

$$\dot{u}(t) = \tilde{C} \cdot \theta(u(t)), \qquad u(\tau) = |x(\tau)|. \tag{$*$}$$

By our comparison theorem 4.5, $|x(t)| \leq u(t)$ for all $t \in [\tau, \beta(x_0, t_0))$. But from lemma 4.78, we can easily deduce that the unique solution to $(*)$ is

$$u(t) = \exp\Big(\exp\Big(\tilde{C}(t - \tau) + \ln\big(1 + \ln\big(|x(\tau)|\big)\big)\Big) - 1\Big) \qquad \forall t \in [\tau, \beta(x_0, t_0)).$$

In particular, $u(t)$ and thus $|x(t)|$ remains bounded as $t \nearrow \beta(x_0, t_0)$, which gives the desired contradiction.

By similar arguments, we can show that in fact, $\alpha(x_0, t_0) = -\infty$. Hence, we see that the solution $x(t)$ exists globally, as claimed.

□

Remark 4.66. From local uniqueness, one can easily prove that $\Phi(\,\cdot\,, t, t_0)$ is bijective for all $t, t_0 \in \mathbb{R}$, and that $(\Phi(\,\cdot\,, t, t_0))^{-1} = \Phi(\,\cdot\,, t_0, t)$, a result which will frequently be used in the following section. This relation does also show that Φ and Φ^{-1} have *the same regularity properties*. In general, the regularity of Φ in $x \in \mathbb{R}^d$ is only as good as the regularity of F in $x \in \mathbb{R}^d$, i.e. under the hypotheses of the previous theorem, Φ depends continuously on $x \in \mathbb{R}^d$. If F is continuously differentiable w.r.t. $x \in \mathbb{R}^d$, then so is Φ, a fact which will be needed in section 4.9. The proof of these statements can be found, for instance, in chapter V of [20].

4.9 Liouville's theorem

Liouville's theorem states that in the sense of definition 4.31, the flows associated to a large class of ODEs, including the typical evolution equations in classical mechanics, are measure preserving (see definition 4.31) w.r.t. Lebesgue measure on phase space. One can find various versions of Liouville's theorem in textbooks on (mathematical) physics, see e.g. [5, p. 68f.] for a "classical derivation" and [16, p. 343] for a derivation within the formalism of symplectic manifolds. However, in the proofs which one usually encounters, it is required that the ODE is autonomous and the r.h.s. is at least continuously differentiable, which is in general not the case for the forces we typically encounter in this thesis. Note that the latter assumption is often hidden in the requirement that the Hamiltonian of the corresponding physical system be smooth or at least a $\mathcal{C}^2$-function. In order to make Liouville's theorem applicable to a system where the pair interaction is a Coulomb force with cut-off as introduced in section 3.2, which is only Lipschitz-continuous, or a mean field Coulomb force, which is time-dependent and in general not even Lipschitz continuous, we need to prove the statement in a more general setting.

So, let us consider a physical system governed by the evolution equations

$$\dot{q}(t) = F_1(p(t), t), \qquad \dot{p}(t) = F_2(q(t), t),$$

where $F_{1,2} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$. As usual, we combine these equations to an ODE on phase space $\mathbb{R}^{2d} = \mathbb{R}^d \times \mathbb{R}^d$, letting $x(t) := (q(t), p(t))$ and obtaining $\dot{x}(t) = F(x(t), t)$ with

$$F : \mathbb{R}^{2d} \times \mathbb{R} \to \mathbb{R}^{2d}, \qquad (q, p, t) \mapsto \big(F_1(p, t), F_2(q, t)\big).$$

Observe that typically, F_1 is smooth and linearly bounded in p uniformly in t because it has the form $F_1(p_1, \dots, p_d, t) = \big(\frac{p_1}{m_1}, \dots, \frac{p_d}{m_d}\big)$ where m_i, $i \in [d]$, are the particle masses. On the other hand, the forces which are relevant in the main text are, as already discussed above, continuous, bounded and (log-) Lipschitz continuous in q uniformly in t. Let us take the weakest of all these conditions, namely that F_1, F_2 and hence F are continuous with $F_1(\,\cdot\,, t), F_2(\,\cdot\,, t)$ and therefore $F(\,\cdot\,, t)$ satisfying a log-Lip-condition uniformly in t. From our results in section 4.8, we already know that also under this quite weak assumption, the ODE $\dot{x}(t) = F(x(t), t)$ admits a unique, global flow. Now, the crucial observation is that by construction of F, for all $j \in [2d]$, the j-th component $(F(\,\cdot\,, t))_j$ of F is independent of x_j. With this in mind, we are ready to formulate the desired generalization of Liouville's theorem. Note that the dimension d in the theorem corresponds to $2d$ in the preceding discussion.

Theorem 4.67. *Let $d \in \mathbb{N}$ and $F \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ such that F satisfies a log-Lip-condition uniformly in t. Moreover, assume that for all $j \in [d]$ and $t \in \mathbb{R}$, the j-th component $F_j(\,\cdot\,, t) : \mathbb{R}^d \to \mathbb{R}$ of $F(\,\cdot\,, t)$ does not depend on the j-th coordinate, i.e. for all $y \in \mathbb{R}^{d-1}$, the map*

$$F_{j,t,y} : \mathbb{R} \to \mathbb{R}, \qquad s \mapsto \big(F(y_1, \dots, y_{j-1}, s, y_j, \dots, y_{d-1}, t)\big)_j$$

is constant. Let $\Phi : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^d$ denote the unique global flow associated with the ODE $\dot{x}(t) = F(x(t), t)$, the existence of which is guaranteed by theorem 4.65. Then for every $t, t_0 \in \mathbb{R}$, the map $\Phi_{t,t_0} : \mathbb{R}^d \to \mathbb{R}^d$, $x \mapsto \Phi(x, t, t_0)$ is measure preserving w.r.t. d-dimensional Lebesgue measure.

Proof. The proof is divided into two steps: First, we show the claim in case we additionally have $F \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ for all $t \in \mathbb{R}$, and second we relax these assumptions to match our hypotheses.

For the whole proof, $t_0 \in \mathbb{R}$ will be fixed, so in the following, we will omit the t_0-dependence of Φ in our notation.

(1) As announced, suppose that $F \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$. Then by [20, p. 95-98], the formal computation

$$\partial_t(D_x\Phi(x,t)) = D_x(\partial_t\Phi(x,t)) = D_x(F(\Phi(x,t),t)) = D_xF(\Phi(x,t),t) \cdot D_x\Phi(x,t) \qquad (4.21)$$

is justified in the sense that the mixed partial derivatives $\partial_t\partial_{x_i}\Phi(x,t)$ and $\partial_{x_i}\partial_t\Phi(x,t)$ exist and agree for all $i \in [d]$. Now, fix $t \in \mathbb{R}$, then as we have already discussed in remark 4.66, $\Phi_t : \mathbb{R}^d \to \mathbb{R}^d$, $x \mapsto \Phi(x,t)$ is bijective and continuous with continuous inverse and consequently a diffeomorphism on $\mathbb{R}^d$. By comparing the *substitution rule* (also known as *transformation formula*)

$$\int_{\mathbb{R}^d} f \,\mathrm{d}x = \int_{\Phi_t^{-1}(\mathbb{R}^d)} (f \circ \Phi_t) \cdot |\det(D\Phi_t)| \,\mathrm{d}x = \int_{\mathbb{R}^d} (f \circ \Phi_t) \cdot |\det(D\Phi_t)| \,\mathrm{d}x,$$

