TECHNISCHE UNIVERSITÄT WIEN

DISSERTATION

Rigorous derivations of diffusion systems from moderately interacting particle models

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften unter der Leitung von

Univ. Prof. Dr. Ansgar Jüngel
E101 – Institut für Analysis und Scientific Computing, TU Wien

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Dipl.-Ing. Alexandra Holzinger, BSc
Matrikelnummer: 09871354

Diese Dissertation haben begutachtet:

1. **Prof. Dr. Ansgar Jüngel**
   Institut für Analysis und Scientific Computing, TU Wien
2. **Prof. Dr. Li Chen**
   Institut für Mathematik, Universität Mannheim
3. **Prof. Dr. Christian Kühn**
   Fakultät für Mathematik, Technische Universität München (TUM)

Wien, am 8. Mai 2023

Kurzfassung

Diese Doktorarbeit beschäftigt sich, ausgehend von stochastischen Interaktionsmodellen, mit der rigorosen Herleitung spezieller nicht-linearer partieller Differentialgleichungen.
Die verwendeten Methoden stützen sich auf das mathematische Konzept der sogenannten ‚mean-field limits', ein Konzept, das sich nicht nur auf dem Gebiet der reinen Mathematik, sondern auch in der fächerübergreifenden Forschung in den Bereichen Populationsdynamik, Physik, Neurowissenschaften, Deep Learning und Wirtschaftsforschung steigender Beliebtheit erfreut.
Die Grundidee dieser speziellen Partikel-Grenzwerte liegt darin, dass unter bestimmten Voraussetzungen das Partikelsystem trotz Interaktionen zwischen den Partikeln im Grenzwert (Anzahl der Partikel strebt gegen Unendlich) durch eine Dichtefunktion approximiert werden kann, die wiederum als Lösung einer partiellen Differentialgleichung aufgefasst werden kann. Diese Eigenschaft wird in der Fachliteratur auch ‚propagation of chaos' genannt.
In der vorliegenden Arbeit werden nur sogenannte „diffusive Partikelsysteme" betrachtet, welche im Grenzwert zu partiellen Differentialgleichungen mit positiver Diffusionskonstante führen. Speziell werden in dieser Doktorarbeit Interaktionssysteme betrachtet (auch „moderate Interaktionssysteme" genannt), bei denen der Interaktionskern mit der Anzahl der Partikel skaliert. Im Gegensatz zu klassischen Mean-field-Modellen (auch „schwache Mean-field-Modelle" genannt), führt das moderate Regime zu lokalen partiellen Differentialgleichungen.
Die Arbeit gliedert sich in drei Teile: Im ersten Teil der Doktorarbeit wird - ausgehend von einem moderaten stochastischen Teilchenmodell - eine verallgemeinerte Version des sogenannten SKT-Systems hergeleitet, welches ein Mehr-Spezies-Modell mit Kreuzdiffusionsstruktur in der Populationsdynamik darstellt. Ebenso enthält auch der folgende zweite Teil der Arbeit eine rigorose Herleitung einer fraktionellen Poröse-Mediums-Gleichung mit moderat interagierenden Partikeln. Aufgrund der verwendeten Techniken im moderaten Regime enthalten diese beiden ersten Teile der Arbeit auch Abschätzungen von nicht-lokalen Approximationsmodellen der eben genannten lokalen partiellen Differentialgleichungen. Der dritte Teil der vorliegenden Doktorarbeit enthält eine neue mathematische Technik, um - ausgehend von diffusiven Partikeln unter dem Einfluss von Aggregation - ein bedingtes Konvergenzresultat in L^2-Norm herzuleiten. Dieses Resultat kann als erster Schritt zu einem Fluktuations-Resultat im Kontext von aggregierenden Mean-field-Partikelsystemen gesehen werden.

Abstract

This thesis is concerned with the derivation of certain types of nonlinear partial differential equations from stochastic interacting particle systems. The underlying methods are within the framework of mean-field limits, a well-known mathematical concept which has become an emerging tool of interdisciplinary research due to the increasing number of applications in population dynamics, physics, neuroscience, deep learning and others.

The basic idea of these types of particle limits is to show that even though the particles are interacting with each other – under certain conditions – in the large particle limit, the system can be approximated by a density function which solves a partial differential equation: This is also called 'propagation of chaos'. Throughout this thesis, the case of diffusive particle systems is considered leading to partial differential equations with positive diffusion parameters. Special focus in this work is put on moderately interacting particle systems, a technique where the interaction kernel of the particle system scales with the number of particles. In contrast to the classical mean-field limit, which is also called weak mean-field limit, the moderate regime leads to local partial differential equations.

The thesis is split into three parts: In the first part, a rigorous derivation of a generalised version of the so-called SKT-system – a multi-species model from population dynamics – from moderately interacting particles is shown. In the second part, the method of moderately interacting particles is used to derive a porous media equation with fractional diffusion. Due to technical issues which occur in the moderate regime, rigorous estimates of non-local approximations of the particular partial differential equations are shown in those two chapters, as well. The third part of this work shows a new technique for proving a conditional quantitative L^2-convergence result for diffusive particles under the effect of aggregation, which can be seen as a step towards the proof of fluctuations around the mean-field limit in the setting of aggregating particles.

I can't carry it for you, but I can carry you[1]

— J.R.R. Tolkien

Danksagung

Ich bedanke mich an dieser Stelle besonders bei Prof. Ansgar Jüngel für die treffende und interessante Themenstellung und seine fachliche Unterstützung während meines Doktorats.

Weiters bedanke ich mich bei Prof. Jüngel und der Technischen Universität Wien, dass für die Finanzierung meiner Doktoratsstelle die entsprechenden Mittel bereitgestellt wurden. Außerdem durfte ich Teil der vom österreichischen Wissenschaftsfonds (FWF) finanzierten Doktoratsschule *Dissipation and Dispersion in Nonlinear PDEs* (grant W1245) sein, die mir einige Konferenzteilnahmen und längere Forschungsaufenthalte in Deutschland möglich machte. Besonderer Dank gilt dabei Prof. Li Chen, die mich mehrmals an der Universität Mannheim willkommen geheißen hat. Zusätzlich bedanke ich mich bei Prof. Christian Kühn für sein Interesse an meiner Arbeit und für seine Bereitschaft, diese auch zu begutachten.

Neben der fachlichen und finanziellen Unterstützung hat mich vor allem die mentale Unterstützung meines Freundeskreises in und außerhalb der Universität unglaublich getragen – euer Anteil an dem Gelingen dieses Projektes ist nicht hoch genug zu schätzen! Am Anfang meines Doktorats wurde mir gesagt „*Ein Doktorat ist kein Sprint, sondern ein Marathon*" und in diesem Sinne bedanke ich mich bei allen Menschen, die mich auf diesem Weg begleitet haben: Manche sind die gesamten Strecke mit mir gelaufen, andere die ersten Kilometer und ein paar von euch habe ich erst auf den letzten Metern kenngelernt. All eure Unterstützung ist hoch geschätzt!

Schlussendlich möchte ich meiner Familie für ihre mentale und finanzielle Unterstützung während meiner gesamten Ausbildungsphase danken.

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, ERC Advanced Grant no. 101018153.

[1]In: *The Lord of The Rings: The Return of the King.*

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 8. Mai 2023

Alexandra Holzinger

Contents

Das gute Gelingen ist zwar nichts Kleines, fängt aber mit Kleinigkeiten an.

— Sokrates[1]

1 Introduction

Contents

It is one of the fundamental aims in science to understand the effect of *particles* or *parts* merging together to one quantity, like birds forming a swarm or gas particles forming an entity. Aristotele once said *"the whole is greater than the sum of its parts"*[2], which raises questions like

- *which properties of the individual parts are inherited by the whole* and
- *which properties can be only observed if we look at the whole quantity.*

In applied mathematics, we speak of *microscopic levels* if we are talking about the individual parts (that later form one quantity) and *macroscopic levels* if we are looking at the whole quantity.

Usually, in applications coming from natural sciences and economy – this thesis will only consider such applications – the macroscopic level can be described by a *density function*, which, roughly speaking, indicates in which areas there are more particles and in which areas there are less. Since time is an essential part of most processes arising in science, it is crucial for understanding the macroscopic level to (mathematically) describe its evolution in time, i.e. how it changes in time – for instance *where is the swarm of birds moving to* or *will the gas be equally distributed in a room after we wait a certain amount of time.*

In fact, amongst others we are interested whether we can describe the time evolution of the density function of the particles by a *partial differential equation* which then allows us to simulate and study properties of the macroscopic level. In the words of the famous physicist Paul Dirac (1902-1984), who said *"I consider that I understand an equation when I can predict the properties of its solutions, without actually solving it"*[3], although in most cases we lack a concrete formula of the solution, we can describe it by its properties. Additionally, in modern times numerical simulations of solutions of partial differential equations have become an essential part of applied mathematical research. This shows an advantage of studying – from a mathematical but also from an applied point of view – the connection between microscopic (particle) levels and its corresponding macroscopic equation since particle systems arising from physics often consider a large number of particles ($N \sim 10^{20}$).

[1] Attributed to Sokrates

[2] [Das Ganze ist mehr als die Summe seiner Teile] In: Aristoteles, *Metaphysik VII,17.*

[3] Quoted in: Frank Wilczek, Betty Devine, *Longing for the harmonies (1988).*

Hence, the numerical simulations of the microscopic level are often too costly (in terms of computation time) and therefore – due to efficient numerical schemes for partial differential equations – the macroscopic equations play an important role for simulations. Additionally, because of the high complexity, interesting questions like long-term clustering of particles are difficult to answer if we only consider the microscopic dynamics.
Certainly, there are different mathematical approaches used to motivate or rigorously prove connections between microscopic particle systems and macroscopic equations. In this thesis, so-called *mean-field limits* and the associated concept of *propagation of chaos* is used in order to rigorously show connections between certain particle systems and partial differential equations. In recent years, mean-field limits have become a growing field of mathematical research. The reason lies in the fact that those particle limits can be used in a broad variety of applications, like swarm modelling [18], deep learning [106], neuroscience [2], evolutionary biology [23] and economy [98], to name a few. The particles can for instance represent molecules, neurons, bacteria, plants or humans. Caused by this broad variety of applications, the topic has also become an emerging tool for modern interdisciplinary research.
The origin lies in the late 19th century, where Boltzmann already proposed that the particle dynamics of a large class of particle systems can be captured by one macroscopic partial differential equation (PDE), see [7]. Heuristically, this means that (for the considered cases) for large systems the particles behave like 'one' and become independent in the limit; see Section 1.1 for a mathematical definition of this intuition. In 1900, at the International Congress of Mathematicians, David Hilbert (1862-1943) famously addressed this question in his sixth problem, where he claimed that providing an axiomatic mathematical framework for Boltzmann's considerations should be one of the goals in modern mathematics, [57], [85]. However, caused by a lack of suitable mathematical techniques at that time, like important results in probability theory, it took more than fifty years until Mark Kac [66] made significant progress in this matter by mathematically formalising the notion of *chaos* for the Boltzmann equation, see Section 1.1.1 for details.
Despite the lack of a mathematical framework at that time, Boltzmann's idea to consider particle systems which become 'independent' if the number of particles becomes large, still forms the core motivation of mean-field limits, where the limiting macroscopic equation (represented by a non-linear PDE) of a large system of interacting particles is studied. The microscopic particle system is described by a large system of (stochastic) ordinary differential equations. The interaction between one particle with all other particles is incorporated into the system by using a weighted sum over all interactions - a *mean value* - which motivates the name *mean-field limit*. Since the particle system in mean-field theory is usually represented by a stochastic system of interacting particles and the macroscopic dynamics are represented by deterministic partial differential equations, which model the *typical particle* (since the particles behave like 'one' in the limit), the topic of mean-field limits lies on the border between two mathematical disciplines: Stochastics and Partial Differential Equations (PDEs). The challenge is to take advantage of different techniques from those two mathematical fields despite those two disciplines often times having different notations and aims. As a matter of fact, the techniques used in this thesis will cover both disciplines, however, they strongly rely on analytical techniques like uniform estimates of solutions of (non-local and local) non-linear partial differential equations, which are crucial

especially for the results in Chapter 2 and 3. In Chapter 4, a (more technical) result is shown by proving a connection between two different notions of convergence of particle systems for a diffusion model with aggregation. However, even if at the first glimpse it looks like a purely probabilistic result, the new technique used there is inspired by estimates from classical PDE theory.

1.1 Notion of chaos

There are different notions of chaos and hence *propagation of chaos*, the standard one being derived by the framework given by Mark Kac in [66]. Originally, the notion of chaos in [66] was considered to be suitable for the kinetic Boltzmann theory, however, the lecture notes of H.P. McKean [82] addressed ten years later that Kac's concept of *chaos* and *propagation of chaos* can also be used in the framework of a broad class of (nonlinear) diffusion models of the form

$$\partial_t u = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial_{x_i}\partial_{x_j}}[a_{ij}(u)u] - \sum_{i=1}^{d} \frac{\partial}{\partial_{x_i}}[b_i(u)u] \quad \text{on } (0,\infty)\times\mathbb{R}^d, \tag{1.1}$$

where the diffusion coefficients $a_{ij}(u)$ as well as the drift coefficients $b_i(u)$ depend (in a nonlinear and nonlocal way via an integral formulation) on the solution u and on the spatial variable x. The precise form of the diffusion and drift coefficients a_{ij} and b_i will be discussed in Section 1.2, where we will illustrate some ideas given in [82] for a simple example. Before we discuss this toy example, in the following part of the thesis, we introduce the general concept of *chaos* and *propagation of chaos* which form the core idea and motivation of mean-field limits.

1.1.1 Kac: The introduction of a mathematical framework

The scope of the present thesis lies in the derivation of non-linear partial differential equations with diffusion (and aggregation) phenomena arising from physics and biology from stochastic particle systems and not in kinetic theory. However, since the concept of those particle derivations and mean-field limits strongly relies on the concept of the so-called *Boltzmann property* which was first introduced by Kac in [66], in this section we present a short summary of the seminal work [66], which is not only an important work for kinetic theory but for particle derivations of partial differential equations in general. This section is based on [66] and the recent articles [56] and [85]. For more information on Kac's work and implications in kinetic theory we refer to the latter two papers.
Based on Boltzmann's work, [7] and his well-known '*Stosszahlansatz*', [66, Section 2], Kac developed a mathematical framework for Boltzmann's intuition and ideas for kinetic theory for dilute gases. In the setting of the spatially homogeneous Boltzmann equation of the form

$$\partial_t f(t,v) = Q(f,f) \quad \text{for } t \geq 0 \ \ , v \in \mathbb{R}^d,$$

where v denotes the velocity, Q denotes a so-called *collision operator* and $f(t,v)$ is a distribution function of a dilute gas, where we assume that the exchange of energy between

the gas molecules only happens through collisions. This spatially homogeneous equation was derived by Kac in [66] in a simplified setting (i.e. $d = 1$ and other simplifications of the model) by using Poisson-like jump processes for the mutual collisions between two molecules on the microscopic level, see [66, Sections 2-3] for a complete description. In his BASIC THEOREM, [66, Section 3], Kac showed that the so-called *Boltzmann property* (see definition below) propagates in time in the following way:
Let us assume that $d = 1$ and define $V = (v_1, \ldots, v_N) \in \mathbb{R}^N$ the vector of velocities of the N gas molecules in the system on spheres $\mathcal{S}_N := \{V : v_1^2 + \ldots + v_N^2 = N\}$[4] and let $\phi_N(V, t)$ fulfil the so-called *master-equation*, which is a PDE that describes the change in time of the distribution of points V under the influence of (random) collisions, see [66, Equation (3.4)]. Additionally, set

$$f_N^{(k)}(v_1, \ldots, v_k, t) := \int_{x_{k+1}^2 + \ldots + x_N^2 = N - v_1^2 - \ldots - v_k^2} \phi_N(V, t) \mathrm{d}S,$$

where we integrate over spheres which fulfil $x_{k+1}^2 + \ldots + x_N^2 = N - v_1^2 - \ldots - v_k^2$. The distribution functions $f_N^{(k)}$ are called *k-th contraction* of ϕ_N in [66]. Then the BASIC THEOREM in [66] says that if at time $t = 0$, the symmetric distribution function $\phi_N(V, 0)$ with $V \in \mathcal{S}_N$ fulfils the *Boltzmann property*, i.e. for all $k \in \mathbb{N}$

$$\lim_{N\to\infty} f_N^{(k)}(v_1, \ldots, v_k, 0) = \prod_{i=1}^{k} \lim_{N\to\infty} f_N^{(1)}(v_i, 0),$$

then it also holds at any time $t > 0$

$$\lim_{N\to\infty} f_N^{(k)}(v_1, \ldots, v_k, t) = \prod_{i=1}^{k} \lim_{N\to\infty} f_N^{(1)}(v_i, t).$$

For a proof of this statement, we refer the reader to [66, Section 4]. In the upcoming sections we will see that the observation that the Boltzmann property concerning the finite '*contractions*' of the distribution function ϕ_N propagates in time (under simplifying assumptions for the Boltzmann equation) forms the basic concept of mean-field limits and propagation of chaos. Based on [66], McKean [82] made use of this general concept by applying it to a class of parabolic nonlinear partial differential equations. The definition stayed close to the original one in [66] – see Definition 1 –, however, in modern literature, the name *Boltzmann property* changed into *u-chaos* or *Kac's chaos*[5] .