where $f \in L^1(\mathbb{R}^d)$, see e.g. [4, p. 195], with the characterization (4.9) of measure-preserving maps, we see that we need only show that $|\det(D\Phi_t(x))| = 1$ for all $x \in \mathbb{R}^d$. Since $\Phi_{t_0} = \mathrm{id}_{\mathbb{R}^d}$, we have $\det(D\Phi_{t_0}(x)) = \det(\mathbb{1}^{d\times d}) = 1$ for all $x \in \mathbb{R}^d$, so it suffices to prove that for all $x \in \mathbb{R}$, the map $s \mapsto \det(D\Phi_s(x))$ is differentiable with $\frac{\mathrm{d}}{\mathrm{d}s}\det(D\Phi_s(x)) \equiv 0$. Indeed, since Φ_s is a diffeomorphism, $D\Phi_s(x) \in \mathrm{GL}(d,\mathbb{R})$ for all $x \in \mathbb{R}^d$, $s \in \mathbb{R}$. Moreover, we just saw in (4.21) that for fixed $x \in \mathbb{R}^d$, the map $s \mapsto D\Phi_s(x) = D_x\Phi(x,s)$ is differentiable with derivative $\partial_s D\Phi_s(x) = D_xF(\Phi_s(x),s) \cdot D\Phi_s(x)$ Using Jacobi's formula (theorem 4.69) for the time derivative of a determinant, we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\det(D\Phi_s(x)) &= \det(D\Phi_s(x)) \cdot \mathrm{tr}\left(\partial_s(D\Phi_s(x)) \cdot (D\Phi_s(x))^{-1}\right)\\ &= \det(D\Phi_s(x)) \cdot \mathrm{tr}\left(D_xF(\Phi_s(x),s)\right)\\ &= \det(D\Phi_s(x)) \cdot \sum_{i=1}^d \partial_{x_i}F_i(\Phi_s(x),s) = 0,\end{aligned}$$

where in the last step we finally used that by assumption, F_i is independent of x_i.

(2) Now, let $F \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ satisfy a log-Lip-condition uniformly in t, i.e. there is some $C > 0$, which w.l.o.g. may be assumed to satisfy $C \geq 1$, such that

$$|F(x,t) - F(y,t)| \leq C \cdot \theta(|x-y|) \qquad \forall x,y \in \mathbb{R}^d,\ t \in \mathbb{R},$$

where $\theta : \mathbb{R} \to \mathbb{R}$ is defined in 4.77. The following proof is inspired by [6, p. 61-66], but the fact that we do not even have Lipschitz continuity of $F(\,\cdot\,,t)$ requires some major adjustments. However, the central idea of the proof, namely to approximate F by smooth maps F^n, apply step (1) to the flows corresponding to the smooth maps and show that the property of being measure preserving survives in the limit $n \to \infty$ because the flows associated with F^n converge to the flows associated with F in an adequate sense as $n \to \infty$, remains unchanged.

Let us again fix $t \in \mathbb{R}$. First, we need to argue that Φ_t is measurable. In fact, Φ_t is even continuous: This follows from the discussion in remark 4.66 with continuity of F w.r.t. the x-component, however, it yields some insights to prove this by hand here: For $x, y, \in \mathbb{R}^d$ with $x \neq y$, one has (corollary 4.19)

$$\partial_t^+\big|\Phi_t(x) - \Phi_t(y)\big| \leq \big|F(\Phi_t(x),t) - F(\Phi_t(y),t)\big| \leq C \cdot \theta\big(\big|\Phi_t(x) - \Phi_t(y)\big|\big) \qquad \forall t \in \mathbb{R}.$$

By the comparison theorem 4.5 and lemma 4.78, noting that the solution $x(t)$ of the IVP $\dot{x}(t) = C\cdot\theta(x(t))$, $x(t_0) = |x-y| = |\Phi_{t_0}(x) - \Phi_{t_0}(y)|$ given in the lemma is the unique and therefore also the maximal solution, it follows that $|\Phi_t(x) - \Phi_t(y)| \le x(t) \to 0$ as $|x-y| \to 0$.

We want to prove that Φ_t preserves d-dimensional Lebesgue measure λ^d. Since it is common knowledge that $\mathcal{E} := \{E \subset \mathbb{R} : E \text{ compact}\}$ is a $\cap$-stable generator of $\mathcal{B}(\mathbb{R}^d)$, by [12, p. 39] it suffices to prove that $\lambda^d|_{\mathcal{E}} = \lambda^d \circ \Phi_t^{-1}|_{\mathcal{E}}$. Moreover, again from remark 4.66, we know that Φ_t is invertible with continuous and therefore measurable inverse. With remark 4.32, we conclude that we may prove $\lambda^d(E) = \lambda^d(\Phi_t(E))$ for all $E \in \mathcal{E}$ instead.

For this, let $E \subset \mathbb{R}^d$ be compact, then E is bounded by the Heine-Borel theorem, i.e. there is some $r > 0$ such that $E \subset \overline{B}_r(0) := \left\{x \in \mathbb{R}^d : |x| \le r\right\}$ (as usual, we equip $\mathbb{R}^d$ with the maximum norm $|\cdot|$). We claim that there is some $R > 0$ such that for all $G \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R}; \mathbb{R}^d)$ with $\|F - G\|_\infty \le 1$, there is a unique global flow Ψ associated to the ODE $\dot{x}(t) = G(x(t), t)$ satisfying $\Psi(\overline{B}_r(0) \times [t_0, t]) \subset \overline{B}_R(0)$. Indeed, there is always a unique local solution (G is continuous and continuously differentiable and in particular locally Lipschitz w.r.t. the first variable, so the Picard-Lindelöf theorem applies), and for $s \in [t_0, t]$ and ψ a local solution with $\psi(s) \in B_1^{\mathrm{c}}(0)$,

$$\begin{aligned}|G(\psi(s), s)| &\le |G(\psi(s), s) - F(\psi(s), s)| + |F(\psi(s), s) - F(0, s)| + |F(0, t)| \\ &\le 1 + C\cdot\theta(|\psi(s)|) + \sup_{s\in[t_0,t]} |F(0, s)| \le \overline{C}\cdot\theta(|\psi(s))| \,.\end{aligned}$$

Thus, it is easy to see from the same argument as in the second part of the proof of theorem 4.65 that ψ does in fact exist globally, and from the same estimate, one can deduce that for

$$R := \exp\left(\exp\left(\overline{C}(t - t_0) + \ln(1 + r)\right) - 1\right),$$

$\psi(s) \in \overline{B}_R(0)$ for all $s \in [t_0, t]$ (it certainly suffices to consider the parts of the trajectories which are contained in $B_1^{\mathrm{c}}(0)$). We conclude that a unique, global flow Ψ for the ODE $\dot{x}(t) = G(x(t), t)$ does indeed exist. However, reviewing the argumentation which we just applied shows that actually, the behaviour of G on $B_R^{\mathrm{c}}(0) \times [t_0, t]$ is not relevant for trajectories starting in $\overline{B}_r(0)$, so the conclusions hold for any $G \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R})$ which allows for a global flow and satisfies $\sup\left\{\left|G(x, s) - F(x, s) : (x, s) \in \overline{B}_R(0) \times [t_0, t]\right|\right\} \le 1$. We will need this later on.

Let $i \in [d]$. Since $\overline{B}_R(0) \times [t_0, t] \subset \mathbb{R}^{d+1}$ is compact and $F_i : \mathbb{R}^d \times \mathbb{R} \cong \mathbb{R}^{d+1} \to \mathbb{R}$ is by assumption continuous and does not depend on the $i+1^{\text{st}}$ coordinate, lemma 4.70 guarantees for all $n \in \mathbb{N}$ the existence of smooth functions (even polynomials) $\tilde{F}_i^n : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ not depending on x_i such that

$$\sup\left\{\left|F_i(x, s) - \tilde{F}_i^n(x, s)\right| : (x, s) \in \overline{B}_R(0) \times [t_0, t]\right\} < \tfrac{1}{n} \qquad \forall i \in [d].$$

Define $\tilde{F}^n := (\tilde{F}_1^n, \ldots, \tilde{F}_d^n) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, then $\tilde{F}^n$ is by construction smooth and

$$\sup\left\{\left|F(x, s) - \tilde{F}^n(x, s)\right| : (x, s) \in \overline{B}_R(0) \times [t_0, t]\right\} < \tfrac{1}{n}.$$

However, a global flow to the ODE with r.h.s. $\tilde{F}^n$ only exists if $\tilde{F}^n$ does not grow to fast at infinity; we ensure this by making $\tilde{F}^n$ compactly supported in the x-component as follows: Let $\eta \in \mathcal{C}^\infty(\mathbb{R}^d; [0, 1])$ such that

$$\eta|_{\overline{B}_R(0)} \equiv 1, \qquad \eta|_{\left(\overline{B}_{2R}(0)\right)^{\mathrm{c}}} \equiv 0.$$

Existence of such a map is a standard exercise in various lectures. For $(x, s) \in \mathbb{R}^d \times \mathbb{R}$, let $F^n(x, s) := \eta(x) \cdot \tilde{F}^n(x, s)$, then by construction, $F^n(\,\cdot\,, s)$ and $\tilde{F}^n(\,\cdot\,, s)$ coincide on $\overline{B}_R(0)$ for all $s \in \mathbb{R}$ and $n \in \mathbb{N}$. F^n is by construction smooth and compactly supported in the first argument, in particular, the derivatives w.r.t the first argument are bounded uniformly in $s \in [t_0, t]$. Therefore, F^n is Lipschitz-continuous in x uniformly in t and thus satisfies the hypotheses of theorem 4.65. It follows that the ODE $\dot{x}(t) = F^n(x(t), t)$ admits a unique, global flow Φ^n. Moreover, it is clear that

$$\sup\left\{\left|F^n(x, s) - F(x, s)\right| : (x, s) \in \overline{B}_R(0) \times [t_0, t]\right\} < \frac{1}{n} \leq 1.$$

From our above discussion, we deduce that $\Phi^n_s(\overline{B}_r(0)) \subset \overline{B}_R(0)$ for all $s \in [t_0, t]$, $n \in \mathbb{N}$. We now claim that Φ^n_t is close to Φ_t in the sense that

$$\sup\left\{|\Phi_t(x) - \Phi^n_t(x)| : x \in \overline{B}_r(0)\right\} \xrightarrow{n\to\infty} 0. \tag{4.22}$$

Indeed, for any $x \in \overline{B}_r(0)$ and $n \in \mathbb{N}$, we have already argued that $\Phi^n_s(x), \Phi_s(x) \in \overline{B}_R(0)$ for all $s \in [t_0, t]$, and for these s,

$$\begin{aligned}
\partial^+_s\left|\Phi^n_s(x) - \Phi_s(x)\right| &\leq \left|F^n(\Phi^n_s(x), s) - F(\Phi_s(x), s)\right| \\
&\leq \left|F^n(\Phi^n_s(x), s) - F(\Phi^n_s(x), s)\right| + \left|F(\Phi^n_s(x), s) - F(\Phi_s(x), s)\right| \\
&\leq \frac{1}{n} + C \cdot \theta\big(\left|\Phi^n_s(x) - \Phi_s(x)\right|\big).
\end{aligned}$$

Also, we have that $\left|\Phi^n_{t_0}(x) - \Phi_{t_0}(x)\right| = |x - x| = 0$. Using theorem 4.5, we may conclude that $|\Phi^n_t(x) - \Phi_t(x)| \leq u_n(t)$ were $u_n(t)$ is the unique and therefore maximal solution to the IVP

$$\dot{u}_n(t) = C \cdot \theta(u_n(t)) + \frac{1}{n}, \qquad u_n(0) = 0.$$

By lemma 4.79, $\lim_{n\to\infty} u_n(t) = 0$. In particular, given $\delta > 0$ arbitrary, there is some $N \in \mathbb{N}$ such that for all $n \geq N$, $0 \leq u_n(t) \leq \delta$. Consequently, for $n \geq N$, $|\Phi^n_t(x) - \Phi_t(x)| \leq \delta$, i.e. indeed,

$$\sup\left\{\left|\Phi^n_t(x) - \Phi_t(x)\right| : x \in \overline{B}_r(0)\right\} \leq \delta \qquad \forall n \geq N,$$

which proves (4.22).

It remains to transfer the property of being measure preserving from the Φ^n_t to Φ_t. In the following, we will write $|E| := \lambda^d(E)$. For $\delta > 0$ and $U \subset \mathbb{R}^d$, let

$$B_\delta(U) := \left\{x \in \mathbb{R}^d : \operatorname{dist}(x, U) := \inf_{u \in U} |x - u| < \delta\right\}.$$

Then $B_\delta(U) \searrow \overline{U}$ as $\delta \searrow 0$: for $0 \leq \delta_1 \leq \delta_2$, $B_{\delta_1}(U) \subset B_{\delta_2}(U)$ and $\bigcap_{\delta>0} B_\delta(U) = \overline{U}$. Since E is compact, so is $\Phi_t(E)$, and thus in particular, $\Phi_t(E)$ is closed and bounded. Hence, $B_\delta(\Phi_t(E))$ is also bounded for every $\delta > 0$, and it is easily checked to be open and hence measurable. This implies that $|B_\delta(\Phi_t(E))| < \infty$ for any $\delta > 0$, and continuity of the measure from above implies that $|B_\delta(\Phi_t(E))| \searrow |\Phi_t(E)|$ as $\delta \searrow 0$. In particular, we can find some $\delta > 0$ such that $|B_\delta(\Phi_t(E))| \leq |\Phi_t(E)| + \varepsilon$. By (4.22), we may find some $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\sup\left\{\left|\Phi_t(x) - \Phi^n_t(x)\right| : x \in E\right\} \leq \sup\left\{\left|\Phi_t(x) - \Phi^n_t(x)\right| : x \in \overline{B}_R(0)\right\} < \delta.$$

This shows that for all $n \geq N$, $\Phi_t^n(E) \subset B_\delta(\Phi_t(E))$. Consequently, using that Φ_t^n is measure preserving,

$$|\Phi_t(E)| + \varepsilon \geq |B_\delta(\Phi_t(E))| \geq |\Phi_t^n(E)| = |E| \,.$$

Since $\varepsilon > 0$ was arbitrary, we obtain $|\Phi_t(E)| \geq |E|$. However, all arguments that we have used yet do also to $(\Phi(\,\cdot\,, t, t_0))^{-1} = \Phi(\,\cdot\,, t_0, t)$, too, so we may also conclude that

$$|E| = \left|\Phi_t^{-1}\left(\Phi_t(E)\right)\right| \geq |\Phi_t(E)| \,.$$

This finally proves that $|\Phi_t(E)| = |E|$, i.e. Φ_t is indeed measure preserving.

□

4.10 Miscellanea

In this section, we collect various auxiliary results which do not really fit into one of the previous sections. We start with proving Cramer's rule:

Lemma 4.68 (Cramer's rule)**.** *Let $A \in \mathrm{GL}(d, \mathbb{R})$ and $b \in \mathbb{R}^d$. Then for $i \in [d]$, the i-th component x_i of the unique solution $x \in \mathbb{R}^d$ to the equation $A \cdot x = b$ is given by*

$$x_i = \frac{1}{\det A} \cdot \det \begin{pmatrix} | & \cdots & | & | & | & \cdots & | \\ a_1 & \cdots & a_{i-1} & b & a_{i+1} & \cdots & a_d \\ | & \cdots & | & | & | & \cdots & | \end{pmatrix},$$

where A has the columns $a_1, \ldots, a_d \in \mathbb{R}^d$.