1.1.2 Propagation of chaos

Based on the well-known framework in [66], the following section contains important definitions and notions used in every chapter of this thesis. This section follows [62] and [113] with additional insights into further notions of chaos from [56].

[4]Later called 'Kac's spheres', [56].
[5]In this thesis we will use the name u-chaos, see Definition 1.

First, we state the definition of *chaos* used in the classical framework of mean-field limits:

Definition 1 (u-chaotic, Definition 2.1. in [113])**.** *Let u be a probability measure on a separable metric space $\mathcal{M}$. Then, a sequence f_N of symmetric probability measures*[6] *on the product space $\mathcal{M}^N$ is said to be u-chaotic if for any finite number $k \in \mathbb{N}$ and bounded, continuous functions $\phi_1, \dots, \phi_k \in C_b(\mathcal{M})$ it holds that*

$$\lim_{N\to\infty} \langle f_N, \phi_1 \otimes \ldots \otimes \phi_k \otimes 1 \otimes \ldots \otimes 1 \rangle = \prod_{i=1}^{k} \langle u, \phi_i \rangle. \tag{1.2}$$

Remark 1.1 (Weak convergence of k-marginals)**.** *Note that condition (1.2) means that for all $k \in \mathbb{N}$ the k-marginal of the sequence f_N of symmetric probability measures on $\mathcal{M}^N$ converges weakly to the product measure $u^{\otimes k}$.*

Before we continue with stating the definition of *propagation of chaos*, we have to fix some ideas and the framework of (interacting) particle systems. We will explain more about the specific form of mean-field particle systems, in the following section (Section 1.2), however, in order to properly state the definition of propagation of chaos, we need the following concept of *empirical measures*: Let $\mathcal{X}_N(t) := (X_1(t), \dots, X_N(t))$ be a sequence of N (interacting) indistinguishable[7] particles at a certain time $t \geq 0$. Then, we define the associated *empirical measure* at time $t \geq 0$ as the following random distribution

$$\mu_{\mathcal{X}_N}(t, x) := \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i(t)}(x), \quad t \geq 0, \tag{1.3}$$

where $\delta_X(\cdot)$ denotes the Dirac delta at point X.

Proposition 1.2 (Equivalent statements to u-chaotic, Proposition 2.2 in [113])**.** *Under the assumption of Definition 1, the following statements are equivalent*

(i) *f_N is u-chaotic*

(ii) *Condition (1.2) holds for $k = 2$, which means that it is sufficient to show convergence of the second marginal of f_N towards the product measure $u^{\otimes 2}$; (see Remark 1.1)*

(iii) *If $\mathcal{X}_N := (X_1, \dots, X_N)$ is distributed according to f_N, i.e. Law$(X_1, \dots, X_N) = f_N$ for all $N \in \mathbb{N}$, then the associated empirical measure $\mu_{\mathcal{X}_N}(\cdot)$ converges in law towards the deterministic measure u, where the empirical measure is defined in (1.3).*

For a proof of this proposition, we refer to [113]. We note that for showing convergence in law of the empirical measure $\mu_{\mathcal{X}_N}$ is enough to prove that for any test function $\phi \in C_b(\mathcal{M})$, which is bounded and continuous, it holds that $\mathbb{E}(|\frac{1}{N}\sum_{i=1}^{N} \phi(X_i) - \int_{\mathcal{M}} \phi(x) \mathrm{d}u(x)|) \to 0$ for $N \to \infty$, see [113] and [83].

Remark 1.3 (Law of large numbers)**.** *Since statement (iii) in Proposition 1.2 shows that the empirical measure $\mu_{\mathcal{X}_N}$ converges to a deterministic measure u, this can be seen as a version of 'Law of large numbers' for particle systems.*

[6]'Symmetric' means invariant under permutations of the coordinates

[7]The assumption of indistinguishability implies that the joint law is symmetric

Now, we have all definitions in hand to define *propagation of chaos*. Let

$$\mathcal{X}_N(t) := (X_1(t), \dots, X_N(t))$$

be a sequence of N interacting indistinguishable particles and $t \geq 0$. Additionally, let $u(t)$ be a solution to a partial differential equation with initial condition u_0 (the concrete form of this PDE will be made explicit in Section 1.2; to understand the concept it is not important).

Definition 2 (Propagation of chaos for a particle system, Definition 4 in [62])**.**
If at time $t = 0$ the joint distribution $f_N(0)$ of $\mathcal{X}_N(0)$ is u_0-chaotic, then we say that propagation of chaos holds, if at any time $t > 0$ the joint distribution $f_N(t)$ of $\mathcal{X}_N(t)$ is $u(t)$-chaotic.

Due to Proposition 1.2 this implies that $\mu_{\mathcal{X}_N}(t) \to u(t)$ for $N \to \infty$ holds true in law at any time, which shows that in this case the particle dynamics converge at any time to the deterministic law $u(t)$.

Remark 1.4. *Choosing $X_i(0) = \xi_i$ with independent and identically distributed random variables on $\mathcal{M}$ such that $Law(\xi_i) = u_0$, implies that $f_N(0)$ is trivially u_0-chaotic.*

Heuristically, we can interpret the *propagation of chaos property* in the setting of interacting particle systems in the following way: Let us start with independent and identically distributed random variables at time $t = 0$. At any point $t > 0$ – since the particles are interacting (see Section 1.2) – they are not independent any more. However, as the number of particles N grows, this property of independence (which implies a factorised law) can be recovered in the large particle limit for any time $t > 0$.

Multi-species propagation of chaos

In Chapter 2 of this thesis, we show a propagation of chaos result for a multi-species model, where we use an extended variant of Definition 1 that was already used for instance in [2] in the framework of a multi-species neural network. Due to the symmetry assumption on the probability measures in Definition 1 (and therefore the assumption of indistinguishability of all particles)– which is not true for multi-species models – Definition 1 has to be adapted for a multi-species case. Let $N \in \mathbb{N}$ be the total number of particles and $n \in \mathbb{N}$ the number of species. We denote the particle dynamics with

$$\mathcal{X}_N(t) := (X_1^1(t), \dots, X_N^n(t)),$$

where the upper index denotes the species and assume that particles within one species are indistinguishable. Then, we say that at time $t \geq 0$, the sequence of joint laws f_N of $\mathcal{X}_N$ is $u(t) = (u_1(t), \dots, u_n(t))$-chaotic, if for any $k \in \mathbb{N}$, the law of the k-tuple $(X_{i_1}^{s_1}(t), \dots, X_{i_k}^{s_k}(t))$ converges weakly to the product measure $\prod_{i=1}^k u_{s_i}(t)$. Here, each species has a different limiting process, however chaos still propagates in time. In Chapter 2, we show propagation of chaos for a multi-species model in a different (stronger) sense, by path-wise estimates, see Section 1.2 for an introduction to this *coupling technique*.

Other notions of chaos

Definition 1 and Definition 2 go back to the classical work [66]. However, it might be useful to also use other notions of chaos. In this section, we give some insight into other notions of propagation of chaos, which are governed by using different norms of convergence. The idea however stays the same as in Definition 2. We follow the review paper [62] and the article [56].

Convergence in Monge-Kantorovich-Wasserstein distance, [62]. In some applications, propagation of chaos is shown with respect to the p-MKW-distance: Let ρ_1, ρ_2 be two probability measures with finite p-th moment, then the p-MKW distance for $p \geq 1$ is defined as follows

$$\mathcal{W}_p(\rho_1, \rho_2) := \inf_{\substack{(X,Y), \mathrm{Law}(X)=\rho_1 \\ \mathrm{Law}(Y)=\rho_2}} \Big(\mathbb{E}(|X - Y|^p)\Big)^{1/p},$$

see [62] for instance. Hence, one can look at *propagation of chaos in MKW-distance* in the following way:

Definition 3. *Let the assumptions in Definition 2 hold. Let $f_N^k(t)$ denote the distribution of the k-marginal of $f_N(t)$ for any $t \geq 0$ and*

$$\mathcal{W}_p(f_N^k(0), u_0^{\otimes k}) \to 0 \quad \text{for } N \to \infty.$$

Then, propagation of chaos holds in p-MKW norm if for any time $t > 0$

$$\mathcal{W}_p(f_N^k(t), u^{\otimes k}(t)) \to 0 \quad \text{for } N \to \infty.$$

We refer the reader also to Section 1.2, where we show that convergence in expectation of the second moment (using so-called coupling techniques) implies propagation of chaos in 2-MKW-distance. See also article [56], where different implications between the notation of chaos by Kac and propagation of chaos in 1-MKW-distance are shown. In the present thesis, with regard to the MKW-distance, only the above mentioned implication (1.14) shown in the next section is of relevance, especially in Chapter 2 and 3.

Another notion of chaoticity worth mentioning in this section is concerned with convergence of the entropy functional:

Definition 4 (u-entropy-chaotic, [62, 56])**.** *Let all assumptions of Definition 1 hold. By defining the Boltzmann-entropy for f_N as follows*

$$\mathcal{H}_N(f_N) := \frac{1}{N} \int_{\mathcal{M}^N} f_N \log f_N \mathrm{d}x_1 \dots \mathrm{d}x_N,$$

we say that f_N is u-entropy chaotic if $H_1(u) < \infty$ and

$$H_N(f_N) \to H_1(u).$$

Indeed, one can show that the notion of entropic chaos is stronger than the chaoticity defined in Definition 1:

Proposition 1.5 (Theorem 1.4 (iii) and (iv) in [56]). *Let all assumptions of Definition 1 and Definition 4 hold and $\mathcal{M} = \mathbb{R}^d$. If f_N is u-entropy chaotic, then it is u-chaotic.*

For a proof we refer to [56], where property (1.2) is referred to as *Kac-chaotic*. We refer the reader to [62] and [56] and the references therein for further discussions on other notions of chaos, like *Fisher-information-chaotic*, which is even stronger than entropy-chaotic, [56, Theorem 1.4].

1.1.3 Outlook: Fluctuations around the mean-field limit

As mentioned in Remark 1.3, proving a mean-field limit result and the associated propagation of chaos property can be seen as a *law of large numbers result* on the level of the empirical measure of the particle dynamics. However, by approximating a stochastic interacting particle system by a deterministic measure, some information inherited by the stochasticity of the system gets lost. This is the reason why the study of *fluctuations around the mean-field limit*, which can be seen as *next order correction* to the mean-field behaviour, is of high interest. Questions associated with fluctuations in mean-field settings have already been studied by Braun and Hepp [9], Rost [100], Dawson [39], Sznitman [112], Oelschläger [90], Lewicki [74] as well as Jourdain and Méléard [65] in the last century and recently by [105] (in the setting of neural networks) and Wang et al. [117] for instance.
In this section, we give a motivational introduction:
If by denoting with $\mu_{\mathcal{X}_N(t)}$ empirical measure of the particle dynamics at time $t > 0$, see (1.3), 'propagation of chaos' means $\mu_{\mathcal{X}_N(t)} \to u(t)$ in law for a deterministic measure $u(t)$.[8] Since this corresponds to the law of large numbers, it is a natural question to ask, whether the quantity

$$\xi_N(t) := \sqrt{N}(\mu_{\mathcal{X}_N(t)} - u(t)) \tag{1.4}$$

associated with the well-known *central limit theorem* from standard probability theory converges (in a distributional sense). For mean-field interacting particle systems, the random measure $\xi_N(t)$ is called *fluctuation process around the mean-field limit $u(t)$.*
Assuming that there exists a limiting distribution of ξ_N, denoted by ξ, then formally, this shows why studying the limiting behaviour of ξ_N can be seen as 'next order correction': By writing $\mu_{\mathcal{X}_N(t)} = u(t) + \frac{1}{\sqrt{N}}\xi_N \approx u(t) + \frac{1}{\sqrt{N}}\xi$, the term $\frac{1}{\sqrt{N}}\xi$ gives us a correction of the limiting behaviour measured by the deterministic measure $u(t)$ that vanishes at scale $N^{-1/2}$. By recalling the classical central limit theorem for independent random variables, see [69, Theorem 17.10] and [41] for a formulation for empirical measures, we note that in case of independent particles, the limiting distribution ξ is Gaussian. However, as already mentioned in the section before, for interacting particles independence can clearly not be expected. In the spirit of Definition 2, one could ask a (formal) question like: If at time $t = 0$ the limit $\xi_N(0) \to \xi_0$ for $N \to \infty$ towards a Gaussian distribution holds, does it hold

[8] In the next section we will see that in case of mean-field interacting particles $u(t)$ solves a partial differential equation.

at any time $t > 0$ in the limit, i.e. $\xi_N(t) \to \xi_t$ for $N \to \infty$ and $t > 0$, where ξ_t is a Gaussian distribution[9]?
In general, one can not expect this question to be answered positively for all particle systems fulfilling the propagation of chaos property, see the work of Dawson [39] where he showed a phase transition result for the fluctuation process in a specific mean-field setting. However, for instance in [46] Fernandez and Méléard showed a general result for McKean-Vlasov dynamics where a *central limit theorem* holds.
When studying the limiting behaviour of ξ_N defined in (1.4), the convergence rate of the propagation of chaos property $\mu_{\mathcal{X}_N(t)} \to u(t)$ is of great importance, however, it is a challenging task to show the *optimal rate* of convergence. In some situations, changing the scaling in (1.4) from $N^{1/2}$ to a sequence $e_N < N^{1/2}$ might be fruitful, see for example [65] for a result in the moderate regime (see Section 1.2.2 for a definition of *moderate regime*) where e_N is chosen to be logarithmic in N. In this case, the limiting distribution of the fluctuation process is deterministic and not Gaussian.

Connection to this thesis:

- In Chapter 2, we derive a cross-diffusion system (multi-species) with linear diffusion from a mean-field interacting particle system, which implies a propagation of chaos result. Interestingly, if we set the number of species to one, the limiting partial differential equation reduces to a porous media equation with additional diffusion $\sigma > 0$:

$$\partial_t u = \sigma \Delta u + \frac{1}{2}\Delta(u^2) = \sigma \Delta u + \operatorname{div}(u \nabla u). \tag{1.5}$$

  We note that (1.5) was already derived almost 15 years before in [47] by Figalli and Philipowski with a different particle system. The main difference between the two derivations can be heuristically explained form an analytical point of view by the fact that we can write the Laplace-Operator Δ as $\operatorname{div}(\nabla)$ or interpret it as 'pure diffusion'.[10]

  This shows that there exist two different particle dynamics (and hence two different empirical measures $\mu_{\mathcal{X}^1_N(t)}(\cdot)$ and $\mu_{\mathcal{X}^2_N(t)(\cdot)}$) converging in law to the same deterministic measure $u(t)$, which solves (1.5).

  We expect that the fluctuation processes of the two particle systems respectively show different limiting behaviours, which would help us to understand the difference of the particle models from a modelling point of view. Partial results have been derived by Oelschläger in [90], where he was able to prove a central limit theorem for a *corrected fluctuation process*[11] in the setting of Figalli and Philipowski and by Jourdain and Méléard [65] who where able to show convergence of the fluctuation process with a different scaling than $N^{1/2}$. However, a complete picture is still missing in the literature.

[9] We do not specify here what *Gaussian* means in this context; in fact we are talking about generalized Ornstein-Uhlenbeck processes, see [58] for a definition.
[10] In terms of equation (1.7) of the following section: $V_1 = 0$ in Chapter 2, $V_2 = 0$ in [47].
[11] Corrected means $\xi_N = \sqrt{N}(\mu_{\mathcal{X}_N(t)} - u(t) + c_N)$, where c_N is a deterministic correction which fulfils $c_N \to 0$ for $N \to \infty$.

- Motivated by the findings in Chapter 2 described above, in Chapter 4, we make a step towards extending Oelschläger's technique of [90] for showing a *(corrected) fluctuation result* in a more general setting, i.e. we also allow aggregating particles instead of only repulsive particles in [90]. In terms of the limiting partial differential equation, we allow both signs $\pm$ in front of the non-linear term:

$$\partial_t u = \sigma \Delta u \pm \frac{1}{2}\Delta(u^2) = \sigma \Delta u \pm \operatorname{div}(u \nabla u). \tag{1.6}$$

However, a complete *fluctuation result* for aggregating particles in the setting of (1.6) is still an open question, since the new technique developed in Chapter 4 only gives a partial result by assuming that at least a propagation of chaos result by coupling methods holds in probability (see Section 1.2.1 for an introduction to coupling techniques). Nonetheless, the method developed in Chapter 4 is expected to hold also in models of *Keller-Segel-type*, see the appendix of Chapter 4 (Section 4.A), and also forms an important step towards fluctuation results in the framework of cross-diffusion models. We refer to Chapter 4 and the summary in Section 1.3 for a more detailed introduction.