Proof. Let $i \in [d]$. Writing out the equation $A \cdot x = b$, we obtain $x_1 \cdot a_1 + \ldots + x_d \cdot a_d = b$, i.e.

$$x_1 \cdot a_1 + \ldots + x_{i-1} \cdot a_{i-1} + (x_i \cdot a_i - b) + x_{i+1} \cdot a_{i+1} + \ldots + x_d \cdot a_d = 0.$$

This shows that the d vectors $a_1, \ldots, a_{i-1}, x_i \cdot a_i - b, a_{i+1}, \ldots, a_d \in \mathbb{R}^d$ are linearly dependent. In particular, the $d \times d$ matrix containing these vectors as columns has zero determinant. By multilinearity of det, we obtain

$$\begin{aligned} 0 &= \det \begin{pmatrix} | & \cdots & | & | & | & \cdots & | \\ a_1 & \cdots & a_{i-1} & x_i \cdot a_i - b & a_{i+1} & \cdots & a_d \\ | & \cdots & | & | & | & \cdots & | \end{pmatrix} \\ &= x_i \cdot \det \begin{pmatrix} | & \cdots & | \\ a_1 & \cdots & a_d \\ | & \cdots & | \end{pmatrix} - \det \begin{pmatrix} | & \cdots & | & | & | & \cdots & | \\ a_1 & \cdots & a_{i-1} & b & a_{i+1} & \cdots & a_d \\ | & \cdots & | & | & | & \cdots & | \end{pmatrix}, \end{aligned}$$

which after solving for x_i yields the desired result. □

With Cramer's rule at hand, we get a useful expression, sometimes called *Jacobi's formula*, for the time derivative of the determinant map composed with a smooth curve in the space of invertible square matrices:

Theorem 4.69. *Let $I \subset \mathbb{R}$ open and $A : I \to \mathrm{GL}(d, \mathbb{R})$ a $\mathcal{C}^1$-curve. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t} \det(A(t)) = \det(A(t)) \cdot \mathrm{tr}\left(\dot{A}(t) \cdot (A(t))^{-1}\right), \tag{4.23}$$

where $\dot{A}(t) := \left(\dot{a}_{ij}(t)\right)_{ij}$ and tr denotes the trace operator.

Proof. Using multilinearity of the determinant map, we compute (cf. [3, p. 178])

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\det(A(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\det\begin{pmatrix} | & \cdots & | \\ a_1(t) & \cdots & a_d(t) \\ | & \cdots & | \end{pmatrix} \\
&= \sum_{i=1}^{d}\det\begin{pmatrix} | & \cdots & | & | & | & \cdots & | \\ a_1(t) & \cdots & a_{i-1}(t) & \dot{a}_i(t) & a_{i+1}(t) & \cdots & a_d(t) \\ | & \cdots & | & | & | & \cdots & | \end{pmatrix} \\
&= \det(A(t))\cdot\sum_{i=1}^{d}\frac{1}{\det(A(t))}\cdot\det\begin{pmatrix} | & \cdots & | & | & | & \cdots & | \\ a_1(t) & \cdots & a_{i-1}(t) & \dot{a}_i(t) & a_{i+1}(t) & \cdots & a_d(t) \\ | & \cdots & | & | & | & \cdots & | \end{pmatrix} \\
&= \det(A(t))\cdot\sum_{i=1}^{d}x_{i,i}(t).
\end{aligned}$$

In the last step, we used Cramer's rule (lemma 4.68), with $x_{i,i}(t)$ denoting the i-th component of the unique vector $x_i(t) \in \mathbb{R}^d$ which satisfies $A(t)\cdot x_i(t) = \dot{a}_i(t)$. Writing the d equations $A(t)\cdot x_i(t) = \dot{a}_i(t)$ as matrix equation, we get that $x_{i,i}(t)$ is the (i,i)-th component of the (by invertibility of $A(t)$ unique) solution $X(t)$ to the equation $A(t)\cdot X(t) = \dot{A}(t)$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\det(A(t)) = \det(A(t))\cdot\sum_{i=1}^{d}\left[(A(t))^{-1}\cdot\dot{A}(t)\right]_{i,i} = \det(A(t))\cdot\operatorname{tr}\left((A(t))^{-1}\cdot\dot{A}(t)\right).$$

The claim now follows from the invariance of tr under cyclic permutations. □

Next, we prove that the familiar Stone-Weierstraß theorem "preserves" the number of relevant variables in the following sense:

Lemma 4.70. *Let $C \subset \mathbb{R}^d$ be compact, $i \in [d]$ and $f : C \to \mathbb{R},\ x \mapsto f(x)$ a continuous map which does not depend on the i-th component x_i of $x \in C$. Let $\varepsilon > 0$. Then there is some polynomial $p : \mathbb{R}^d \to \mathbb{R}$ which also does not depend on x_i such that* $\sup_{x\in C}|f(x) - p(x)| < \varepsilon$.

Proof. If $C = \emptyset$, there is nothing to show. Let $z \in C$ and consider the map

$$\widehat{\pi}_i : \mathbb{R}^d \to \mathbb{R}^{d-1}, \qquad (x_1, \ldots, x_d) \mapsto (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d).$$

Then $\widehat{\pi}_i$ is obviously continuous, and therefore $\widehat{\pi}_i(C) \subset \mathbb{R}^{d-1}$ is compact. Define

$$\widehat{f}_i : \widehat{\pi}_i(C) \to \mathbb{R}, \qquad (y_1, \ldots, y_{d-1}) \mapsto f(y_1, \ldots, y_{i-1}, z_i, y_i, \ldots, y_{d-1}).$$

Note that by assumption, $\widehat{f}_i$ does not depend on the choice of $z \in C$. Then $\widehat{f}_i$ is clearly continuous, and by the Stone-Weierstrass theorem, cf. [2, p. 394 f.], we can find a polynomial $\widehat{p}_i : \mathbb{R}^{d-1} \to \mathbb{R}$ such that

$$\sup_{y\in\widehat{\pi}_i(C)}\left|\widehat{f}_i(y) - \widehat{p}_i(y)\right| < \varepsilon.$$

Let us finally set

$$p : \mathbb{R}^d \to \mathbb{R}, \qquad (x_1, \ldots, x_d) \mapsto \widehat{p}_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d).$$

Then p is obviously a polynomial which does not depend on x_i, and it is clear by construction that

$$\sup_{x\in C}\left|f(x) - p(x)\right| = \sup_{y\in\widehat{\pi}_i(C)}\left|\widehat{f}_i(y) - \widehat{p}_i(y)\right| < \varepsilon.$$ □

Let us now briefly digress on properties of the Coulomb potential with and without cut-off which are important in section 3.2. We want to prove that *away from the singularity*, the Coulomb force is Lipschitz continuous, and find good bounds on the local Lipschitz constants. By rotational symmetry of the Coulomb force, it is convenient to use $|\cdot| := |\cdot|_2$ with $|x|_2 := (x_1^2 + x_2^2 + x_3^2)^{\frac{1}{2}}$ the euclidean norm for once in the following two lemmata.

Lemma 4.71. *Let*

$$k : \mathbb{R}^3 \setminus \{0\} \to \mathbb{R}^3, \qquad q \mapsto \frac{q}{|q|^3}$$

the Coulomb force. Then there is some $C > 0$ such that for all $q, \tilde{q} \in \mathbb{R}^3$ with $q, \tilde{q} \neq 0$,

$$|k(q) - k(\tilde{q})| \leq \frac{C}{\big(\min\{|q|, |\tilde{q}|\}\big)^2 \cdot \max\{|q|, |\tilde{q}|\}} \cdot |q - \tilde{q}| \,.$$

Proof. For $i, j \in [3]$ and $q \neq 0$,

$$\partial_i k_j(q) = \partial_i \frac{q_j}{|q|^3} = \frac{\delta_{ij} \cdot |q|^3 - q_j \cdot 3\,|q|^2 \cdot \frac{q_i}{|q|}}{|q|^6} = \frac{\delta_{ij} \cdot |q|^2 - 3 q_i q_j}{|q|^5}.$$

Since $|ab| \leq \frac{1}{2}\left(a^2 + b^2\right)$ for all $a, b \in \mathbb{R}$, we obtain

$$|\partial_i k_j(q)| \leq \frac{|q|^2 + \frac{3}{2}|q|^2}{|q|^5} = \frac{5}{2\,|q|^3} \qquad \forall q \in \mathbb{R}^3 \setminus \{0\}.$$

Now, let $q, \tilde{q} \in \mathbb{R}^3 \setminus \{0\}$; w.l.o.g we may assume that $|q| \leq |\tilde{q}|$. Then for any two $\mathcal{C}^1$-curves $\gamma_{1,2} : [0,1] \to \mathbb{R}^3 \setminus \{0\}$ with $\gamma_1(0) = q$, $\gamma_1(1) = \gamma_2(0)$, $\gamma_2(1) = \tilde{q}$, by the fundamental theorem of calculus, we obtain

$$\begin{aligned}
\left|k_j(\tilde{q}) - k_j(q)\right| &\leq \left|k_j(\gamma_2(1)) - k_j(\gamma_2(0))\right| + \left|k_j(\gamma_1(1)) - k_j(\gamma_1(0))\right| \\
&= \left|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(k_j(\gamma_2(t))\right) \mathrm{d}t\right| + \left|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(k_j(\gamma_1(t))\right) \mathrm{d}t\right|.
\end{aligned}$$

Switching to appropriate spherical coordinates, for $r = |q|$, $R = |\tilde{q}|$ and some $\varphi \in [0, \pi]$ we may write $q = (r, 0, 0)$ and $\tilde{q} = (R\cos(\varphi), R\sin(\varphi), 0)$. Choose

$$\gamma_1(t) := \begin{pmatrix} r + t(R - r) \\ 0 \\ 0 \end{pmatrix}, \qquad \gamma_2(t) := \begin{pmatrix} R\cos(t\varphi) \\ R\sin(t\varphi) \\ 0 \end{pmatrix}.$$

Then obviously $\gamma_1(0) = q$, $\gamma_1(1) = \gamma_2(0) = (R, 0, 0)$, $\gamma_2(1) = \tilde{q}$ and $\gamma_{1,2} \in \mathcal{C}^1([0,1]; \mathbb{R}^3 \setminus \{0\})$. Moreover, by the chain rule and the Cauchy-Schwarz-inequality, we obtain

$$\begin{aligned}
\left|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\left(k_j(\gamma_1(t))\right) \mathrm{d}t\right| &= \left|\int_0^1 \nabla k_j(\gamma_1(t)) \cdot \dot{\gamma}_1(t)\,\mathrm{d}t\right| \leq \int_0^1 \left|\nabla k_j(\gamma_1(t))\right| \cdot |\dot{\gamma}_1(t)|\,\mathrm{d}t \\
&\leq \frac{5\sqrt{3}}{2} \int_0^1 \big(r + t(R - r)\big)^{-3} \cdot (R - r)\,\mathrm{d}t = \frac{5\sqrt{3}}{2}\left[-\frac{1}{2}\big(r + t(R - r)\big)^{-2}\right]_0^1 \\
&= \frac{5\sqrt{3}}{4}\left(\frac{1}{r^2} - \frac{1}{R^2}\right) = \frac{5\sqrt{3}}{4}\,\frac{R^2 - r^2}{r^2 R^2} = \frac{5\sqrt{3}}{4}\,\frac{(R - r)(R + r)}{r^2 R^2} \\
&\leq \frac{5\sqrt{3}}{2 r^2 R} \cdot (R - r)\,.
\end{aligned}$$

On the other hand,

$$\left|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\big(k_j(\gamma_2(t))\big)\,\mathrm{d}t\right| = \left|\int_0^1 \nabla k_j(\gamma_2(t))\cdot\dot\gamma_2(t)\,\mathrm{d}t\right| \le \int_0^1 \big|\nabla k_j(\gamma_2(t))\big|\cdot|\dot\gamma_2(t)|\,\mathrm{d}t$$
$$\le \frac{5\sqrt{3}}{2}\int_0^1 R^{-3}\cdot R\varphi\,\mathrm{d}t \le \frac{5\sqrt{3}}{2rR^2}\cdot r\varphi.$$

Observe that $|q-\tilde q|^2 = r^2+R^2-2rR\cos(\varphi)$, so using the trigonometric identity

$$\cos(\varphi) = \cos\left(\tfrac{\varphi}{2}+\tfrac{\varphi}{2}\right) = \cos^2\left(\tfrac{\varphi}{2}\right)-\sin^2\left(\tfrac{\varphi}{2}\right) = 1-2\sin^2\left(\tfrac{\varphi}{2}\right),$$

we see that

$$|q-\tilde q|^2 = r^2+R^2-2rR\big(1-2\sin^2\left(\tfrac{\varphi}{2}\right)\big) = (R-r)^2+4rR\sin^2\big(\tfrac{\varphi}{2}\big).$$

In particular, $R-r\le|q-\tilde q|$ and

$$r\varphi = \pi\sqrt{r^2}\cdot\tfrac{2}{\pi}\tfrac{\varphi}{2} \le \pi\sqrt{rR}\cdot\sin\big(\tfrac{\varphi}{2}\big) = \tfrac{\pi}{2}\sqrt{4rR\sin^2\big(\tfrac{\varphi}{2}\big)} \le \tfrac{\pi}{2}\,|q-\tilde q|\,,$$

where we employed the elementary inequality $\frac{2}{\pi}\phi\le\sin\phi$ for $\phi\in[0,\frac{\pi}{2}]$, which can be proved by concavity of sin on $[0,\frac{\pi}{2}]$. The cases $\varphi=0$ and $r=R$, $\varphi=\pi$ show that actually, both inequalities are sharp. Putting everything together, we finally arrive at

$$|k(q)-k(\tilde q)| \le \sqrt{3}\cdot\max\{|k_j(q)-k_j(\tilde q)| : i\in[3]\} \le \sqrt{3}\cdot\left(\frac{5\sqrt{3}}{2\,|q|^2\,|\tilde q|}\cdot|q-\tilde q| + \frac{5\sqrt{3}\pi}{4\,|q|\,|\tilde q|^2}\cdot|q-\tilde q|\right)$$
$$\le \frac{C}{|q|^2\,|\tilde q|}\cdot|q-\tilde q|$$

where $C = 3(\frac{5}{2}+\frac{5\pi}{4})$. □

Remark 4.72. In the previous proof, we used that (with $|\cdot|\equiv|\cdot|_2$)) $|\nabla k_j(q)|\le\frac{5\sqrt{3}}{2}|q|^{-3}$. With a slightly more involved calculation, one can show that actually, $|\nabla k_j(q)|\le 2\,|q|^{-3}$ and hence further improve C.

Note that in the proof we just saw, the particular form of ∇k_j was not important, the only thing we needed was that $|\nabla k_j(q)|\le C\,|q|^{-3}$ for all $q\in\mathbb{R}^3\setminus\{0\}$. Hence, we can do a similar computation for the Coulomb force with cut-off. Of course, since the (almost everywhere defined) gradient of the Coulomb force with cut-off f defined in (3.2) is bounded by our above estimate for the gradient of the Coulomb force k (recall that $|\nabla f(q)| = S^{3\delta}\le|q|^{-3}$ on $B_{S^{-\delta}(0)}\setminus\{0\}$), we can always use the estimate from the previous lemma. Better estimates can only be achieved if we have good control on the distance of q and $\tilde q$, since then, knowing that, say, q is near the cut-off region, we can "localize" the paths γ_1, γ_2 and see that a large part of them is contained in the cut-off region, which leads to a "cut-off local Lipschitz bound":

Lemma 4.73. *Let $f:\mathbb{R}^3\to\mathbb{R}^3$ be the Coulomb force with cut-off defined in* (3.2). *Then there is some $g:\mathbb{R}^3\to\mathbb{R}$ satisfying a S^3_δ-condition such that for all $q\in\mathbb{R}^3$ and all $\xi\in\mathbb{R}^3$ with $|\xi|\le\frac{1}{2}S^{-\delta}$,*

$$|f(q+\xi)-f(q)|\le g(q)\cdot|\xi|\,. \tag{$*$}$$

Proof. We distinguish between two cases:

(i) Let $|q| \leq S^{-\delta}$. If $|q+\xi| \geq |q|$, by using the same paths as in the proof of lemma 4.71, if necessary with an additional splitting when moving radially outwards and crossing the sphere of radius $S^{-\delta}$, and using the rough estimate $|\nabla f(q)| \leq CS^{3\delta}$ for all $q \in B_{2S^{-\delta}}(0)$ with $q \neq 0$ and $|q| \neq S^{-\delta}$, we obtain that in this case,

$$|f(q+\xi) - f(q)| \leq CS^{3\delta} \cdot |\xi| .$$

For $|q+\xi| < |q|$, we can interchange the roles of q and $q+\xi$ and see that the statement remains true.