1.2 Particle systems of mean-field type

In order to simplify the notation, in this section we will always consider a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, even when not specifically written.
Throughout this thesis we will consider the particle system to follow a stochastic differential equation of *mean-field type*, i.e. the position of the i-th particle changes in time according to the following system of stochastic differential equations

$$\begin{aligned} \mathrm{d}X_i^N(t) &= \nabla U(X_i^N(t))\mathrm{d}t + \frac{1}{N}\sum_{j=1}^N V_1(X_i^N(t) - X_j^N(t))\mathrm{d}t \\ &\quad + \left(\sigma + \frac{1}{N}\sum_{j=1}^N V_2(X_i^N(t) - X_j^N(t))\right)^{1/2} \mathrm{d}W_i(t) \qquad i = 1, \dots, N, \end{aligned} \tag{1.7}$$

where $(W_i(t))_{i=1}^N$ is a family of independent Brownian motions and $X_i^N(t) \in \mathbb{R}^d$ denotes the position of the i-th particle in $\mathbb{R}^d$ at time $t \geq 0$.

- $\nabla U : \mathbb{R}^d \to \mathbb{R}^d$ can be seen as an *environmental potential*,
- $V_1 : \mathbb{R}^d \to \mathbb{R}^d$ is considered as *interaction kernel* of the drift part, whereas $V_2 : \mathbb{R}^d \to \mathbb{R}^d$ measures the interaction in the diffusion part. We remark that in this thesis the interaction only depends on the spatial difference of the particles, however, also more general interactions can be considered, see the lecture notes [113], [83] and the review papers [21], [22] for discussion on a more general framework.

In this thesis, we will always consider diffusive particle systems, which means that the diffusion parameter $\sigma > 0$ is strictly positive. However, one of the interaction kernels

V_1, V_2 might vanish: In Chapter 2, we only consider interaction in the diffusion part in order to show a convergence result in the setting of cross-diffusion models, i.e. $V_1 = 0$ in that chapter. This structure is crucial for deriving the so-called SKT-model. In Chapter 3, which is concerned with the derivation of a fractional porous media equation, we only consider interaction in the drift part, i.e. $V_2 = 0$. Additionally, we want to remark that in Chapter 3, we let $\sigma \to 0$ in the end on the level of partial differential equations. On the level of interacting particles we always consider the case that diffusion is present. In Chapter 4, which deals with local diffusion-aggregation models, aggregation is incorporated in the model by mean-field interaction in the drift part by using a gradient structure, see the following section and equation (1.18) for an introduction.

1.2.1 Coupling and Itô's formula

In the introduction of this thesis, we have mentioned that the general concept of mean-field limits is to show convergence of a (stochastic) particle system towards a solution of a deterministic partial differential equation (PDE). In Section 1.1, we gave a mathematical definition of propagation of chaos which connects the finite-time marginals of a particle system with the product measure of a limiting distribution, which solves a PDE - in this section we will discuss the specific form of this PDE in more detail. Additionally, since often times in the present work, we show convergence of the microscopic dynamics not by proving the *propagation of chaos property* (Definition 2) directly, but by using a technique which we refer to as *coupling techniques*, we will also provide general information about the concept of *coupling* in this section. This technique is based on introducing an additional particle system, the so-called *non-linear process*, which is *not* an interacting particle system anymore. Instead the particles are independent from each other and have a common density function, which – under suitable assumptions on the initial data and the interaction kernels – solves certain PDE. Then, we show convergence of a particle system towards a solution of this PDE by showing convergence of a particle X_i^N towards its limiting non-linear process $\overline{X}_i$. In order to fix this idea, in the following part we summarise the toy example that was presented in the well-known lecture notes by Alain-Sol Sznitman, [113].

Toy example for coupling techniques

This section mainly follows the lecture notes [113], but also incorporates aspects of the lecture notes by Sylvie Méléard [83] and the review paper by Jabin and Wang [62]. In the subsequent we consider (for simplicity) the following particle system of mean-field type:

$$\mathrm{d}X_i^N(t) = \frac{1}{N}\sum_{j=1}^{N} V_1(X_i^N(t) - X_j^N(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \tag{1.8}$$

$$X_i^N(0) = \xi_i, \quad \text{on } \mathbb{R}^d \qquad i = 1, \dots, N, \tag{1.9}$$

which corresponds to (1.7) with $\nabla U, V_2 = 0$ and where we assume that the initial conditions ξ_i are independent and identically distributed on $\mathbb{R}^d$ with density function u_0. Additionally, we assume that $V_1 : \mathbb{R}^d \to \mathbb{R}^d$ is globally Lipschitz continuous and bounded.

Next, we introduce the so-called *non-linear process* associated with (1.8):

$$\mathrm{d}\overline{X}_i(t) = \Big[V_1 * u(t, \overline{X}_i(t))\Big]\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \tag{1.10}$$

$$\overline{X}_i(0) = \xi_i \quad i = 1, \dots, N, \tag{1.11}$$

where $*$ denotes the convolution in space on $\mathbb{R}^d$ and $u(t)$ denotes the law of the non-linear process $\overline{X}_i(t)$. The existence and uniqueness of the solution to (1.10) is shown in [113, Theorem 1.1] under the boundedness and Lipschitz assumption on V_1 stated in the beginning of this section. However, the boundedness assumption can be weakened by assuming boundedness of the second moment of u_0, see [83, Theorem 2.2]. Interestingly – and important for this thesis – under suitable assumptions on the initial data, the density function of the law $u(t)$, which we also denote by $u(t)$, can be written as a solution to the following PDE

$$\partial_t u = \sigma\Delta u - \operatorname{div}\big((V_1 * u)u\big), \quad u(0) = u_0. \tag{1.12}$$

This can be (formally) seen by using Itô's formula, which we recall for the reader's convenience here:

Theorem 1.6 (Itô's formula, in Theorem 4.2.1 [95])**.**
Let the function $\phi(t, x) = (\phi_1(t, x), \dots, \phi_k(t, x)) \in C^2([0, \infty) \times \mathbb{R}^d; \mathbb{R}^k)$ *and* $X(t) = (X_1(t), \dots, X_d(t))$ *be an d-dimensional Itô process which fulfils the following SDE*

$$\mathrm{d}X(t) = b(t)\mathrm{d}t + s(t)\mathrm{d}W(t).$$

Then $Y(t) = \phi(t, X(t))$ *is a k-dimensional Itô process which fulfils*

$$\mathrm{d}Y_j = \partial_t\phi_j(t, X)\mathrm{d}t + \sum_{i=1}^d \partial_{x_i}\phi_j(t, X)\mathrm{d}X_i + \frac{1}{2}\sum_{i,\ell=1}^d \partial_{x_i x_\ell}\phi_j(t, X)\mathrm{d}X_i\mathrm{d}X_\ell,$$

where $\mathrm{d}W_i\mathrm{d}W_j = \delta_{ij}\mathrm{d}t, \mathrm{d}t\mathrm{d}W_i = \mathrm{d}W_i\mathrm{d}t = \mathrm{d}t\mathrm{d}t = 0.$

Let $C_b^2(\mathbb{R}^d)$ denote the space of bounded and twice continuously differentiable functions on $\mathbb{R}^d$ taking values in $\mathbb{R}$, where all derivatives up to the second order are bounded. Then, one sees that Itô's formula implies that for every $\phi \in C_b^2(\mathbb{R}^d)$ it holds that

$$\begin{aligned}\phi(\overline{X}_i(t)) - \phi(\overline{X}_i(0)) = &\int_0^t \nabla\phi(\overline{X}_i(s))\mathrm{d}W_i(s) \\ &+ \int_0^t \sigma\Delta\phi(\overline{X}_i(s)) + V_1 * u(s, \overline{X}_i(s))\nabla\phi(\overline{X}_i(s))\mathrm{d}s.\end{aligned}$$

Taking the expectation and using the regularity of the test function ϕ and that $u(t)$ is the law of $\overline{X}_i(t)$ leads to a weak formulation of the PDE (1.12). Sometimes in literature, this formulation is called *very weak formulation*, since all derivatives are on the test function. Note that in order to show the general concept of coupling techniques used in this thesis, the arguments here are not rigorous since we do not justify that the law of the nonlinear process is indeed absolutely continuous with respect to the Lebesgue measure and that we have sufficient regularity of the density function u, see Chapter 2 where such an argument is performed (for multi-species models) in a more rigorous way.

For this toy example (1.8) the following convergence result towards the non-linear process $\overline{X}_i$ holds:

Theorem 1.7 (Convergence to nonlinear process, Theorem 1.4. in [113])**.** *Under the assumptions on V_1 made in the beginning of this section, for any $T > 0$ it holds that*

$$\sqrt{N}\mathbb{E}\left(\sup_{0\le t\le T}\left|X_i^N(t) - \overline{X}_i(t)\right|\right) \le C,$$

where $C > 0$ is a constant not dependent on the number of particles N, but can be dependent on the Lipschitz constant of V_1.

The proof can be done in a straightforward way by exploiting the independence of system (1.10) and using a Gronwall-type argument, see [113, Theorem 1.1]. In [83, Theorem 2.3] a similar result is shown by proving convergence of the second moment

$$\mathbb{E}\left(\sup_{0\le t\le T}\left|X_i^N(t) - \overline{X}_i(t)\right|^2\right) \le C/N, \tag{1.13}$$

where the constant $C > 0$ also depends on the Lipschitz bound of V_1 under the additional assumption that u_0 has finite second moment.

An important implication of Theorem 1.7 and (1.13) is the propagation of chaos property: First, it is easy to see that (1.13) implies propagation of chaos in 2-Monge-Kantorovich-Wasserstein distance, since by denoting with $f_N^k(t)$ the distribution of the k-th marginal of the common distribution of $(X_1^N(t), \ldots, X_N^N(t))$ if follows

$$\mathcal{W}_2^2(f_N^k(t), u^{\otimes k}(t)) \le \mathbb{E}(|(X_1^N - \overline{X}_1, \ldots, X_k^N - \overline{X}_k)(t)|^2) \le Ck/N \to 0 \qquad \text{for } N \to \infty. \tag{1.14}$$

A similar result holds true with $\mathcal{W}_1$ distance for the result in Theorem 1.7.
Second, we can show propagation of chaos in the sense of Definition 2 and Proposition 1.2. By defining the empirical measure of (1.8) via $\mu_N(t) := \sum_{i=1}^N \delta_{X_i^N(t)}$, see (1.3), a short calculation (which is also presented in [62, Section 3.1]), shows that Theorem 1.7 (and hence (1.13)) already implies the weak convergence of the empirical measure towards the limiting solution of the nonlinear PDE (1.12):
For any test function $\phi \in C_b^1(\mathbb{R}^d)$

$$\begin{aligned}\mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^N \phi(X_i^N(t)) - \int_{\mathbb{R}^d}\phi(x)u(t,x)\mathrm{d}x\right|\right) &\le \mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^N \phi(X_i^N(t)) - \frac{1}{N}\sum_{i=1}^N \phi(\overline{X}_i(t))\right|\right)\\ &+ \mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^N \phi(\overline{X}_i(t)) - \int_{\mathbb{R}^d}\phi(x)u(t,x)\mathrm{d}x\right|\right) \to 0,\end{aligned} \tag{1.15}$$

where the first term converges to zero due to Theorem 1.7 and the Lipschitz continuity of ϕ and the second term converges to zero due to the independence of $\overline{X}_i$ (having density function u) which implies that the law of large numbers holds in that case.
The above calculations show that by using this coupling approach, we do not only show propagation of chaos in the level of empirical measures and hence according to Definition 2, but a path-wise estimate on the level of particles, which can be seen as a stronger version of propagation of chaos than in Definition 2 and Definition 3.

Remark 1.8. *Following the lecture notes by McKean [82], one can see that this toy example can be extended for interaction in the diffusion part. Indeed, in Chapter 2 we will use the coupling method for a model with interaction in the diffusion part in a more complicated (multi-species) setting.*

Other notions of convergences: For the presented toy example we have shown with Theorem 1.7 and (1.13) two results where the convergence of the particles of the interacting system (1.8) towards the non-linear system (1.10) is shown with respect to expectation. However, in some situations it is useful to 'reduce' the type of convergence. For example, instead of showing convergence in expectation, one could show convergence in probability, i.e. for all $\alpha > 0$

$$\sup_{0<t<T} \mathbb{P}\big(|X_i^N(t) - \overline{X}_i(t)| > \alpha\big)^{12} \to 0 \quad \text{ as } N \to \infty. \tag{1.16}$$

One advantage of using a different notion of convergence – besides technical reasons – , might be that by using a *weaker notion of convergence* one might get better convergence rates in N. As used for instance in the works by Peter Pickl and co-authors [6, 72, 31] for Vlasov-type equations, one could also use the *cut-off parameter* α depending on the number of particles $N \in \mathbb{N}$, such that $\alpha(N) \to 0$ if $N \to \infty$. Heuristically, one can interpret this notion of convergence in the following way: We allow a *bad set* – where we allow particles to have a distance $\alpha(N)$ – with positive probability, however, as the number of particles increases, the probability of the set converges to zero. For more details, we refer to the above mentioned articles.

Connection to the present thesis: In Chapter 2 of this thesis, we extend the approach of classical coupling methods for multi-species cross-diffusion models with interaction in the diffusion part in the so-called *moderate regime* (see the following Section 1.2.2 for an explanation of this concept). In Chapter 3, we use coupling methods for showing convergence towards a fractional version of the porous media equation, where in contrast to the cross-diffusion setting, the interaction is only considered in the drift part but we deal with singularity of the kernel of $(-\Delta)^{-s}$ for $0 < s < 1$.
In Chapter 4, we also use the concept of coupling: Inspired by the techniques used by Pickl and co-workers [6, 72, 31])we show that under the assumption that convergence in probability holds (similar to (1.16)) with a certain cut-off rate $\alpha(N) > 0$ and an algebraic

[12] For simplicity we use here the euclidian norm as a measure of the difference between the particles; however, different *notions of distance* can be used; see [6], [72] and [31]

convergence rate, also convergence in L^2-norm of the smoothed empirical measures holds at rate $N^{-1/2-\varepsilon}$ with $\varepsilon > 0$. This result can be seen as a step towards extending Oelschläger's techniques for proving a fluctuation result in a more general setting, see Section 1.1.3 for an introduction of the concept of fluctuations and Chapter 4 for the exact result.

1.2.2 Strength of interactions - The concept of moderately interacting particles

In the standard mean-field setting, see particle system (1.7), the interaction potentials V_1, V_2 do not depend on the number of particles. As motivated in Section 1.2.1, by using the standard coupling techniques, the limiting PDE structure (1.12) will be of non-local type since it contains a convolution with the interaction kernel. Nevertheless, many partial differential equations arising from biology, physics and other applications are of *local type*, i.e. the partial differential equation at a point x does not depend on the values of the solution in a neighbourhood but solely on the point x. In order to derive such equations - which do not contain convolution or integral terms - one has to extend the classical concept of mean-field limits through introducing the so-called *strength of interaction* of the particles. We distinguish between *weakly, moderately* and *strongly interacting particles.* In particular, in order to derive partial differential equations of local type, in this thesis the concept of moderately interacting particles will be used in all chapters.

For this section, we follow the classification by Karl Oelschläger in [91] and consider the two different types of particle systems: The *classical diffusion setting* (1.17) and a particle system with *gradient structure* (1.18). Those two settings have to be treated slightly differently when it comes to the strength of interaction.

I. Classical Diffusion Process. First, we consider the particle dynamic for N particles on $\mathbb{R}^d$, with $d \geq 1$, where the equation for the i-th particle reads as follows

$$\mathrm{d}X_i^N(t) = \frac{1}{N}\sum_{\substack{j=1\\ j\neq i}}^{N} V_N(X_i^N(t) - X_j^N(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \qquad i = 1, \ldots, N, \tag{1.17}$$

where as usual by $(W_i(t))_{i=1}^N$ we denote a family of independent Brownian motions. We do not specify the initial condition since it is not important for the classification; the reader could just think of independent and identically distributed initial data.
However, we want to strongly emphasize that in difference to (1.7), the interaction kernel V_N can depend on the number of particles N, but we always assume that the scaling is in such a way that $\|V_N\|_{L^1(\mathbb{R}^d)} = 1$. We are now interested in the so-called *strength of interaction*, i.e. the influence of the interaction term $N^{-1}V_N(X_i^N(t) - X_j^N(t))$ between particle i and j on the dynamics for particle i in terms of N. We consider three cases:

1. **Weak Interaction.** If V_N does not depend on the number of particles, i.e. $V_N = V$ for all $N \in \mathbb{N}$, then the *strength of interaction* scales with N^{-1}, since the influence of the j-th particle on the movement of particle i can be measured by $N^{-1}V(X_i^N(t) - X_j^N(t))$.

In [91], this case is called regime of *weakly interacting particles*, which can be motivated by particle physics, [91]. However, one can also simply say that the strength in terms of the number of particles is lower than in the other two cases (strongly and moderately interacting particles) discussed below. The regime of weak interaction has been studied in many different settings. It goes back to the work of McKean [82], see also works by Braun and Hepp [9] (for Vlasov dynamics), Sznitman [113] and Méléard [83]. It leads – as shown in the section before – to non-local partial differential equations of convolution-type.

2. **Strong Interaction.** We speak of *strongly interacting particles*, if the interaction potential V_N scales in a way that the scaling N^{-1} cancels out, i.e.

$$V_N(x) = NV(N^{1/d}x) \quad \text{for all } x \in \mathbb{R}^d,$$

for a smooth function V with $\|V\|_{L^1(\mathbb{R}^d)} = 1$. In this case the strength of interaction is $O(1)$. In the limit, this then leads to an approximation of Poisson point processes, see [113, Chapter II.], which will not be covered in the present thesis.

3. **Moderate Interaction.** Analogous to the setting of strongly interacting particles, we let V be a smooth and normalised function on $\mathbb{R}^d$. Considering the following scaling for the interaction potential for $0 < \beta < 1/d$

$$V_N(x) = N^{\beta d}V(N^{\beta}x) \quad \text{for all } x \in \mathbb{R}^d,$$

the *strength of interaction* for each particle becomes $O(N^{-1+\beta d})$ which – for $0 < \beta < 1/d$ – is stronger than $O(N^{-1})$ but weaker than $O(1)$. Since the strength of interaction lies 'between' weakly and strongly interacting particles, this regime is called *moderately interacting* particles. As shown in [113, Chapter II.],[91] in this regime the term a_N, which can be seen as *variance of the mean-field force* since $V_N \to \delta_0$ in distributional sense, defined as follows

$$a_N := \mathbb{E}\left(\left[\frac{1}{N-1}\sum_{j=2}^{N} V_N(X_1^N(t) - X_j^N(t)) - u(t, X_1^N(t))\right]^2\right)$$

converges to 0 if and only if $\beta < 1/d$, where $u(t, \cdot)$ solves a local PDE[13]. This shows that in this case the mean-field interaction part approaches a local force. For the critical case $\beta = 1/d$ we have non-vanishing variance (and hence fluctuations) leading to a strong regime, as discussed before. The idea of *moderate interaction* can be also generalised by using a (not necessarily algebraic) scaling in N: Let $\eta(N)$ be a function in N with

$$V_N(x) = \eta(N)^{-d}V(\eta(N)^{-1}x) \quad \text{for all } x \in \mathbb{R}^d,$$

where $\eta(N) \to 0$ if $N \to \infty$ is in such a way that $0 < \eta(N)^{-1} < O(N^{1/d})$, see for example [65] where in comparison to Oelschläger's work [91] a logarithmic connection

[13]The concrete shape of the PDE is discussed at the end of this section; see (1.22).

between η and N is used. Then, the strength of interaction is $O(N^{-1}\eta(N)^{-d})$, which also lies 'between' weakly and strongly interacting particles and justifies that we also use the name *moderately interacting particles* in this case. This concept with logarithmic connection is used in Chapter 2.

In Chapter 3 and 4, where we connect η and N in a logarithmic and an algebraic way respectively, we use a general version of the following *gradient diffusion process*:

II. Gradient Diffusion Process. Similar to (1.17) we consider a particle system with *gradient structure*:

$$\mathrm{d}X_i^N(t) = \frac{1}{N}\sum_{\substack{j=1\\ j\neq i}}^{N} \nabla V_N(X_i^N(t) - X_j^N(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \qquad i = 1, \dots, N. \tag{1.18}$$

As mentioned before, this gradient structure will be used in Chapter 3 (in a more complicated setting) and Chapter 4. In this setting, we will also distinguish between *weakly, strongly and moderately interacting particles.* However, due to the gradient structure of the interaction kernel the classification changes for strongly and hence moderately interacting particles:

1. **Weak Interaction.** If V_N does not depend on the number of particles, i.e. $V_N = V$ for all $N \in \mathbb{N}$, then analogously as for the non-gradient structure the influence scales with N^{-1}, which we refer to as the regime of *weakly interacting particles.*

2. **Strong Interaction.** In analogous way as in the setting of classical diffusion processes, we use the notation

$$V_N(x) = N^{\beta d} V(N^{\beta} x) \quad \text{for all } x \in \mathbb{R}^d,$$

for some $\beta > 0$. The regime of strong interaction changes for gradient systems (1.18) in comparison to (1.17). This is motivated in [91] by the fact that – assuming heuristically that the particles $X_i^N(t)$ are already independent at any time $t > 0$ with common density function u – the variance of the 'force field' $F^N(t,x) := \frac{1}{N}\sum_{i=1}^N \nabla V_N(x - X_i^N(t))$ of the particle system (1.18) has variance of order $O(N^{-1+\beta(d+2)})$ at any point $x \in \mathbb{R}^d$ and time $t > 0$. If $\beta = 1/(d+2)$, the variance does not vanish for $N \to \infty$ leading to a regime of *strongly interacting particles* and non-trivial fluctuations of the force in the limit. To the best of the author's knowledge, the large particle limit has not yet been determined for this choice of β in the gradient case.

3. **Moderate Interaction.** As mentioned for the case of strongly interacting particles, the variance of the 'force field' $F^N(t,x) := \frac{1}{N}\sum_{i=1}^N \nabla V_N(x - X_i^N(t))$ of the particle system (1.18) at any point $x \in \mathbb{R}^d$ and time $t > 0$ has variance of order $O(N^{-1+\beta(d+2)})$. From this fact one can see that for any $0 \le \beta < 1/(d+2)$, the variance of the force field vanishes for $N \to \infty$, which heuristically shows that the particle system is converging to a system with deterministic 'force'. This leads to the regime of *moderately interacting particles*, which is used in Chapter 4. In a similar way as for

particle system (1.17), the idea of moderate interaction can be also generalised by using a not necessarily algebraic scaling in N:

$$V_N(x) = \eta(N)^{-d} V(\eta(N)^{-1}x) \quad \text{for all } x \in \mathbb{R}^d,$$

where $\eta(N) \to 0$ if $N \to \infty$ is such a way that $0 < \eta(N)^{-1} < O(N^{1/(d+2)})$. Then, the strength of interaction is $O(N^{-1}\eta(N)^{-d})$, which lies 'between' weakly and strongly interacting particles. This regime is considered in Chapter 3 of the present thesis.

In the following Table 1.1 we recall the different regimes of *weak, moderate* and *strong* interaction for a interaction kernel scaled in N via $V_N(x) = N^{\beta d} V(N^{\beta} x)$ for all $x \in \mathbb{R}^d$.

	Classical Diffusion Process (1.17)	Gradient Diffusion Process (1.18)
Weak Regime	$\beta = 0$	$\beta = 0$
Moderate Regime	$0 < \beta < 1/d$	$0 < \beta < 1/(d+2)$
Strong Regime	$\beta = 1/d$	$\beta = 1/(d+2)$

Table 1.1: Classification of the strength of interaction according to [91].

Remark 1.9. *Despite the fact that the classification of [91] was done in a framework where interaction is only present in the drift part of the particle system (this corresponds to $V_2 = 0$ in (1.7)), the term 'moderately interacting particles' is also used in a more general situation, where interaction is also part of the diffusion part, see [65] or Chapter 2 of this thesis.*

Moderate interaction and the connection to local partial differential equation

In all three following chapters (Chapter 2, Chapter 3 and Chapter 4) of this thesis, we work in the regime of moderate interaction (either using a gradient structure similar to (1.18) in the last two chapters or the classical diffusion model with interaction in the diffusion part in Chapter 2), since all chapters of this thesis are concerned with the derivation of *local* partial differential equations from interacting particle systems. For illustrative reasons, let us start with the classical diffusion model (1.17): By using a moderate scaling we see that

$$V_N(x) \to \delta_0(x) \quad \text{for } N \to \infty \qquad \text{in distributional sense,} \tag{1.19}$$

where we recall that $V_N(x) = \eta(N)^{-d} V(\eta(N)^{-1}x)$ for $x \in \mathbb{R}^d$ with V a symmetric, non-negative, smooth function with $\int_{\mathbb{R}^d} V(x)\mathrm{d}x = 1$ and $\eta(N) \to 0$ for $N \to \infty$ where $0 < \eta(N)^{-1} < O(N^{1/d})$. Additionally, for simplicity we assume that V is compactly supported on the unit-ball in $\mathbb{R}^d$ denoted by $B_1(0)$[14].
By (1.19), we see (at least formally) that the limiting equation (1.12) becomes local, since in distributional sense $V_N * u \to u$ for $N \to \infty$.

[14] In many applications this assumption can be weakened by assuming bounded moments of V.

Intermediate Levels: The proofs given in Chapters 2, 3 and 4 are based on the concept of a so-called *intermediate level:*
Note that we illustrate the idea of an *intermediate level* in a basic setting (particle level according to (1.18)), see Chapters 2, 3 and 4 for different settings. The underlying idea is the following:

- For the first step, we '*ignore*' the dependence of $\eta(N)$ on N, take $\eta > 0$ fixed and define the interaction kernel as $V^\eta(x) := \eta^{-d}V(\eta^{-1}x)$. Hence, the interaction kernel now does not depend on N. By looking at the interacting particle system (1.17) but with $V_N = V^\eta$, this corresponds to weakly interacting particles with interaction kernel V^η. By coupling methods, we know that (see (1.8) - (1.10)) for fixed $\eta > 0$ our particle system converges to the nonlinear process (which now depends on $\eta > 0$)

  $$\mathrm{d}\overline{X}_i^\eta(t) = \Big[V^\eta * u_\eta(t, \overline{X}_i^\eta(t))\Big]\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t) \qquad i = 1, \dots, N, \tag{1.20}$$

  where u_η solves the non-local PDE $\partial_t u_\eta = \sigma\Delta u_\eta - \mathrm{div}((V^\eta * u_\eta)u_\eta)$. Here, we want to remind the reader that particles $\overline{X}_i^\eta$ are independent and identically distributed with density function u_η.

  System (1.20) will be called *intermediate level* throughout this thesis. At the end of this step, we need to establish estimates for the difference between $X_i^N(t)$ and $\overline{X}_i^\eta(t)$ in a suitable norm, like convergence in expectation as in Theorem 1.7. However, the mathematical difficulty is that in comparison to Theorem 1.7, we have to keep track of the dependence of η, since in the last step of this guideline we want to let $\eta \to 0$.

- Second, we compare the non-linear process (1.20) with the following *local non-linear process*

  $$\mathrm{d}\overline{X}_i(t) = \Big[u(t, \overline{X}_i(t))\Big]\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t) \qquad i = 1, \dots, N, \tag{1.21}$$

  where u (formally) solves the local partial differential equation

  $$\partial_t u = \sigma\Delta u - \mathrm{div}(u^2)^{15}. \tag{1.22}$$

  Particle system (1.21) corresponds to the *macroscopic level* and the local partial differential equation which we want to derive. Note that the macroscopic level (1.21) as well as intermediate level (1.20) are not interacting particle systems but already independent from the other particles in the two systems. At the end of this second step we wish to derive estimates of the differences $|\overline{X}_i - \overline{X}_i^\eta|$ for $i = 1, \dots, N$ depending on η. Usually, in this step the main difficulty are analytical error estimates between the solution u_η to the non-local PDE and u, the solution to the local PDE.

- Finally, we compare the particle dynamics

  $$\mathrm{d}X_i^N(t) = \frac{1}{N}\sum_{j=1}^N \eta(N)^{-d}V(\eta(N)^{-1}(X_i^N(t) - X_j^N(t)))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \quad i = 1, \dots, N.$$

[15] In one dimension this is a Burger's type equation, [113] equation (2.3).

with the local macroscopic dynamics (1.21) by using estimates between the particle dynamics and the intermediate system (by carefully taking track on the dependence on $\eta > 0$) and between the intermediate system and the local macroscopic dynamics (1.21). By letting $\eta \to 0$ and $N \to \infty$ at the same time (with a connection between η and N) the desired propagation of chaos result towards a *local* partial differential equation is obtained.

We illustrate the concept with the following Figure 1.2.2:

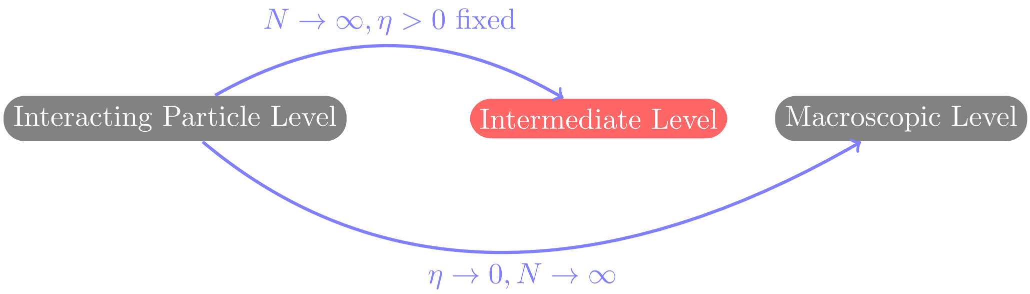

Figure 1.1: Schematic idea of an intermediate level

Interaction radius: We also want to remark by assuming that V has compact support on $B_1(0)$, the scaled interaction potential V_N has support on the ball with radius $\eta(N)$. Hence, the parameter η can be also interpreted as *interaction radius* of the particles. In the limit, as the number of particles converges to infinity, this interaction radius converges to zero leading to a *local* macroscopic level.

The idea of using an *intermediate level* in the moderate regime and hence exploiting well-known techniques for mean-field limits in the weak regime is not newly invented in this thesis, see [65] for instance, where the intermediate level is called *mollified version*. However, especially in Chapters 2 and 3 careful estimates on the non-local PDE level(s) are incorporated in the estimates between the intermediate level and the macroscopic level as well as between the intermediate level and the microscopic level, which shows the value of classical PDE theory in the context of (stochastic) mean-field limits. To summarise, in this section we have seen that by using coupling techniques in the moderate regime, on the PDE level estimates of *local* and *non-local* partial differential equations are of great importance.

1.3 Outline of this thesis

The mathematical results of this thesis are structured in three parts. In the following, we shortly illustrate the main goal of each of the chapters and provide associated key-words; a more detailed description is given in the Sections 1.3.1- 1.3.3:

- In Chapter 2, we derive the well-known so-called *SKT model* – which is a multi-species cross-diffusion model – from a moderate stochastic interaction model. The results of

this chapter have been already published in the *Journal of Nonlinear Science*; [25].
key-words: *cross-diffusion, non-local cross-diffusion models, moderate regime, logarithmic scaling*

Declaration of authorship: The topic of this article was brought to me by Ansgar Jüngel, Esther Daus and Li Chen. Determining the shape of the particle model and writing the proofs was mainly my work which came along together with fruitful discussions and two research stays with my co-authors who helped me throughout this project with valuable advice.

- Chapter 3 is devoted to the derivation of a fractional porous media equation from a stochastic interacting particle model. The results of this chapter have been already published in *Communications in Partial Differential Equations*; [30].
key words: *fractional diffusion, nonlocal porous media equation, moderate regime, smoothed singular kernel, vanishing diffusion*

  Declaration of authorship: The topic of this article was brought to me by Li Chen. My expertise at that time was clearly on designing the particle models and writing the proofs concerning the mean-field derivation part. This was done by many discussions during the pandemic via Zoom with Li Chen. Additionally, I contributed with discussions and did proof-reading of the sections concerning estimates of the solution(s) of the partial differential equation(s) considered in this article and corrected mistakes within these sections. Finalizing the manuscript was split equally amongst the authors Li Chen, Ansgar Jüngel, Nicola Zamponi and myself.

- In the last chapter of this thesis (Chapter 4), we extend techniques used by K. Oelschläger [90] for a fluctuation result in the setting of repulsive particles. In Chapter 4 we show that the essential L^2 convergence with rate $N^{-1/2-\varepsilon}$ used by Oelschläger also holds in case of aggregating particles, given that propagation of chaos holds in probability. This is an ongoing work together with Ansgar Jüngel and Li Chen; in the appendix (Section 4.A) a result concerning propagation of chaos in probability with a singular kernel of Coulomb-type is presented which is close to submission and part of a joint work with Li Chen, Veniamin Gvozdik and Yue Li; [28].
key words: *aggregating particles, L^2-convergence, smoothed empirical densities, convergence in probability*

  Declaration of authorship: Li Chen and myself already discussed the topic of fluctuations around the mean-field limit in the beginning of my PhD. The idea of using convergence results in probability – inspired by Peter Pickl and co-workers – was brought to me by Li Chen. Writing the proofs of this chapter was my work which would not have been possible without the fruitful discussions and advice of my co-workers. The proof of the result in the appendix of this chapter was done by myself based on many calculations I did together with Li Chen in Mannheim on the whiteboards in her office.