(ii) If $|q| > S^{-\delta}$, then $|q+\xi| \geq |q| - |\xi| \geq \frac{1}{2}|q|$, and using that the gradient of the Coulomb force with cut-off is, where existent, bounded by the gradient of the Coulomb force without cut-off, by virtually the same computation as in lemma 4.71, again possibly with a splitting of the radial path at the radius $S^{-\delta}$, we obtain

$$|f(q+\xi) - f(q)| \leq \frac{C}{(\frac{1}{2}|q|)^3} \cdot |\xi| \leq \frac{8C}{|q|^3} \cdot |\xi| .$$

This shows that for

$$g(q) := 8C \cdot \begin{cases} S^{3\delta}, & |q| \leq S^{-\delta} \\ |q|^{-3}, & |q| > S^{-\delta}, \end{cases} \tag{4.24}$$

$(*)$ holds, and obviously g satisfies a S^3_δ-condition. □

Next, we want to derive a special form of the mean value theorem of differentiation in multi-variable calculus:

Lemma 4.74. *Let $a, b \in \mathbb{R}$ with $a < b$, $f \in \mathcal{C}^2([a,b])$ and $x \in [a,b]$. Then for all $h > 0$ such that $[x, x+h] \subset [a,b]$, there is some $\theta \in (0,1)$ such that*

$$\frac{1}{h}\big(f(x+h) - f(x)\big) - f'(x) = \frac{h}{2} f''(x+\theta h).$$

Proof. We use Taylor's theorem with the Lagrangian form of the remainder (see e.g. [2, p. 341]): for some $\xi \in (x, x+h)$,

$$\frac{1}{h}\big(f(x+h) - f(x)\big) - f'(x) = \frac{1}{h}\big(f(x) + hf'(x) + \frac{h^2}{2} f''(\xi) - f(x)\big) - f'(x) = \frac{h}{2} f''(\xi).$$

Letting $\theta := \frac{\xi - x}{h} \in (0,1)$ yields the desired result. □

Corollary 4.75. *Let $d \in \mathbb{N}$, $U \subset \mathbb{R}^d$ open and $f \in \mathcal{C}^2(U;\mathbb{R})$. Then for $i \in [d]$, $x, y \in U$ such that $\{x + ty : t \in [0,1]\} \subset U$, there is some $\theta \in (0,1)$ such that*

$$\frac{1}{h}\big(f(x+hy) - f(x)\big) - \nabla f(x) \cdot y = \frac{h}{2} \sum_{i,j=1}^{d} \partial_i \partial_j f(x+\theta y) \cdot y_i y_j.$$

Proof. Apply lemma 4.74 to the map $f \circ \gamma$, where $\gamma : [0,1] \to \mathbb{R}^d,\ t \mapsto x + ty$, and use the chain rule. □

Morover, we need some results on a few functions which are relevant in the main text.

Lemma 4.76. *For $C, \varepsilon > 0$, consider the maps $u, v, w : [1, \infty) \to \mathbb{R}$ where*

$$u(s) := \frac{\sqrt{\ln(s)}}{s^\varepsilon}, \qquad v(s) := \frac{1 + \ln(s)}{s^\varepsilon}, \qquad w(s) := \frac{e^{C\sqrt{\ln(s)}}}{s^\varepsilon}.$$

Then u, v, w are bounded.

Proof. Clearly, boundedness of v implies boundedness of u. By L'Hospital's rule,

$$\lim_{s\to\infty} v(s) = \lim_{s\to\infty} \frac{s^{-1}}{\varepsilon \cdot s^{\varepsilon-1}} = \varepsilon^{-1} \cdot \lim_{s\to\infty} s^{-\varepsilon} = 0.$$

Since v is non-negative and continuous, boundedness follows immediately.

Observe that for all $s \geq 1$,

$$v(s) = e^{C\sqrt{\ln(s)}} = e^{\ln(s)\cdot \frac{C}{\sqrt{\ln(s)}}} = s^{\frac{C}{\sqrt{\ln(s)}}},$$

so using $\lim_{s\to\infty} \frac{C}{\sqrt{\ln(s)}} - \varepsilon = -\varepsilon < 0$ we obtain that $\frac{C}{\sqrt{\ln(s)}} - \varepsilon < -\frac{\varepsilon}{2}$ for s big enough, and hence

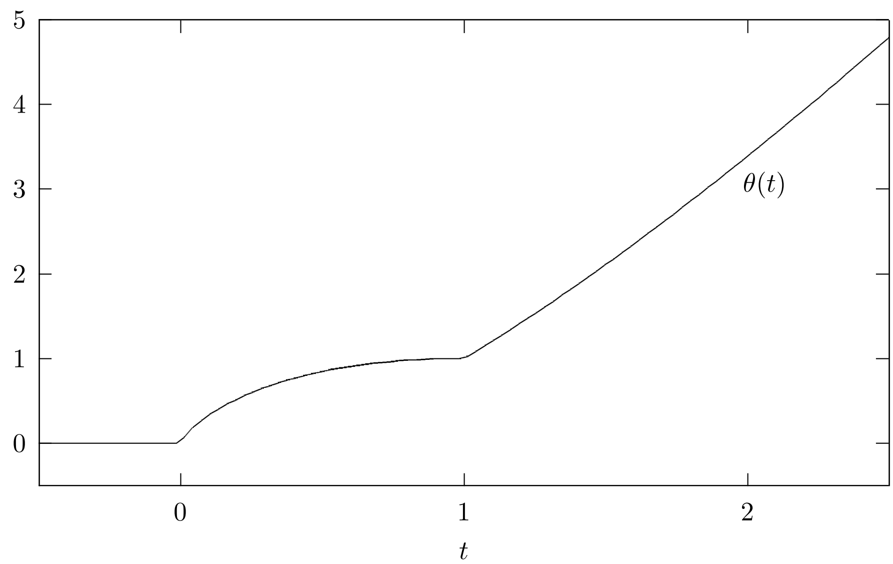
$$\lim_{s\to\infty} v(s) = \lim_{s\to\infty} \frac{e^{C\sqrt{\ln(s)}}}{s^\varepsilon} = \lim_{s\to\infty} s^{\frac{C}{\sqrt{\ln(s)}}-\varepsilon} = 0.$$

With the same arguments which were applied to v, we see that w is bounded as well. $\square$

Lemma 4.77. *Consider the map*

$$\theta : \mathbb{R} \to \mathbb{R}, \qquad t \mapsto \begin{cases} 0, & t \leq 0 \\ t \cdot (1 + |\ln(t)|), & t > 0. \end{cases}$$

Then $\theta \in \mathcal{C}(\mathbb{R})$, $\theta(t) > 0$ for $t > 0$, θ is monotonously increasing and $\lim_{\varepsilon \searrow 0} \int_\varepsilon^1 \frac{1}{\theta(t)}\, \mathrm{d}t = \infty$.



Proof. Note that for $t \in (0,1)$, $|\ln(t)| = -\ln(t)$. By L'Hospital's rule,

$$\lim_{t \searrow 0} t \cdot |\ln(t)| = \lim_{t \searrow 0} \frac{-\ln(t)}{\frac{1}{t}} = \lim_{t \searrow 0} \frac{-\frac{1}{t}}{-\frac{1}{t^2}} = \lim_{t \searrow 0} t = 0.$$

Hence, θ is continuous in 0. Continuity in all other points is clear. That $\theta(t) > 0$ for $t > 0$ is obvious. For $t < 0$, $\theta(t)$ is constant and therefore its derivative equals 0. For $t \in (0,1)$ resp. $t \in (1,\infty)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta(t) = \frac{\mathrm{d}}{\mathrm{d}t}\big(t \cdot (1 - \ln(t)\big) = -\ln(t) > 0 \quad \text{resp.} \quad \frac{\mathrm{d}}{\mathrm{d}t}\theta(t) = \frac{\mathrm{d}}{\mathrm{d}t}\big(t \cdot (1 + \ln(t)\big) = 2 + \ln(t) > 0,$$

which by continuity shows that θ is non-decreasing.