1.3.1 Cross-diffusion system of SKT-type, Chapter 2

In this chapter, we give a rigorous proof that the following cross-diffusion system of SKT-type, which is a well-known model in population dynamics for n interacting species,

$$\partial_t u_i = \operatorname{div}(u_i \nabla U_i) + \Delta\big(\sigma u_i + u_i \sum_{j=1}^{n} f(a_{ij}u_j)\big) \quad \text{on } \mathbb{R}^d, \quad i = 1, \dots, n, \tag{1.23}$$

with $u_i(0) = u_0$, $\sigma > 0, a_{ij} \geq 0$, can be derived in the moderate regime for $N \to \infty$ from a particle system of size nN. In this microscopic system of mean-field type, the interaction between particles of the same and of different species is modelled via interaction kernels in the diffusion part of the stochastic differential equation (SDE). Equation (1.23) was first introduced by Shigesada, Kawasaki and Teramoto [104] in the late 1970s to model interacting insects under the effects of inter- and intra-species population pressures. The functions U_i model environmental potentials, which are assumed to be 'dispersive' and the non-linearity $f \geq 0$ has to be at least *locally* Lipschitz continuous.
The approach used in this part of the thesis is based on two articles in the regime of *moderately interacting particles* - the work [91] by K. Oelschläger for single-species models, which was later extended by Jourdain and Méléard [65] by also considering interactions in the diffusion part of the particle system.
As explained in Section 1.2.2, in the regime of moderate interaction, the interaction kernel depends on the number of particles via the *interaction radius* $\eta = \eta(N)$ (support of the kernel) and approximates a Dirac distribution for $\eta \to 0$ for $N \to \infty$. In our case, we choose a logarithmic connection between N and η, namely $\eta \sim C\log(N)^{-1/(2d+2)}$, and show the (strong) convergence of the second moment of the particles towards the solution of system (1.23), which forms the main result of this chapter:
Let $X_{k,i}^{N,\eta}$ denote the k-th particle of the i-th species of the microscopic level and $\widehat{X}_{k,i}$ the k-th particle of the i-th species of the corresponding macroscopic particle systems (obtained by coupling methods). Then, the following holds true:

Main Theorem (Chapter 2, Theorem 2.5)**.** Under suitable conditions on the initial datum u_0 the convergence

$$\sup_{k=1,\dots,N} \mathbb{E}\Bigg(\sum_{i=1}^{n} \sup_{0<s<T} \big|(X_{k,i}^{N,\eta} - \widehat{X}_{k,i})(s)\big|^2\Bigg) \leq C(T, n, \sigma_1, \dots, \sigma_n)\eta^{2(1-\alpha)} \to 0 \tag{1.24}$$

holds for $\eta \to 0$ and $N \to \infty$, where $\alpha > 0$ is an approximation parameter which vanishes for globally Lipschitz continuous functions f.

In order to prove the mean-field limit rigorously, we use the following non-local system

$$\partial_t u_{\eta,i} = \operatorname{div}(u_{\eta,i} \nabla U_i) + \Delta\Bigg(\sigma_i u_{\eta,i} + u_{\eta,i} \sum_{j=1}^{n} f_\eta(B_{ij}^\eta * u_{\eta,j})\Bigg) \text{ on } \mathbb{R}^d, \quad i = 1, \dots, n, \tag{1.25}$$

for fixed $\eta > 0$ as an *intermediate system*, where f_η is a suitable approximation of f and B_{ij}^η denotes the interaction kernel with radius $\eta > 0$ between species i and j.

This non-local system, whose existence and uniqueness analysis is also included in Chapter 2, can be viewed as mean-field approximation for fixed η in the *weakly interacting regime*. Crucial to the particle derivation in this chapter are $L^\infty((0,T)\times\mathbb{R}^d)$-estimates of the PDE solutions (and their derivatives) to the local system (1.23) and the non-local system (1.25), which are given in Chapter 2, Theorems 2.2 and 2.3, respectively. Those estimates rely on the fact that we derive solutions of the PDEs involved in this derivation in $L^\infty(0,T;H^s(\mathbb{R}^d))$ with $s > d/2+1$.
The novelty of this derivation is threefold: First, it is – to the best of the author's knowledge – the first rigorous derivation of the SKT system from stochastic interacting particles of mean-field type. Second, we extend the concept of moderate interactions for multiple species by considering interaction in the diffusion part of the particle system. Third, we allow for *non-globally* Lipschitz interactions by the non-linearity f. The trade-off of using a non-globally Lipschitz function f is a slower convergence rate than in [65]. However, for globally Lipschitz interactions, we gain exactly the same convergence rate as in the single species case by Jourdain and Méléard [65].
At the end of this chapter, numerical experiments are shown where we compare the results concerning segregation behaviour with the cross-diffusion particle system used in [26].

Outlook: Interesting follow-up questions to this chapter can be

- Since the convergence rate in (1.24) is only logarithmic in N, it is an interesting question whether this can be improved; possibly with a different notion of convergence?
- Similar to the question above: Can we allow for an algebraic scaling of $\eta(N)$ in N?
- Does a fluctuation theorem (in the spirit of Section 1.1.3) hold for this multi-species model?

1.3.2 Porous-media equation with fractional diffusion, Chapter 3

In the third chapter of this thesis, the so-called *porous-medium equation with fractional diffusion*

$$\partial_t\rho = \operatorname{div}(\rho\nabla P(\rho)), \quad P(\rho) = (-\Delta)^{-s}f(\rho) \quad \text{on } \mathbb{R}^d \;\; d\geq 2, \tag{1.26}$$

where f is a non-decreasing function with $f(0)=0$ and for $0<s<1$ we let $(-\Delta)^{-s}u = \mathcal{K}*u$ with the singular kernel $\mathcal{K}(x) = C(d,s)|x|^{2s-d}$, is studied.
It is shown rigorously that (1.26) can be derived from a stochastic interacting particle system using mean-field limit techniques by showing a propagation of chaos result with moderately interacting particles. The two main difficulties of the derivation of this non-local porous medium equation are the singularity of the convolution kernel $\mathcal{K}$ and that we allow for a large class of (possible non-globally Lipschitz continuous) functions $f(\rho)$. A guiding example would be $f(\rho)=\rho^\alpha$ for $\alpha\geq 1$. Both of them can be overcome by using suitable approximating sequences on the particle level which are specifically tailored for the structure of equation (1.26). We use the following regularisation parameters

- $\beta > 0$: Using ideas of the general concept of moderately interacting particles, we define a interaction kernel via $\mathcal{W}_\beta(x) = \beta^{-d}\mathcal{W}_1(|x|/\beta)$ for a smooth, symmetric, non-negative and normalised function $\mathcal{W}_1$ on $\mathbb{R}^d$, where $\beta = \beta(N)$ depends on the number of particles in a logarithmic way, i.e. $\beta \sim (\log(N))^{-\mu}$ for some $\mu > 0$. At this point, we want to remark that caused by the special structure of the parabolic-elliptic system (1.26), $\beta > 0$ does not take the role of an interaction radius, like $\eta(N) > 0$ in Section 1.2.2, but still has a similar idea since

$$\mathcal{W}_\beta \to \delta_0$$

in distribution for $N \to \infty$ which implies $\beta \to 0$. Because of this we still call the regime used in this chapter *moderately interacting regime*.

- $\zeta > 0$: On the microscopic and intermediate particle level, we use a combination between cut-off and convolution techniques in order to approximate $\mathcal{K}$ by a sequence of smooth and compactly supported kernels $\mathcal{K}_\zeta$. In the limit, we let $\zeta \to 0$ where we connect ζ with the number of particles N in an algebraic way, i.e. $\zeta \sim N^{-\nu}$ for some $1/4 > \nu > 0$. The concrete value of ν depends on the choice of the parameter $s \in (0, 1)$.

- $\sigma > 0$: On the one hand, this parameter is used in order to add additional diffusion to the system (1.26). On the other hand, we also use it in order to approximate the non-linearity f by a sequence of smooth functions. In contrast to ζ and β, we do not connect σ with the number of particles.

Similarly to Chapter 2, we use the concept of *intermediate levels* represented by non-local equations, which is important when we deal with moderate interactions. Different to the cross-diffusion case, equation (1.26) does not contain pure diffusion. Therefore, we need one additional stochastic level, which leads to the following hierarchy of SDE levels:

I. **Microscopic Level:** On this level, we consider $N \in \mathbb{N}$ interacting particles – denoted by X_i^N – with all regularisation parameters strictly positive, i.e. $\zeta > 0, \beta > 0, \sigma > 0$.

II. **Intermediate Level:** This (technical) level follows the general approach of *intermediate systems* (see Section 1.2.2 for a general introduction), by letting $N \to \infty$, but 'ignoring' the dependence of $\zeta > 0$ and $\beta > 0$ and keeping them fixed. This level is represented by an uncoupled system of SDEs, where all particles have a common density function, which solves the following non-linear PDE

$$\partial_t \rho_{\sigma,\beta,\zeta} = \sigma \Delta \rho_{\sigma,\beta,\zeta} + \operatorname{div}\left(\rho_{\sigma,\beta,\zeta} \nabla \mathcal{K}_\zeta * f_\sigma(\mathcal{W}_\beta * \rho_{\sigma,\beta,\zeta})\right) \quad \text{in } \mathbb{R}^d. \tag{1.27}$$

III. **Macroscopic Level** (with additional diffusion):

$$\partial_t \rho_\sigma = \sigma \Delta \rho_\sigma + \operatorname{div}(\rho_\sigma \nabla (-\Delta)^{-s}(f_\sigma(\rho_\sigma))) \qquad \text{in } \mathbb{R}^d, \tag{1.28}$$

where we recall that the approximation of f is denoted by f_σ and also diffusion is added to the system.

The limit $\sigma \to 0$ is performed on the PDE level and not via coupling methods. We are able to show that there exists a subsequence such that $\rho_\sigma \to \rho$ strongly in $L^1(\mathbb{R}^d \times (0,T))$, where ρ solves (1.26). We remark that, since we can not show uniqueness of (1.26) - which is still an open question - the propagation of chaos result only holds up to a subsequence:

Main Theorem (Chapter 3, Theorem 3.2)**.** For suitable scaling of β and ζ with respect to N, the following holds: Let $\mathrm{P}^k_{N,\sigma,\beta,\zeta}(t)$ be the joint distribution of $(X_1^N(t),\dots,X_k^N(t))$ for $k \geq 1$ and $t \in (0,T)$. Then there exists a subsequence in σ such that

$$\lim_{\sigma\to 0} \lim_{\substack{N\to\infty \\ (\beta,\zeta)\to 0}} \mathrm{P}^k_{N,\sigma,\beta,\zeta}(t) = \rho^{\otimes k}(t),$$

where the limit is understood in the weak sense and is locally uniform in time and ρ solves the fractional porous media equation (1.26).

In this chapter, we also present existence results to the equations (1.26), (1.28) and (1.27), including error estimates between the different PDE approximations, see Theorem 3.1, Proposition 3.14 and Proposition 3.4 .

Outlook: Except for the open question of uniqueness of the fractional diffusion equation, regarding the particle derivation the following questions can be of interest:

- When it comes to singular kernels for mean-field type derivations of partial differential equations, a natural question would be whether we can allow the singular kernel to be used on the particle level, see for instance the recent frameworks developed by Jabin and Wang [61], [63] and Serfaty and Duerinckx [43]. Is the regularisation $\mathcal{K}_\zeta$ necessary in order to derive a propagation of chaos result?
- Another natural question to ask is whether we are able to derive better rates of convergence between the particle levels and can we derive a fluctuation result for singular kernels of Riesz type which are used in this chapter?

1.3.3 Aggregation-diffusion equation, Chapter 4

In the last chapter of this thesis, we consider the following local diffusion model with aggregation for $\kappa = \pm 1$

$$\partial_t u = \sigma \Delta u - \kappa \operatorname{div}(u \nabla u) \quad \text{for } t > 0, \quad u(0) = u_0 \quad \text{in } \mathbb{R}^d. \tag{1.29}$$

It is well-known [91, 27] that – under suitable assumptions on the initial data and the interaction kernel – (1.29) can be derived from a system of interacting particles in the moderate regime:

$$\mathrm{d}X_i^{N,\eta}(t) = \frac{\kappa}{N}\sum_{j=1}^N \nabla V^\eta(X_i^{N,\eta}(t) - X_j^{N,\eta}(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \tag{1.30}$$
$$X_i^{N,\eta}(0) = \zeta_i \quad \text{in } \mathbb{R}^d, \quad i = 1,\dots,N,$$

where the parameter κ models the type of the dynamics: $\kappa = -1$ corresponds to repulsive interactions and $\kappa = 1$ to aggregating particles. As usual in the moderate regime $V^\eta(x) := \eta^{-d}V(|x|/\eta)$, for a smooth, non-negative and normalized function V, where in the limit $N \to \infty$, the parameter $\eta > 0$ is connected to $N \in \mathbb{N}$, such that $\eta \to 0$ if $N \to \infty$.
As mentioned in Section 1.1.3, it is of particular interest to study so-called *fluctuations around the mean-field limit* since by approximating a stochastic interacting particle system through a deterministic partial differential equation some information induced by the stochasticity of the interaction system gets lost. In the setting of moderately interacting particles, we are interested in the *intermediate fluctuations*, where we do not compare the empirical measure $\mu_{N,\eta}$ associated with (1.30) with the local PDE solution (1.29) but with the intermediate solution $\bar{u}^\eta$, which solves for fixed $\eta > 0$:

$$\partial_t \bar{u}^\eta = \sigma \Delta \bar{u}^\eta - \kappa \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta), \; t > 0, \quad \bar{u}^\eta(0) = u_0 \text{ in } \mathbb{R}^d. \tag{1.31}$$

Following techniques developed by K. Oelschläger for repulsive particles, in order to study the limiting behaviour of the intermediate fluctuations, we show an $L^2(\mathbb{R}^d)$ convergence result for the *smoothed* empirical measure towards the *smoothed* intermediate solution

$$f^{N,\eta}(t,x) := (\mu_{N,\eta}(t) * Z^\eta)(x), \qquad g^\eta(t,x) := (\bar{u}^\eta(t) * Z^\eta)(x),$$

where V^η is assumed to be 'convolutional square', i.e. $V^\eta = Z^\eta * Z^\eta$. This result holds also in case of aggregating particles, which are not included in [91]. For more details on the connection of the $L^2(\mathbb{R}^d)$ convergence result and the limiting behaviour of the intermediate fluctuations, we refer to Section 4.1.2.

The main result of this chapter is an $L^2(\mathbb{R}^d)$ convergence result of the smoothed empirical measure towards the smoothed intermediate solution with rate $N^{-1/-\varepsilon}$. The rate of convergence plays an important role in the study of the fluctuation behaviour; see Section 4.1.2. The theorem reads as follows:

Main Theorem (Chapter 4, Theorem 4.1)**.** Let $\eta = N^{-\beta}$, where $0 < \beta < 1/(10d + 12)$. Then, for any $T > 0$, there exists $\varepsilon > 0$ and a constant $C(\beta, d, T) > 0$ such that for sufficiently large number of particles $N > 0$,

$$\mathbb{E}\left(\sup_{0<t<T} \|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 + \sigma \int_0^T \|\nabla (f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 \mathrm{d}t \right) \leq C(\beta, d, T) N^{-1/2-\varepsilon}.$$

This results holds under suitable assumptions on the initial condition u_0 and by assuming that propagation of chaos holds for $\eta = N^{-\beta}$ at least in probability in the following way: For every $\gamma > 0$ and $T > 0$ there exists a constant $C(\gamma, T)$ such that

$$\mathbb{P}\Big(\max_{i=1,\ldots,N} |X_i^{N,\eta}(t) - \overline{X}_i^\eta(t)| > N^{-\alpha} \Big) \leq C(\gamma, T) N^{-\gamma}, \tag{1.32}$$

for a suitable cut-off rate $\alpha > 0$. In Section 4.A, we discuss this assumption and give a rigorous proof for convergence in probability for interaction kernels approximating the singular Coulomb-kernel. Verifying this result rigorously in a more general setting – for

instance for interaction kernels approximating a Dirac distribution – is an open research question.
In order to give a proof of the main theorem of this section, we develop a new technique, since the result in [91] only works for repulsive particles and can not be extended in a straightforward way for aggregating particles. The idea is the following:
In comparison to the repulsive case, the following term ($\langle \cdot, \cdot \rangle$ denotes the dual bracket)

$$M(t) = \kappa \int_0^t \langle \mu_{N,\eta}, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \rangle \mathrm{d}s$$

can not be ignored due to the positive sign in case $\kappa = 1$ (aggregating case). Hence, we have to estimate it directly. Since we already know that $(\mu_{N,\eta} - \bar{u}^\eta) \to 0$ weakly, [91, 27], where $\bar{u}^\eta$ solves the non-local equation (1.31), inspired by PDE techniques a first attempt would be (similar as for pure PDE estimates) to 'replace' $\mu_{N,\eta}$ with $\bar{u}^\eta$ and use that $\sup_{0<t<T} \|\bar{u}^\eta(t)\|_{L^\infty(\mathbb{R}^d)} < \sigma$ for *small* initial data, which would allow us to absorb this term by diffusion terms.
However, this attempt does not work directly since the remainder $\sup_{0<t<T} \|\mu_{N,\eta}(t) - \bar{u}^\eta(t)\|_{L^\infty(\mathbb{R}^d)}$ can not be bounded uniformly in $\omega \in \Omega$. Therefore, we add and subtract the *empirical measure* of the intermediate particle system (where all particles are independent with common density function $\bar{u}^\eta(t)$) denoted by $\overline{\mu}_{N,\eta}$. This allows us to split the difference $\mu_{N,\eta} - \bar{u}^\eta$ into

(i) *a mean-field estimate* $\mu_{N,\eta} - \overline{\mu}_{N,\eta}$, and

(ii) *a law of large numbers estimate* $\overline{\mu}_{N,\eta} - \bar{u}^\eta$ (since the intermediate particles are already independent).

By exploiting our assumption of convergence in probability as well as a law-of-large numbers estimate in probability (see Chapter 4, Lemma 4.2), we do not estimate $\sup_{0<t<T} \|\mu_{N,\eta}(t) - \overline{\mu}_{N,\eta}(t)\|_{L^\infty(\mathbb{R}^d)}$ and $\sup_{0<t<T} \|\overline{\mu}_{N,\eta}(t) - \bar{u}^\eta(t)\|_{L^\infty(\mathbb{R}^d)}$ directly, but we allow for a set $\mathcal{B} \subset \Omega$ where the respective difference is 'large'. By exploiting the mean-field convergence and the law-of-large numbers estimate we can conclude that the probability of $\mathcal{B}$ is small, which illustrates the main idea of our technique.
Because of an error term we make by manipulating the dual bracket $\langle \cdot, \cdot \rangle$ and the convolution with Z^η, the estimates are very delicate. For more details we refer the reader to Chapter 4.

Outlook: Inspired by the results of Chapter 4, the following open research questions could serve as starting points for future research:

- First, to prove the assumed propagation of chaos property for $\eta = N^{-\beta}$ in case of aggregating particles would provide a more complete picture of the intermediate fluctuations in case of aggregating particles, we comment on the current technical challenges in Section 4.A.2.
- Second, it is interesting to note that the $L^2(\mathbb{R}^d)$ convergence shown in Chapter 4 can be used in order to prove a strong mean-field limit in $L^1(\mathbb{R}^d)$ norm, which has been

recently shown as a consequence of relative entropy (or modulated free energy) estimates for certain interaction systems of Coulomb-type, [63, 10, 43]. Future research will be concerned with the connection between the before mentioned convergence types.

- Further investigations on the fluctuation behaviour of particle system (1.30) are still open. The main theorem of this chapter will serve as an important step towards a better understanding of the limiting behaviour. Additionally, analysing the limiting SPDE structure will be of particular interest and a future research goal.

- Since Section 4.A.1 provides a convergence result in probability, a natural question would be whether it is possible to extend the newly developed techniques in this part of the thesis towards Coulomb interactions in order to show an $L^2(\mathbb{R}^d)$ convergence result and consequently a fluctuation theorem for Keller-Segel-type equations. This would fill a significant gap in literature fo the study of Keller-Segel systems.

- Prospectively, results on fluctuations around the mean-field limit for cross-diffusion models are of particular interest. Since aggregation effects play an important role in cross-diffusion settings, the results of this thesis can serve as an important first step.

Once we accept our limits, we go beyond them.

— Albert Einstein[1]

2 Rigorous derivation of cross-diffusion systems by a moderate model

Contents

This chapter is taken from the article

[25] Li Chen, Esther S. Daus, Alexandra Holzinger, and Ansgar Jüngel. Rigorous derivation of population cross-diffusion systems from moderately interacting particle systems. *Journal of Nonlinear Science*, 31, 1-38, 2021.

2.1 Problem setting

The aim of this chapter is to derive the population cross-diffusion system of Shigesada, Kawasaki, and Teramoto [104] from a stochastic, moderately interacting particle system in a mean-field-type limit. More precisely, we derive the system of equations

$$\partial_t u_i = \operatorname{div}(u_i \nabla U_i) + \Delta\bigg(\sigma_i u_i + u_i \sum_{j=1}^n f(a_{ij}u_j)\bigg), \quad u_i(0) = u_{0,i} \quad \text{in } \mathbb{R}^d,\ t > 0, \tag{2.1}$$

where $i = 1, \dots, n$ is the species index, $d \geq 1$ the space dimension, $u = (u_1, \dots, u_n)$ is the vector of population densities, and $U_i = U_i(x)$ are given environmental potentials. The parameters $\sigma_i > 0$ are the constant diffusion coefficients in the stochastic system, and $a_{ij} \geq 0$ are limiting values of the interaction potentials. In the linear case $f(s) = s$, we obtain the population model in [104]. System (2.1) with nonlinear functions f have also been studied in the mathematical literature; see, e.g., [32, 40, 73]. Such systems can be

[1]Attributed to A. Einstein.

formally derived from random walks on a lattice, where the nonlinearity originates from the transition rates in the random-walk model [119, Appendix A]. Assuming that the transition rates depend in a nonlinear way on the densities leads to equations similar to (2.1). We assume that f is smooth but possibly *not* globally Lipschitz continuous (including power functions). Our results are valid for functions f_i depending on the species type, but we choose the same function for all species to simplify the presentation.
This chapter extends the many-particle limit of [26] leading to the cross-diffusion system

$$\partial_t u_i = \operatorname{div}\left(\sigma_i \nabla u_i + \sum_{j=1}^{n} a_{ij} u_i \nabla u_j\right) \quad \text{in } \mathbb{R}^d,\ t > 0,\ i = 1, \dots, n, \tag{2.2}$$

which differs from (2.1) by the drift term, the nonlinear function f, and the diffusion term

$$\operatorname{div} \sum_{j=1}^{n} a_{ij} u_j \nabla u_i.$$

System (2.2) is the mean-field limit of the particle system for N individuals

$$\begin{aligned} \mathrm{d}Y_{k,i}^{N,\eta} &= -\sum_{j=1}^{n} \frac{1}{N} \sum_{\ell=1}^{N} \nabla B_{ij}^{\eta}\big(Y_{k,i}^{N,\eta} - Y_{\ell,j}^{N,\eta}\big)\mathrm{d}t + \sqrt{2\sigma_i}\,\mathrm{d}W_i^k(t), \\ Y_{k,i}^{N,\eta}(0) &= \xi_i^k, \quad i = 1, \dots, n,\ k = 1, \dots, N, \end{aligned} \tag{2.3}$$

where $(W_i^k(t))_{t\geq 0}$ are d-dimensional Brownian motions and $\xi_i^1, \dots, \xi_i^N$ are independent and identically distributed (iid) random variables with the common probability density function $u_{0,i}$. The functions

$$B_{ij}^{\eta}(x) = \eta^{-d} B_{ij}\left(\frac{|x|}{\eta}\right), \quad x \in \mathbb{R}^d, \tag{2.4}$$

are interaction potentials regularizing the delta distribution δ_0, i.e. $B_{ij}^{\eta} \to a_{ij}\delta_0$ as $\eta \to 0$ in the sense of distributions.
System (2.1) is derived from an interacting particle system for n species with particle numbers $N_1, \dots, N_n$, moving in the whole space $\mathbb{R}^d$. To simplify, we set $N = N_i$ for all $i = 1, \dots, n$. The key idea in this chapter is to consider interacting diffusion coefficients:

$$\begin{aligned} \mathrm{d}X_{k,i}^{N,\eta} &= -\nabla U_i(X_{k,i}^{N,\eta})\mathrm{d}t + \left(2\sigma_i + 2\sum_{j=1}^{n} f_\eta\left(\frac{1}{N}\sum_{\substack{\ell=1 \\ (\ell,j)\neq(k,i)}}^{N} B_{ij}^{\eta}(X_{k,i}^{N,\eta} - X_{\ell,j}^{N,\eta})\right)\right)^{1/2} \mathrm{d}W_i^k(t), \\ X_{k,i}^{N,\eta}(0) &= \xi_i^k, \quad i = 1, \dots, n,\ k = 1, \dots, N, \end{aligned} \tag{2.5}$$

where f_η is a globally Lipschitz continuous approximation of f with a Lipschitz constant smaller or equal than $\eta^{-\alpha}$ for some small $\alpha > 0$. In view of (2.4), we can interpret the scaling parameter η as the interaction radius of each particle.
Equations (2.1) are derived from system (2.5) in the limit $N \to \infty$, $\eta \to 0$, with the scaling relation between η and N given in (2.9) below. First, for fixed $\eta > 0$, we perform a classical

mean-field limit from (2.5) to the following auxiliary *intermediate* system:

$$\begin{aligned} \mathrm{d}\overline{X}^\eta_{k,i} &= -\nabla U_i(\overline{X}^\eta_{k,i})\mathrm{d}t + \Big(2\sigma_i + 2\sum_{j=1}^n f_\eta\big(B^\eta_{ij} * u_{\eta,j}(\overline{X}^\eta_{k,i})\big)\Big)^{1/2}\mathrm{d}W^k_i(t), \\ \overline{X}^\eta_{k,i}(0) &= \xi^k_i, \quad i=1,\dots,n,\ k=1,\dots,N, \end{aligned} \tag{2.6}$$

where we set $u_{\eta,j}(\overline{X}^\eta_{k,i}) = u_{\eta,j}(t,\overline{X}^\eta_{k,i}(t))$ for $j=1,\dots,n$. The function $u_{\eta,j}$ satisfies the nonlocal cross-diffusion system

$$\begin{aligned} \partial_t u_{\eta,i} &= \operatorname{div}(u_{\eta,i}\nabla U_i) + \Delta\Big(\sigma_i u_{\eta,i} + u_{\eta,i}\sum_{j=1}^n f_\eta(B^\eta_{ij} * u_{\eta,j})\Big), \\ u_{\eta,i}(0) &= u^0_i \text{ in } \mathbb{R}^d, \quad i=1,\dots,n, \end{aligned} \tag{2.7}$$

and will be later identified as the probability density function of $\overline{X}^\eta_{k,i}$. Note that we consider N independent copies $\overline{X}^\eta_{k,i}$, $k=1,\dots,N$, and the intermediate system depends on k only through the initial datum.
Then, passing to the limit $N\to\infty$, $\eta\to 0$ in (2.5) leads to the *macroscopic system*

$$\begin{aligned} \mathrm{d}\widehat{X}_{k,i} &= -\nabla U_i(\widehat{X}_{k,i})\mathrm{d}t + \Big(2\sigma_i + 2\sum_{j=1}^n f(a_{ij}u_j(\widehat{X}_{k,i}))\Big)^{1/2}\mathrm{d}W^k_i(t), \\ \widehat{X}^\eta_{k,i}(0) &= \xi^k_i, \quad i=1,\dots,n,\ k=1,\dots,N, \end{aligned} \tag{2.8}$$

where the functions u_i satisfy (2.1) and can be identified as the probability density functions of $\widehat{X}_{k,i}$. In this limit, we assume that there exists $\delta>0$, depending on n, $\min_i\sigma_i$, and T, such that

$$\eta^{-2(d+1+\alpha)} \le \delta\log N \tag{2.9}$$

holds, where $\alpha\ge 0$ depends on the Lipschitz condition of f, see Assumption ($A4$) below, and that the function f and its derivatives or, alternatively the initial data, are sufficiently small (see Section 2.2 for details). The main result in this chapter is the error estimate

$$\sup_{k=1,\dots,N}\mathbb{E}\Big(\sum_{i=1}^n\sup_{0<s<T}\big|X^{N,\eta}_{k,i}(s)-\widehat{X}_{k,i}(s)\big|^2\Big) \le C(T)\eta^{2(1-\alpha)}. \tag{2.10}$$

We prove this estimate for the potential $U_i(x) = -\frac12|x|^2$, but more general functions are possible; see Remark 2.1. Note that estimate (2.10) implies propagation of chaos; see Remark 2.6. In the case $\alpha=0$, our scaling (2.9) for the multi-species case recovers the result in [65], where a single-species, moderately interacting particle system with interaction in the diffusion part was considered. Our strategy is similar to that one of [65] (and based on ideas of Oelschläger [94]). Since we allow for locally Lipschitz continuous nonlinearities only, we obtain a smaller convergence rate compared to [65], which in fact is natural, since we approximate the nonlinearity with functions having a Lipschitz constant of order $\eta^{-\alpha}$. A difference to [65] is that the authors assume that the diffusion matrix in the stochastic part is positive definite. We do not suppose such a condition, but we need a smallness condition on the nonlinearity for the existence proofs of systems (2.1) and (2.7).

Since the underlying method of this chapter are moderately interacting particles and intermediate systems obtained through *coupling methods*, we refer the reader to Section 1.2.2 for an introduction on this matter for single species models.
Next, we present a brief overview on the existing literature concerning mean-field limits and moderately interacting many-particle limits in the context of diffusion equations which are of particular interest for this chapter. For an introductory overview we refer the reader to the introduction of this thesis. Mean-field limits from stochastic differential equations have been investigated since the 1980s; see the reviews [54, 62] and the classical works by Sznitman [112, 113]. Oelschläger proved that in the many-particle limit, weakly interacting stochastic particle systems converge to a deterministic nonlinear process [92]. Later, he generalized his approach for systems of reaction-diffusion equations [94] and porous-medium-type equations with quadratic diffusion [93], by using moderately interacting particle systems. We also refer to the recent work [27], which also includes numerical simulations. As already mentioned, moderate interactions in stochastic particle system with nonlinear diffusion coefficients were investigated for the first time in [65]. Later, Stevens derived the chemotaxis model from a many-particle system [109]. Further works concern the mean-field limit leading to reaction-diffusion equations with nonlocal terms [59], the hydrodynamic limit in a two-component system of Brownian motions to the cross-diffusion Maxwell–Stefan equations [102], and the large population limit of point measure-valued Markov processes to nonlocal Lotka–Volterra systems with cross diffusion [51]. The latter model is similar to the nonlocal system (2.7). The limit from the nonlocal to the local diffusion system was shown in [87] but only for triangular diffusion matrices. The many-particle limit from a particle system driven by Lévy noise to a fractional cross-diffusion system related to (2.2) was recently shown in [38]. Furthermore, the population system (2.1) was derived in [36] from a time-continuous Markov chain model using the BBGKY hierarchy. The main result of this chapter presents, up to our knowledge, the first rigorous derivation of the Shigesada–Kawasaki–Teramoto (SKT) model (2.1) from a stochastic particle system in the moderate many-particle limit.
Porous-medium-type equations can be derived from stochastic interacting particle systems by assuming interactions in the drift term [47] or in the diffusion term [65]. We allow for interactions in the diffusion part but in a multi-species setting. The paper [51] is concerned with a multi-species framework too, but the authors assume bounded Lipschitz continuous interaction potentials and derive a nonlocal cross-diffusion system only. We are able to relax the assumptions and derive the local cross-diffusion system (2.1).
Compared to the work [36], we take the limits $N \to \infty$, $\eta \to 0$ simultaneously. However, our approach also implies the two-step limit. Indeed, we can first perform the limit $N \to \infty$ for fixed $\eta > 0$ and afterwards the limit $\eta \to 0$ on the PDE level; see Lemma 2.9 and Theorem 2.3. The simultaneous limit $N \to \infty$, $\eta \to 0$, satisfying the scaling relation (2.9), gives a more complete picture, since we can prove the convergence in expectation for the difference of the solutions to the stochastic systems (2.5) and (2.8).

Finally, we remark that the cross-diffusion models (2.1) and (2.2) have quite different structural properties; also see [12, 13]. First, system (2.2) has a formal gradient-flow structure for each species separately, while system (2.1) can be written, under the detailed-balance condition [26], only in a vector-valued gradient-flow form. Second, the segregation behavior of both models is different, i.e., segregation is stronger for the solutions to (2.2) than for model (2.1); see the numerical experiments in Section 2.7.

This chapter is organized as follows: We present our assumptions and main results in Section 2.2. The existence of smooth solutions to the cross-diffusion systems (2.1) and (2.7) and an error estimate for the difference of the corresponding solutions is proved in Sections 2.3 and 2.4, respectively. The proofs are based on Banach's fixed-point theorem and higher-order estimations. We present the full proof since the environmental potential $U_i(x) = -\frac{1}{2}|x|^2$ is not square-integrable, which requires some care; see the arguments following (2.22). Section 2.5 is concerned with the identification of the solutions to the local and nonlocal cross-diffusion systems (2.1) and (2.7), respectively, with the probability density functions associated to the particle systems (2.8) and (2.6), respectively. Error estimate (2.10), the main result of this chapter, is proved in Section 2.6. In Section 2.7, we present Monte–Carlo simulations for an Euler–Maruyama discretization of system (2.5) and compare them to the numerical results from the particle system associated to (2.2). In the appendix of this chapter (Section 2.A), we recall some inequalities used within the proofs of this chapter.

2.2 Assumptions and main results

We impose the following assumptions:

(A1) **Data:** $\sigma_i \in (0,\infty)$ and $\xi_i^1,\ldots,\xi_i^N$ are independent and identically distributed (iid) square-integrable random variables with the common density function $u_{0,i}$ for $i = 1,\ldots,n$ on the probability space $(\Omega,\mathcal{F},P)$.

(A2) **Environmental potential:** $U_i(x) = -\frac{1}{2}|x|^2$, $i = 1,\ldots,n$.

(A3) **Interaction potential:** $B_{ij} \in C_0^\infty(\mathbb{R}^d)$ satisfies $\operatorname{supp}(B_{ij}) \subset B_1(0)$, where $B_1(0)$ is the unit ball in $\mathbb{R}^d$ and $i,j = 1,\ldots,n$.

(A4) **Nonlinearity:** $f \in W^{s+1,\infty}_{\rm loc}(\mathbb{R};[0,\infty))$ and $f_\eta \in W^{s+1,\infty}(\mathbb{R},[0,\infty))$ is such that $f_\eta = f$ on $[-a_\eta, a_\eta]$ and the Lipschitz constant of f_η is less than or equal to $\eta^{-\alpha}$ for a fixed $\alpha \in [0,1)$. Here, $s > d/2+1$ and $a_\eta \to \infty$ as $\eta \to 0$. If f is globally Lipschitz continuous, we set $\alpha = 0$ and $f_\eta = f$.