Now, consider the map

$$h : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \begin{cases} 0, & t \leq 0 \\ t + t \cdot \ln(t), & t > 0 \end{cases}$$

By the same arguments as above, h is continuous with derivative $h'(t) = 2 + \ln(t)$ for $t > 0$, so we see that $h'(t) < 0$ on $(0, e^{-2})$, and since $h(0) = 0$ we obtain $h(t) < 0$ for $t \in (0, e^{-2})$. Consequently, $t \leq -t \cdot \ln(t)$ on this interval, and thus $t \cdot (1 - \ln(t)) = t - t \cdot \ln(t) \leq -2t \cdot \ln(t)$ on $(0, e^{-2})$. It follows that for $\varepsilon \in (0,1)$,

$$\begin{aligned} \int_\varepsilon^1 \frac{1}{\theta(t)}\,\mathrm{d}t &= \int_\varepsilon^1 \frac{1}{t \cdot (1 + |\ln(t)|)}\,\mathrm{d}t = \int_\varepsilon^1 \frac{1}{t \cdot (1 - \ln(t))}\,\mathrm{d}t \geq \int_\varepsilon^{e^{-2}} \frac{1}{t - t \cdot \ln(t))}\,\mathrm{d}t \\ &\geq \int_\varepsilon^{e^{-2}} \frac{1}{2t \cdot (-\ln(t))}\,\mathrm{d}t = \frac{1}{2}\big[-\ln(-\ln(t))\big]_\varepsilon^{e^{-2}} \\ &= \frac{1}{2}\big(\ln(-\ln(\varepsilon)) - \ln(2)\big) \xrightarrow{\varepsilon \searrow 0} +\infty. \end{aligned}$$

□

Lemma 4.78. *For $C > 0$ and $t_0, x_0 \in \mathbb{R}$, consider the IVP*

$$\dot{x}(t) = C \cdot \theta(x(t)), \qquad x(t_0) = x_0, \tag{4.25}$$

where again $\theta : \mathbb{R} \to \mathbb{R}$ is defined in lemma 4.77. Then (4.25) *has a unique global solution, which is for $t \geq t_0$ given by*

$$x(t) = \begin{cases} x_0, & x_0 \leq 0, \\ e^{1 - e^{-C(t-t_0) + \ln(1 - \ln(x_0))}}, & x_0 \in (0,1),\ t \leq t_0 + \frac{1}{C} \cdot \ln(1 - \ln(x_0)), \\ e^{e^{C(t-t_0) - \ln(1 - \ln(x_0))} - 1}, & x_0 \in (0,1),\ t \geq t_0 + \frac{1}{C} \cdot \ln(1 - \ln(x_0)), \\ e^{e^{C(t-t_0) + \ln(1 + \ln(x_0))} - 1}, & x_0 \geq 1. \end{cases} \tag{4.26}$$

In particular, for all $t \geq t_0$, $\lim_{x_0 \searrow 0} x(t) = 0$.

Proof. We only sketch the proof because most of the calculations are straightforward. First, observe that since $C \cdot \theta \geq 0$, every solution of (4.25) is non-decreasing. Next, it is obvious that θ is locally Lipschitz-continuous on $\mathbb{R} \setminus \{0\}$, so from the Picard-Lindelöf theorem (see e.g. [40, p. 68]) we deduce that the only possibility for non-uniqueness to appear is at $x(t) = 0$. However, for

$0 \le x < y \le 1$, one has by concavity of the map $C \cdot \theta$ on $(0,1)$ ($\theta''(t) = -\frac{1}{t} < 0$ for $t \in (0,1)$) that

$$\begin{aligned}|G(y) - G(x)| = |C \cdot \theta(y) - C \cdot \theta(x)| &= C \cdot \frac{\theta(y) - \theta(x)}{y - x} \cdot (y - x) \\ &\le C \cdot \frac{\theta(y-x) - \theta(0)}{y - x} \cdot (y - x) = C \cdot \theta(y - x).\end{aligned}$$

This argument was inspired by a similar computation in [1]. With lemma 4.77, it follows that G satisfies all the presuppositions of Osgood's criterion (see e.g. [40, p. 146-147]), i.e. we obtain uniqueness of solutions. For $x_0 \le 0$, $\theta(x_0) = 0$, so x_0 is an equilibrium point and therefore $x(t) = x_0$ the unique solution. For $x_0 \in (0,1)$, we restrict θ to $(0,1)$ and use the separation of variables formula, noting that an indefinite integral of $\frac{1}{x(1-\ln(x))}$ is given by $-\ln(1 - \ln(x))$. One then checks that for $t \nearrow \tilde{t} := t_0 + \frac{1}{C} \cdot \ln(1 - \ln(x_0))$, $x(t) \nearrow 1$. Hence, for $t \ge \tilde{t}$, we continue the solution with a solution of the IVP for θ restricted to $[1, \infty)$. Since an indefinite integral of $\frac{1}{x(1+\ln(x))}$ is given by $\ln(1 + \ln(x))$, the rest of (4.26) follows from easy computations. Finally, note that as $x_0 \searrow 0$, $\tilde{t} \nearrow \infty$ and $\ln(x_0) \searrow -\infty$, and consequently $x(t) \searrow 0$ for fixed $t \ge t_0$ because we can use the second row of (4.26). □

Next, we want to investigate some sort of *stability* of the IVP in the previous lemma.

Lemma 4.79. *Consider the IVP*

$$\dot{x} = C \cdot \theta(x(t)) + \frac{1}{n}, \qquad x(t_0) = 0. \tag{4.27}$$

Then for every $n \in \mathbb{N}$, there is a unique, global solution x_n to (4.27), and $x_n \to 0$ uniformly on compact sets as $n \to \infty$, i.e. for all $T > t_0$, $\sup\{|x_n(s)| : s \in [t_0, T]\} \xrightarrow{n\to\infty} 0$.

Proof. Let $G_n : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, $(x,t) \mapsto C \cdot \theta(x) + \frac{1}{n}$, $G : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, $(x,t) \mapsto C \cdot \theta(x)$. From lemma 4.78, we know that the unique (and therefore maximal) solution of the IVP $\dot{x}(t) = G(x(t), t)$, $x(t_0) = 0$ is given by $x(t) = 0$. Moreover, for $x, y \in (0,1)$, $|G_n(y) - G_n(x)| = |G(y) - G(x)|$, for the same reasons as in lemma 4.78, solutions to the IVP (4.27) are unique (however, we could also work with the maximal solutions without difficulties here). Clearly, solutions to (4.27) are obviously non-decreasing, and since $\frac{1}{n} \le 1 \le \theta(x)$ for $x \ge 1$, they are easily seen to exist globally: existence on $(-\infty, 1]$ is clear because $C \cdot \theta|_{(-\infty,1]}$ is bounded, and for existence on $[1, \infty)$ use that $C \cdot \theta + \frac{1}{n} \le (C+1) \cdot \theta$ on $[1, \infty)$, the comparison theorem 4.5 and lemma 4.78.

Obviously, G_n converges to G uniformly, so by theorem 3.2 in [20, p. 14f.], there is a subsequence $(n_k)_{n \in \mathbb{N}}$ such that $x_{n_k}(t) \to \overline{x}(t)$ uniformly on $[t_0, T]$ as $k \to \infty$, where $\overline{x}(t)$ is a solution of the IVP $\dot{x}(t) = G(x(t), t)$, $x(t_0) = 0$, i.e. by uniqueness $x_{n_k}(t) \to \overline{x}(t) = x(t) = 0$ uniformly in $t \in [t_0, T]$ for any $T > t_0$. But since the r.h.s of the IVP (4.27) decreases as n increases, again using the comparison theorem, we see that the whole sequence $(x_n(t))_{n \in \mathbb{N}}$ is actually non-increasing, and thus we obtain $x_n(t) \to 0$ uniformly on $[t_0, T]$, as claimed. □

Finally, we shall state a very general version of Grønwall's lemma in integral form, which complements its differential counterpart formulated in theorem 4.3.