Remark 2.1 (Discussion)**. *Environmental potential:*** *The sign of U_i guarantees that the populations are dispersed since the drift term becomes $-x\cdot\nabla u_i - u_i$. We have taken a quadratic potential U_i to simplify the presentation. "Dispersive" potentials (i.e. potentials U_i with $\Delta U_i \le 0$) are needed in the analysis, since we cannot bound terms including ΔU_i if $\Delta U_i \ge 0$. It is possible to choose general (dispersive) potentials $U_i \in C^\infty(\mathbb{R}^d)$ such that ∇U_i is globally Lipschitz continuous, $D^kU_i \in L^\infty(\mathbb{R}^d)$ for $k = 2,\ldots,s+2$, the Hessian D^2U_i is negative semidefinite, $\Delta U_i < 0$, and D^kU_i for $k = 3,\ldots,s$ is sufficiently small in the $L^\infty(\mathbb{R}^d)$ norm. Thus, we may choose $U_i(x) = -|x|^2 + g(x)$ and g is a smooth perturbation.*

Nonlinearity: *Since f is not assumed to be globally Lipschitz continuous, we need to approximate the nonlinearity. The condition on the Lipschitz constant of f_η ensures that we have a control on the growth of the Lipschitz constant of f_η in the limit $N \to \infty$ and $\eta \to 0$. This growth condition is needed in the proof of Lemma 2.9; see (2.34) and thereafter. The condition $s > d/2+1$ ensures that the embedding $H^s(\mathbb{R}^d) \hookrightarrow W^{1,\infty}(\mathbb{R}^d)$ is continuous, and this embedding is needed to obtain solutions in $H^s(\mathbb{R}^d)$ and to derive the estimates.*

We introduce some notation. We set

$$a_{ij} = \int_{\mathbb{R}^d} B_{ij}(|x|)\mathrm{d}x, \quad i,j = 1,\dots,n,$$

$B_{ij}^\eta(x) = \eta^{-d} B_{ij}(|x|/\eta)$, $A_{ij} = \|B_{ij}\|_{L^1(\mathbb{R}^d)} = \|B_{ij}^\eta\|_{L^1(\mathbb{R}^d)}$ and $A = \max_{i,j=1,\dots,n} A_{ij}$. Let $C_s > 0$ be the constant of the continuous embedding $H^s(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d)$ and set

$$I = [-2AC_s\|u_0\|_{H^s(\mathbb{R}^d)}, 2AC_s\|u_0\|_{H^s(\mathbb{R}^d)}]. \tag{2.11}$$

Then, for small $\eta > 0$ such that $a_\eta \geq 2AC_s\|u_0\|_{H^s(\mathbb{R}^d)}$, we have $f_\eta = f$ on I.
First, we ensure that the nonlocal and local cross-diffusion systems (2.7) and (2.1), respectively, have global smooth solutions.

Theorem 2.2 (Existence for the nonlocal system)**.** *Let Assumptions (A2) and (A4) hold, $u_0 \in H^s(\mathbb{R}^d;\mathbb{R}^n)$ for $s > d/2+1$, and let $\eta > 0$ be such that $a_\eta \geq 2AC_s\|u_0\|_{H^s(\mathbb{R}^d)}$. There exists $\varepsilon > 0$ depending on u_0 such that if $\|f\|_{C^{s+1}(I)} \leq \varepsilon$, system (2.7) possesses a unique solution $u_\eta = (u_{\eta,1},\dots,u_{\eta,n})$ satisfying*

$$u_{\eta,i} \in L^\infty(0,\infty;H^s(\mathbb{R}^d)) \cap L^2(0,\infty;H^{s+1}(\mathbb{R}^d)),$$
$$\|u_\eta\|^2_{L^\infty(0,T;H^s(\mathbb{R}^d))} + \sigma_*\|\nabla u_\eta\|^2_{L^2(0,\infty;H^s(\mathbb{R}^d))} \leq \|u_0\|^2_{H^s(\mathbb{R}^d)},$$

where $0 < \sigma_* < \sigma_{min} := \min_{i=1,\dots,n}\sigma_i$.

The dependence of ε on u_0 can be made more explicit. The proof shows that we need to choose $0 < \varepsilon < C\sigma_{\min}^{1/2}\|u_0\|^{-s}_{H^s(\mathbb{R}^d)}$, where $C > 0$ is independent of u_0 and σ_i. Thus, if $\|f\|_{C^{s+1}(I)}$ is finite, the global existence result is valid for small initial data.

Theorem 2.3 (Existence for the local system)**.** *Let u_0 and η satisfy the assumptions of Theorem 2.2. Then there exists $\varepsilon > 0$ depending on u_0 such that if $\|f\|_{C^{s+1}(I)} \leq \varepsilon$, system (2.1) possesses a unique solution $u = (u_1,\dots,u_n)$ satisfying*

$$u_i \in L^\infty(0,\infty;H^s(\mathbb{R}^d)) \cap L^2(0,\infty;H^{s+1}(\mathbb{R}^d)), \quad i = 1,\dots,n,$$
$$\|u\|^2_{L^\infty(0,\infty;H^s(\mathbb{R}^d))} + \sigma_*\|\nabla u\|^2_{L^2(0,\infty;H^s(\mathbb{R}^d))} \leq \|u_0\|^2_{H^s(\mathbb{R}^d)},$$

where $0 < \sigma_* < \sigma_{min}$. Moreover, with the solution u_η from Theorem 2.2, it holds that for an arbitrary $T > 0$,

$$\|u - u_\eta\|_{L^\infty(0,T;L^2(\mathbb{R}^d))} + \|\nabla(u-u_\eta)\|_{L^2(0,T;L^2(\mathbb{R}^d))} \leq C(T)\eta.$$

Next, we state an existence result for the stochastic particle systems (2.5), (2.6), and (2.8).

Proposition 2.4. *Let Assumptions (A1)–(A4) hold and let $\eta > 0$, $N \in \mathbb{N}$. Then:*
(i) *There exist unique square-integrable adapted stochastic processes with continuous paths, which are strong solutions to systems* (2.5), (2.6), *and* (2.8), *respectively.*
(ii) *For each $t > 0$, the (nNd)-dimensional random variables $\overline{X}^{\eta}(t)$ and $\widehat{X}(t)$ possess density functions $\overline{u}_{\eta}(t)^{\otimes N}$ and $\widehat{u}(t)^{\otimes N}$ with respect to the Lebesgue measure on $\mathbb{R}^{nNd}$, respectively.*

The proof follows from [67] and [89]. Indeed, Theorem 2.9 in [67, page 289] shows that there exist continuous square-integrable stochastic processes, which are strong solutions to (2.5), (2.6), and (2.8), respectively. Strong uniqueness is guaranteed by Theorem 2.5 in [67, page 287]. We conclude from [89, Theorem 2.3.1] that $\overline{X}_{\eta}(t)$ and $\widehat{X}(t)$ are absolutely continuous with respect to the Lebesgue measure and thus, they possess density functions $\overline{u}_{\eta}(t,x)^{\otimes N}$ and $\widehat{u}(t,x)^{\otimes N}$, respectively. We prove in Section 2.5 that the density functions $\overline{u}_{\eta}$ and $\widehat{u}$ can be identified with u_{η} and u, the solutions to (2.7) and (2.1), respectively.
The following theorem is our main result.

Theorem 2.5. *Let $X_{k,i}^{N,\eta}$ and $\widehat{X}_{k,i}$ be the solutions to* (2.5) *and* (2.8), *respectively. Then there exist parameters $\delta > 0$, depending on n, σ_{min}, and T, and $\varepsilon > 0$, depending on u_0, such that if $\eta^{-2(d+1+\alpha)} \leq \delta \log N$ and $\|f\|_{C^{s+1}(I)} \leq \varepsilon$,*

$$\sup_{k=1,\dots,N} \mathbb{E}\left(\sum_{i=1}^{n} \sup_{0<s<T} \left|(X_{k,i}^{N,\eta} - \widehat{X}_{k,i})(s)\right|^2 \right) \leq C(T,n,\sigma_{min})\eta^{2(1-\alpha)},$$

where $\alpha \geq 0$ is defined in Assumption (A4).

Remark 2.6. *It is well-known that this result implies propagation of chaos in the single-species case; see, e.g., [62, Section 3.1]. In the multi-species case, this generalizes for fixed k to the convergence of the k-marginal distribution $F_k(t)$ of $(X_{j_1,i_1}^{N,\eta}(t), \dots, X_{j_k,i_k}^{N,\eta}(t))$ at any time $t > 0$ towards the product measure $\otimes_{\ell=1}^{k} u_{i_\ell}(\cdot,t)$ as $N \to \infty$, $\eta \to 0$, i.e.*

$$W_2^2\left(F_k(t), \bigotimes_{\ell=1}^{k} u_{i_\ell}(\cdot,t) \right) \leq kC(T,n,\sigma_{min})\eta \to 0,$$

where W_2 denotes the 2-Wasserstein distance. We refer to the introduction of this thesis, in particular Section 1.2.1, where we sketch the connection between a convergence result in expectation and propagation of chaos in Wasserstein-distance. Additionally, a discussion of propagation of chaos for multi-species systems is provided in Section 1.1.2.

2.3 Proof of Theorem 2.2

We prove the global existence of smooth solutions to the nonlocal system (2.7). Since η is fixed in the proof, we omit it for u_{η} to simplify the notation. We split the proof in several steps. In the first step, we prove the existence of local-in-time solutions satisfying $\|u_i(t)\|_{H^s(\mathbb{R}^d)} \leq 2\|u_0\|_{H^s(\mathbb{R}^d)}$ for $0 < t < T(\eta)$ for some (possibly) small $T(\eta) > 0$. Actually, we show in the second step, that the factor 2 can be replaced by one. This uniform estimate allows us in the third step to conclude the global existence.

Step 1: Local existence of solutions. In this step, the smallness conditions on η and f are not needed. The idea is to apply the Banach fixed-point theorem on the space

$$X_T := \big\{v \in L^\infty(0,T;H^s(\mathbb{R}^d;\mathbb{R}^n)) : \|v\|_{L^\infty(0,T;H^s(\mathbb{R}^d))} \le 2\|u_0\|_{H^s(\mathbb{R}^d)}\big\},$$

where $T > 0$ will be determined later in this proof. We define the fixed-point operator $S : X_T \to X_T$, $S(v) = u$, where u is the unique solution to the linear problem

$$\partial_t u_i = \operatorname{div}(u_i \nabla U_i) + \Delta\big(u_i(\sigma_i + K_i(v(t,x)))\big), \quad u_i(0) = u_{0,i} \quad \text{in } \mathbb{R}^d,\ t > 0, \tag{2.12}$$

with $K_i(v) = \sum_{j=1}^n f_\eta(B_{ij}^\eta * v_j) \ge 0$, $i = 1,\dots,n$. We need to show that S is well defined. We infer from Young's convolution inequality (Lemma 2.11) and the embedding $H^s(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d)$ that

$$\begin{aligned}\sup_{0<t<T} \|\nabla K_i(v)\|_{L^\infty(\mathbb{R}^d)} &\le \sum_{j=1}^n \|f'_\eta\|_{L^\infty(\mathbb{R})} \|\nabla B_{ij}^\eta\|_{L^1(\mathbb{R}^d)} \sup_{0<t<T} \|v_j(t)\|_{L^\infty(\mathbb{R}^d)} \\ &\le C(\eta) \sum_{j=1}^n \|v_j\|_{L^\infty(0,T;H^s(\mathbb{R}^d))} < \infty, \end{aligned} \tag{2.13}$$

i.e., $K_i(v)$ is globally Lipschitz continuous. Therefore, a Galerkin argument to verify higher-order regularity shows that, for given $v \in X_T$, there exists a unique solution $u_i \in L^\infty(0,T;H^s(\mathbb{R}^d)) \cap L^2(0,T;H^{s+1}(\mathbb{R}^d))$ to (2.12). It remains to show that $u = (u_1,\dots,u_n) \in X_T$ for some $T > 0$. The estimations are not difficult, but since ∇U_i is not square integrable, some care is needed.
First, we prove higher-order estimates for $K_i(v)$. Let $\alpha \in \mathbb{N}_0^d$ be a multi-index with order $|\alpha| = m \le s$. By Lemma 2.13 and Young's convolution inequality,

$$\begin{aligned}\int_0^T \|D^\alpha K_i(v)\|_{L^2(\mathbb{R}^d)}^2 dt &\le C \int_0^T \sum_{j=1}^n \|f'_\eta\|_{C^{m-1}(\mathbb{R})}^2 \|B_{ij}^\eta * v_j\|_{L^\infty(\mathbb{R}^d)}^{2(m-1)} \|D^\alpha(B_{ij}^\eta * v_j)\|_{L^2(\mathbb{R}^d)}^2 dt \\ &\le C(\eta) \int_0^T \sum_{j=1}^n \|B_{ij}^\eta\|_{L^1(\mathbb{R}^d)}^{2m} \|v_j\|_{L^\infty(\mathbb{R}^d)}^{2(m-1)} \|D^\alpha v_j\|_{L^2(\mathbb{R}^d)}^2 dt \\ &\le C(\eta) \sum_{j=1}^n \int_0^T \|v_j\|_{H^s(\mathbb{R}^d)}^{2m} dt < \infty, \end{aligned} \tag{2.14}$$

where here and in the following, $C > 0$, $C(\eta) > 0$, etc. are generic constants with values changing from line to line. In a similar way, applying Lemmas 2.11 and 2.12,

$$\begin{aligned}\sup_{0<t<T} \|D^\alpha \nabla K_i(v)\|_{L^2(\mathbb{R}^d)}^2 &\le C \sup_{0<t<T} \sum_{j=1}^n \big\|D^\alpha\big(f'_\eta(B_{ij}^\eta * v_j)\nabla B_{ij}^\eta * v_j\big)\big\|_{L^2(\mathbb{R}^d)}^2 \\ &\le C \sup_{0<t<T} \sum_{j=1}^n \Big(\|f'_\eta(B_{ij}^\eta * v_j)\|_{L^\infty(\mathbb{R}^d)} \|\nabla B_{ij}^\eta\|_{L^1(\mathbb{R}^d)} \|D^m v_j\|_{L^2(\mathbb{R}^d)} \\ &\quad + \|D^m(f'_\eta(B_{ij}^\eta * v_j))\|_{L^2(\mathbb{R}^d)} \|\nabla B_{ij}^\eta\|_{L^1(\mathbb{R}^d)} \|v_j\|_{L^\infty(\mathbb{R}^d)}\Big)^2 \le C(\eta), \end{aligned} \tag{2.15}$$

since, according to Lemma 2.13, we can bound $\sup_{0<t<T}\|D^m(f'_\eta(B_{ij}*v_j))\|_{L^2(\mathbb{R}^d)}$ in terms of $\|f_\eta\|_{C^{s+1}(\mathbb{R})}$, $\|B^\eta_{ij}\|_{L^1(\mathbb{R}^d)}$, and $\sup_{0<t<T}\|v_j\|_{H^s(\mathbb{R}^d)}$, and it holds that $\|\nabla B^\eta_{ij}\|_{L^1(\mathbb{R}^d)}\le C(\eta)$.
We proceed with the proof of $u\in X_T$ for some $T>0$. Applying D^α to (2.12), multiplying the resulting equation by $D^\alpha u_i$, and integrating over $(0,\tau)\times\mathbb{R}^d$ for $\tau<T$ yields

$$\frac{1}{2}\int_{\mathbb{R}^d}|D^\alpha u_i(\tau)|^2\mathrm{d}x-\frac{1}{2}\int_{\mathbb{R}^d}|D^\alpha u_{0,i}|^2\mathrm{d}x+\sigma_i\int_0^\tau\int_{\mathbb{R}^d}|\nabla D^\alpha u_i|^2\mathrm{d}x\mathrm{d}t=I_1+I_2+I_3,\quad(2.16)$$

where

$$\begin{aligned}I_1&=-\int_0^\tau\int_{\mathbb{R}^d}\nabla D^\alpha u_i\cdot D^\alpha(u_i\nabla U_i)\mathrm{d}x\mathrm{d}t,\\ I_2&=-\int_0^\tau\int_{\mathbb{R}^d}\nabla D^\alpha u_i\cdot D^\alpha(\nabla u_iK_i(v))\mathrm{d}x\mathrm{d}t,\\ I_3&=-\int_0^\tau\int_{\mathbb{R}^d}\nabla D^\alpha u_i\cdot D^\alpha(u_i\nabla K_i(v))\mathrm{d}x\mathrm{d}t.\end{aligned}$$