Theorem 4.80 (Grønwall's lemma, integral form)**.** *Let $t_0, T \in \mathbb{R}$ with $t_0 < T$ and suppose that $f : [t_0, T] \to \mathbb{R}$ and $g, h : [t_0, T] \to [0, \infty)$ are measurable such that g, $f \cdot g$, $h \cdot g \in L^1([t_0, T])$. If additionally the inequality*

$$f(t) \le h(t) + \int_{t_0}^{t} g(s) \cdot f(s) \,\mathrm{d}s \qquad \forall t \in [t_0, T]$$

is satisfied, then it holds that

$$f(t) \le h(t) + \int_{t_0}^{t} g(s) \cdot h(s) \cdot \exp\left(\int_{s}^{t} g(\tau) \,\mathrm{d}\tau\right) \mathrm{d}s \qquad \forall t \in [t_0, T].$$

In particular, for h non-decreasing and g continuous,

$$f(t) \le h(t) \cdot \exp\left(\int_{t_0}^{t} g(s) \,\mathrm{d}s\right) \qquad \forall t \in [t_0, T].$$

Remark 4.81. In case $f, g, h : [t_0, T] \to \mathbb{R}$ are measurable and bounded, the hypotheses of the first part of the theorem are clearly satisfied.

Proof of theorem 4.80. For the first claim, see [23], theorem A, with $g \equiv 1$ in the notation of the book and $d\alpha = g \,\mathrm{d}x$ where now g denotes the function in theorem 4.80. Note that the theorem is a reformulation of lemma 4 in [25], however, the form in [23] fits better in the context of this thesis. However, the paragraph about the integral form for locally finite measures in the Wikipedia entry about Grønwall's inequality is the more readable reference, and it provides a nice sketch of the proof as well. For the second claim, using the fundamental theorem of calculus (and non-negativity of g, h), we obtain that for all $t \in [t_0, T]$,

$$\begin{aligned} f(t) &\le h(t) + h(t) \cdot \int_{t_0}^{t} g(s) \cdot \exp\left(\int_{s}^{t} g(\tau) \,\mathrm{d}\tau\right) = h(t) + h(t) \cdot \int_{t_0}^{t} -\frac{\mathrm{d}}{\mathrm{d}s} \exp\left(\int_{s}^{t} g(\tau) \,\mathrm{d}\tau\right) \mathrm{d}s \\ &= h(t) - h(t) \cdot \left[1 - \exp\left(\int_{t_0}^{t} g(s) \,\mathrm{d}s\right)\right] = h(t) \cdot \exp\left(\int_{t_0}^{t} g(s) \,\mathrm{d}s\right). \end{aligned}$$

$\square$

Bibliography

[1] Agarwal, R. P. and Lakshmikantham, V. (1993). *Uniqueness and nonuniqueness criteria for ordinary differential equations*, volume 6. World Scientific Publishing Company.

[2] Amann, H. and Escher, J. (2005). *Analysis I.* Birkhäuser Basel.

[3] Amann, H. and Escher, J. (2008). *Analysis II.* Birkhäuser Basel.

[4] Amann, H. and Escher, J. (2009). *Analysis III.* Birkhäuser Basel.

[5] Arnold, V. (1989). *Mathematical Methods of Classical Mechanics.* Springer-Verlag New York, second edition.

[6] Batt, J. (1962). Fixpunktprobleme bei partiellen Differentialgleichungen im Zusammenhang mit dem statistischen Anfangswertproblem der Stellardynamik. *Dissertation Technische Hochschule Aachen.*

[7] Batt, J. (1977). Global symmetric solutions of the initial value problem of stellar dynamics. *Journal of Differential Equations*, 25(3):342–364.

[8] Bogachev, V. I. (2007a). *Measure Theory I.* Springer-Verlag Berlin Heidelberg.

[9] Bogachev, V. I. (2007b). *Measure Theory II.* Springer-Verlag Berlin Heidelberg.

[10] Braun, W. and Hepp, K. (1977). The Vlasov dynamics and its fluctuations in the $1/n$ limit of interacting classical particles. *Comm. Math. Phys.*, 56(2):101–113.

[11] Cerdà, J. (2010). *Linear Functional Analysis*, volume 116 of *Graduate Studies in Mathematics.* American Mathematical Society.

[12] Cohn, D. (2013). *Measure Theory.* Birkhäuser Basel, second edition.

[13] Dudley, R. (2002). *Real Analysis and Probability.* Cambridge University Press.

[14] Evans, L. C. (2010). *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics.* American Mathematical Society, second edition.

[15] Evans, L. C. and Gariepy, R. F. (2015). *Measure Theory and Fine Properties of Functions.* Textbooks in Mathematics. Chapman and Hall/CRC.

[16] Fecko, M. (2006). *Differential Geometry and Lie Groups for Physicists.* Cambridge University Press.

[17] Fournier, N. and Guillin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3–4):707–738.

[18] Günther, N. M. (1957). *Die Potentialtheorie und ihre Anwendung auf Grundaufgaben der mathematischen Physik.* B.G. Teubner Verlagsgesellschaft Leipzig.

[19] Gut, A. (2005). *Probability: A Graduate Course.* Springer Texts in Statistics. Springer-Verlag New York.

[20] Hartman, P. (2002). *Ordinary Differential Equations.* SIAM, second edition.

[21] Horst, E. and Neunzert, H. (1981). On the classical solutions of the initial value problem for the unmodified non-linear Vlasov equation I: general theory. *Mathematical Methods in the Applied Sciences*, 3(1):229–248.

[22] Horst, E. and Neunzert, H. (1982). On the classical solutions of the initial value problem for the unmodified non-linear Vlasov equation II: special cases. *Mathematical Methods in the Applied Sciences*, 4(1):19–32.

[23] Horváth, L. (1992). Gronwall-Bellman type integral inequalities in measure spaces. *Journal of Mathematical Analysis and Applications*, 202(1):183–193.

[24] Illner, R. and Neunzert, H. (1979). An existence theorem for the unmodified Vlasov equation. *Mathematical Methods in the Applied Sciences*, 1(4):530–544.

[25] Jones, G. S. (1964). Fundamental inequalities for discrete and discontinuous functional equations. *J. Soc. Indust. Appl. Math.*, 12(1):43–57.

[26] Klenke, A. (2014). *Probability Theory.* Universitext. Springer-Verlag London, second edition.

[27] Knight, W. J. (1980). Functions with zero right derivatives are constant. *The American Mathematical Monthly*, 87(8):657–658.

[28] Lazarovici, D. and Pickl, P. (2017). A mean field limit for the Vlasov-Poisson system. *Archive for Rational Mechanics and Analysis*, 225(3):1201–1231.

[29] Lieb, E. H. and Loss, M. (2001). *Analysis.* Graduate Studies in Mathematics. American Mathematical Society, second edition.

[30] Lions, P.-L. and Perthame, B. (1991). Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Inventiones mathematicae*, 105(1):415–430.

[31] Loeper, G. (2006). Uniqueness of the solution to the Vlasov-Poisson system with bounded density. *Journal de mathématiques pures et appliquées*, 86(1):68–79.

[32] Pfaffelmoser, K. (1992). Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial data. *Journal of Differential Equations*, 95(2):281–303.

[33] Rein, G. (2007). Collisionless kinetic equations from astrophysics - the Vlasov-Poisson system. *Handbook of differential equations: evolutionary equations*, 3:383–476.

[34] Schaeffer, D. and Cain, J. W. (2016). *Ordinary Differential Equations: Basics and Beyond.* Springer-Verlag New York.

[35] Schaeffer, J. (1991). Global existence of smooth solutions to the Vlasov Poisson system in three dimensions. *Communications in partial differential equations*, 16(8–9):1313–1335.

[36] Schaub, M. (2017). A second order ground state energy expansion for the dilute bose gas. *Master's Thesis Ludwig-Maximilians-Universität München.*

[37] Spohn, H. (1991). *Large Scale Dynamics of Interacting Particles.* Springer-Verlag Berlin Heidelberg.

[38] Terrell, W. J. (2009). *Stability and stabilization: an introduction*. Princeton University Press.

[39] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag New York.

[40] Walter, W. (1998). *Ordinary Differential Equations*. Graduate Texts in Mathematics. Springer-Verlag New York.

Declaration of authorship

I declare that the work presented here is, to my best knowledge and belief, original and the result of my own investigations, except as acknowledged. It has not been submitted, neither in whole nor partly, for a degree at this or any other university. Formulations and ideas taken from other sources are cited accordingly[1].

Munich, June 22, 2018

Johannes Sedlmeir

[1]this formulation was taken from [36]