First, let $|\alpha|=m=0$. Then, integrating by parts in I_1, using Young's inequality, and observing that $U_i(x)=-\frac{1}{2}|x|^2$,

$$\begin{aligned}I_1&=\frac{1}{2}\int_0^\tau\int_{\mathbb{R}^d}u_i^2\Delta U_i\mathrm{d}x\mathrm{d}t=-\frac{d}{2}\int_0^\tau\int_{\mathbb{R}^d}u_i^2\mathrm{d}x\mathrm{d}t\le 0,\\ I_2&=-\int_0^\tau\int_{\mathbb{R}^d}K_i(v)|\nabla u_i|^2\mathrm{d}x\mathrm{d}t\le 0,\\ I_3&\le\frac{\sigma_i}{2}\int_0^\tau\int_{\mathbb{R}^d}|\nabla u_i|^2\mathrm{d}x\mathrm{d}t+\frac{1}{2\sigma_i}\|\nabla K_i(v)\|^2_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\int_0^\tau\|u_i\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t,\end{aligned}$$

where we used $K_i(v)\ge 0$ for I_2. It follows from (2.13) that

$$I_1+I_2+I_3\le\frac{\sigma_i}{2}\int_0^\tau\int_{\mathbb{R}^d}|\nabla u_i|^2\mathrm{d}x\mathrm{d}t+C\int_0^\tau\|u_i\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t,$$

where $C>0$ depends on the $L^\infty(0,T;H^s(\mathbb{R}^d))$ norm of v. Inserting this estimate into (2.16) with $\alpha=0$ and applying the Gronwall inequality, we infer that

$$\int_{\mathbb{R}^d}u_i(\tau)^2\mathrm{d}x+\frac{\sigma_i}{2}\int_0^\tau\int_{\mathbb{R}^d}|\nabla u_i|^2\mathrm{d}x\mathrm{d}t\le C(u_0)e^{C\tau}.$$

This shows that u_i is bounded in $L^\infty(0,T;L^2(\mathbb{R}^d))$ and $L^2(0,T;H^1(\mathbb{R}^d))$.
Now, let $|\alpha|=m\ge 1$. Then, integrating by parts, using $\Delta U_i\le 0$, and applying Young's inequality again,

$$\begin{aligned}I_1&=\frac{1}{2}\int_0^\tau\int_{\mathbb{R}^d}(D^\alpha u_i)^2\Delta U_i\mathrm{d}x\mathrm{d}t-\int_0^\tau\int_{\mathbb{R}^d}\nabla D^\alpha u_i\cdot\big(D^\alpha(u_i\nabla U_i)-D^\alpha u_i\nabla U_i\big)\mathrm{d}x\mathrm{d}t\\ &\le\frac{\sigma_i}{4}\int_0^\tau\int_{\mathbb{R}^d}|\nabla D^\alpha u_i|^2\mathrm{d}x\mathrm{d}t+\sum_{0<|\beta|\le|\alpha|}\int_0^\tau c_\beta\|D^{\alpha-\beta}u_i\|^2_{L^2(\mathbb{R}^d)}\|D^\beta\nabla U_i\|^2_{L^\infty(\mathbb{R}^d)}\mathrm{d}t\\ &\le\frac{\sigma_i}{4}\int_0^\tau\int_{\mathbb{R}^d}|\nabla D^\alpha u_i|^2\mathrm{d}x\mathrm{d}t+C\int_0^\tau\|u_i\|^2_{H^{m-1}(\mathbb{R}^d)}\mathrm{d}t,\end{aligned}$$

where we used the fact that $D^\beta \nabla U_i$ is bounded for $|\beta| = 1$ and vanishes for $|\beta| > 1$. It follows from integration by parts, $K_i(v) \geq 0$, and Lemma 2.14 that

$$\begin{aligned} I_2 &= -\int_0^\tau \int_{\mathbb{R}^d} \nabla D^\alpha u_i \cdot \big(D^\alpha(\nabla u_i K_i(v)) - \nabla D^\alpha u_i K_i(v)\big) \mathrm{d}x\mathrm{d}t \\ &\quad - \int_0^\tau \int_{\mathbb{R}^d} K_i(v)|\nabla D^\alpha u_i|^2 \mathrm{d}x\mathrm{d}t \\ &\leq \frac{\sigma_i}{4} \int_0^\tau \int_{\mathbb{R}^d} |\nabla D^\alpha u_i|^2 \mathrm{d}x\mathrm{d}t + C \int_0^\tau \big(\|DK_i(v)\|_{L^\infty(\mathbb{R}^d)} \|D^{m-1}\nabla u_i\|_{L^2(\mathbb{R}^d)} \\ &\quad + \|D^m K_i(v)\|_{L^2(\mathbb{R}^d)} \|\nabla u_i\|_{L^\infty(\mathbb{R}^d)}\big)^2 \mathrm{d}x\mathrm{d}t. \end{aligned}$$

We infer from estimates (2.13) and (2.14) for $K_i(v)$ and the embedding $H^s(\mathbb{R}^d) \hookrightarrow W^{1,\infty}(\mathbb{R}^d)$ that

$$I_2 \leq \frac{\sigma_i}{4} \int_0^\tau \int_{\mathbb{R}^d} |\nabla D^\alpha u_i|^2 \mathrm{d}x\mathrm{d}t + C \int_0^\tau \|u_i\|^2_{H^s(\mathbb{R}^d)} \mathrm{d}t.$$

Finally, we use Lemma 2.12 and estimates (2.13) and (2.15) to obtain

$$\begin{aligned} I_3 &\leq \frac{\sigma_i}{4} \int_0^\tau \int_{\mathbb{R}^d} |\nabla D^\alpha u_i|^2 \mathrm{d}x\mathrm{d}t + C \int_0^\tau \int_{\mathbb{R}^d} \big(\|u_i\|_{L^\infty(\mathbb{R}^d)} \|D^m \nabla K_i(v)\|_{L^2(\mathbb{R}^d)} \\ &\quad + \|D^m u_i\|_{L^2(\mathbb{R}^d)} \|\nabla K_i(v)\|_{L^\infty(\mathbb{R}^d)}\big)^2 \mathrm{d}x\mathrm{d}t \\ &\leq \frac{\sigma_i}{4} \int_0^\tau \int_{\mathbb{R}^d} |\nabla D^\alpha u_i|^2 \mathrm{d}x\mathrm{d}t + C(\eta) \int_0^\tau \|u_i\|^2_{H^s(\mathbb{R}^d)} \mathrm{d}t. \end{aligned}$$

Inserting these estimates into (2.16) and summing over $|\alpha| \leq s$, we arrive at

$$\|u_i(\tau)\|^2_{H^s(\mathbb{R}^d)} + \frac{\sigma_i}{4} \int_0^\tau \|\nabla u_i\|^2_{H^s(\mathbb{R}^d)} \mathrm{d}t \leq \|u_{0,i}\|^2_{H^s(\mathbb{R}^d)} + C(\eta) \int_0^\tau \|u_i\|^2_{H^s(\mathbb{R}^d)} \mathrm{d}t.$$

Summing over $i = 1, \ldots, n$ and applying Gronwall's inequality gives

$$\|u(\tau)\|^2_{H^s(\mathbb{R}^d)} \leq \|u_0\|^2_{H^s(\mathbb{R}^d)} e^{C(\eta)\tau} \leq \|u_0\|^2_{H^s(\mathbb{R}^d)} e^{C(\eta)T}.$$

Choosing $T > 0$ sufficiently small, we can ensure that $\|u(\tau)\|_{H^s(\mathbb{R}^d)} \leq 2\|u_0\|_{H^s(\mathbb{R}^d)}$ for all $0 < \tau < T$. This shows that $u \in X_T$, i.e., the operator is well-defined.
Next, we prove that $S : X_T \to X_T$ is a contraction. Let v, $w \in X_T$ and set $\bar{v} = S(v)$ and $\bar{w} = S(w)$. Taking the difference of equations (2.12) satisfied by $\bar{v}_i$ and $\bar{w}_i$, respectively, using the test function $\bar{v}_i - \bar{w}_i$, and integrating by parts, it follows that

$$\frac{1}{2} \int_{\mathbb{R}^d} (\bar{v}_i - \bar{w}_i)(\tau)^2 \mathrm{d}x + \sigma_i \int_0^\tau \int_{\mathbb{R}^d} |\nabla(\bar{v}_i - \bar{w}_i)|^2 \mathrm{d}x\mathrm{d}t = I_4 + I_5 + I_6, \tag{2.17}$$

where

$$I_4 = \frac{1}{2} \int_0^\tau \int_{\mathbb{R}^d} \Delta U_i (\bar{v}_i - \bar{w}_i)^2 \mathrm{d}x\mathrm{d}t \leq 0,$$

$$I_5 = -\int_0^\tau \int_{\mathbb{R}^d} \nabla\big((\bar v_i - \bar w_i)K_i(v)\big)\cdot\nabla(\bar v_i - \bar w_i)\mathrm{d}x\mathrm{d}t,$$
$$I_6 = -\int_0^\tau \int_{\mathbb{R}^d} \nabla\big(\bar w_i(K_i(v) - K_i(w))\big)\cdot\nabla(\bar v_i - \bar w_i)\mathrm{d}x\mathrm{d}t.$$

Because of $K_i(v) \geq 0$ and estimate (2.13) for $\nabla K_i(v)$, we find that, by Young's inequality,

$$\begin{aligned}
I_5 &= -\int_0^\tau \int_{\mathbb{R}^d} K_i(v)|\nabla(\bar v_i - \bar w_i)|^2\mathrm{d}x\mathrm{d}t - \int_0^\tau \int_{\mathbb{R}^d} (\bar v_i - \bar w_i)\nabla K_i(v)\cdot\nabla(\bar v_i - \bar w_i)\mathrm{d}x\mathrm{d}t \\
&\leq \frac{\sigma_i}{4}\int_0^\tau \|\nabla(\bar v_i - \bar w_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C(\sigma_i)\int_0^\tau \|\bar v_i - \bar w_i\|^2_{L^2(\mathbb{R}^d)}\|\nabla K_i(v)\|^2_{L^\infty(\mathbb{R}^d)}\mathrm{d}t \\
&\leq \frac{\sigma_i}{4}\int_0^\tau \|\nabla(\bar v_i - \bar w_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C(\eta)\int_0^\tau \|\bar v_i - \bar w_i\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t.
\end{aligned}$$

It follows again from Young's inequality that

$$\begin{aligned}
I_6 &\leq \frac{\sigma_i}{4}\int_0^\tau \|\nabla(\bar v_i - \bar w_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C(\sigma_i)\int_0^\tau \|\nabla \bar w_i\|^2_{L^\infty(\mathbb{R}^d)}\|K_i(v) - K_i(w)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t \\
&\quad + C(\sigma_i)\int_0^\tau \|\bar w_i\|^2_{L^\infty(\mathbb{R}^d)}\|\nabla(K_i(v) - K_i(w))\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t. \qquad (2.18)
\end{aligned}$$

Since $\bar w \in X_T$, we have $\|\nabla \bar w_i\|_{L^\infty(\mathbb{R}^d)} \leq C\|\bar w_i\|_{H^s(\mathbb{R}^d)} \leq C(u_0)$ and $\|\bar w_i\|_{L^\infty(\mathbb{R}^d)} \leq C(u_0)$. We use the fact that f_η and f'_η are globally Lipschitz continuous:

$$\|K_i(v) - K_i(w)\|_{L^2(\mathbb{R}^d)} \leq C(\eta)\sum_{j=1}^n \|B^\eta_{ij} * (v_j - w_j)\|_{L^2(\mathbb{R}^d)} \leq C(\eta)\|v - w\|_{L^2(\mathbb{R}^d)},$$

$$\begin{aligned}
\|\nabla(K_i(v) - K_i(w))\|_{L^2(\mathbb{R}^d)} &\leq \sum_{j=1}^n \|(f'_\eta(B^\eta_{ij} * v_j) - f'_\eta(B^\eta_{ij} * w_j))B^\eta_{ij} * \nabla v_j\|_{L^2(\mathbb{R}^d)} \\
&\quad + \sum_{j=1}^n \|f'_\eta(B^\eta_{ij} * w_j)\nabla B^\eta_{ij} * (v_j - w_j)\|_{L^2(\mathbb{R}^d)} \\
&\leq C(\eta)\sum_{j=1}^n \|v_j - w_j\|_{L^2(\mathbb{R}^d)}\|B^\eta_{ij}\|_{L^1(\mathbb{R}^d)}\|\nabla v_j\|_{L^\infty(\mathbb{R}^d)} \\
&\quad + C(\eta)\sum_{j=1}^n \|\nabla B^\eta_{ij}\|_{L^1(\mathbb{R}^d)}\|v_j - w_j\|_{L^2(\mathbb{R}^d)} \\
&\leq C(\eta)\|v - w\|_{L^2(\mathbb{R}^d)}.
\end{aligned}$$

Inserting these inequalities into (2.18) and summarizing the estimates for I_4, I_5, and I_6, we conclude from (2.17) and summation over $i = 1, \ldots, n$ that

$$\begin{aligned}
&\frac{1}{2}\|(\bar v - \bar w)(\tau)\|^2_{L^2(\mathbb{R}^d)} + \sum_{i=1}^n \frac{\sigma_i}{4}\int_0^\tau \|\nabla(\bar v_i - \bar w_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t \\
&\quad \leq C_1\int_0^\tau \|\bar v - \bar w\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C_2\tau\|v - w\|^2_{L^\infty(0,\tau;L^2(\mathbb{R}^d))}.
\end{aligned}$$

We apply Gronwall's inequality and the supremum over $0 < \tau < T$ to find that

$$\|\bar{v} - \bar{w}\|^2_{L^\infty(0,T;L^2(\mathbb{R}^d))} \leq 2C_2 e^{2C_1 T} T \|v - w\|^2_{L^\infty(0,T;L^2(\mathbb{R}^d))}.$$

Thus, choosing $T > 0$ such that $2C_2 e^{2C_1 T} T < 1$, we infer that $S : X_T \to X_T$ is a contraction. By Banach's fixed-point theorem, there exists a unique solution $u \in L^\infty(0,T;H^s(\mathbb{R}^d)) \cap L^2(0,T;H^{s+1}(\mathbb{R}^d))$ to (2.7).

Step 2: A priori estimates. Let $u = u_\eta$ be the unique solution to (2.7). We know from Step 1 that $\|u_i(t)\|_{L^\infty(\mathbb{R}^d)} \leq C_s \|u_i(t)\|_{H^s(\mathbb{R}^d)} \leq 2C_s \|u_0\|_{H^s(\mathbb{R}^d)}$ for any $0 < t < T$. Recall that $T = T(\eta)$ and hence we do not have uniform estimates in η even for small $T > 0$ at this step. We show in this step the estimate $\|u_i(t)\|_{H^s(\mathbb{R}^d)} \leq \|u_0\|_{H^s(\mathbb{R}^d)}$, which allows us to conclude that the end time T can be arbitrary and actually does not depend on η. We apply D^α to (2.7) (with $|\alpha| = m \leq s$), multiply the resulting equation by $D^\alpha u_i$, and integrate over $(0,\tau) \times \mathbb{R}^d$ for $\tau < T$, similarly to the corresponding estimate in Step 1:

$$\frac{1}{2}\int_{\mathbb{R}^d} |D^\alpha u_i(\tau)|^2 \mathrm{d}x - \frac{1}{2}\int_{\mathbb{R}^d} |D^\alpha u_{0,i}|^2 \mathrm{d}x + \sigma_i \int_0^\tau \int_{\mathbb{R}^d} |\nabla D^\alpha u_i|^2 \mathrm{d}x\mathrm{d}t = I_7 + I_8 + I_9, \quad (2.19)$$

where

$$\begin{aligned} I_7 &= -\int_0^\tau \int_{\mathbb{R}^d} \nabla D^\alpha u_i \cdot D^\alpha(u_i \nabla U_i)\mathrm{d}x\mathrm{d}t, \\ I_8 &= -\int_0^\tau \int_{\mathbb{R}^d} \nabla D^\alpha u_i \cdot D^\alpha(\nabla u_i K_i(u))\mathrm{d}x\mathrm{d}t, \\ I_9 &= -\int_0^\tau \int_{\mathbb{R}^d} \nabla D^\alpha u_i \cdot D^\alpha(u_i \nabla K_i(u))\mathrm{d}x\mathrm{d}t, \end{aligned}$$

and we recall that $K_i(u) = \sum_{j=1}^n f_\eta(B^\eta_{ij} * u_j)$.

First, let $m = 0$. Arguing similarly as for I_1 and I_2, we find that $I_7 \leq 0$ and $I_8 \leq 0$. We estimate $\nabla K_i(u) = \sum_{j=1}^n f'_\eta(B^\eta_{ij} * u_j) B^\eta_{ij} * \nabla u_j$:

$$\|\nabla K_i(u)\|_{L^2(\mathbb{R}^d)} \leq A \sum_{j=1}^n \|f'_\eta(B^\eta_{ij} * u_j)\|_{L^\infty(\mathbb{R}^d)} \|\nabla u_j\|_{L^2(\mathbb{R}^d)}, \quad (2.20)$$

recalling that $A = \max_{i,j=1,\ldots,n} \|B^\eta_{ij}\|_{L^1(\mathbb{R}^d)}$. This gives for $m = 0$:

$$\begin{aligned} I_9 &\leq \|u_i\|_{L^\infty(0,\tau;L^\infty(\mathbb{R}^d))} \int_0^\tau \|\nabla u_i\|_{L^2(\mathbb{R}^d)} \|\nabla K_i(u)\|_{L^2(\mathbb{R}^d)} \mathrm{d}t \\ &\leq C\|u_0\|_{H^s(\mathbb{R}^d)} \sum_{j=1}^n \|f'_\eta(B^\eta_{ij} * u_j)\|_{L^\infty(0,\tau;L^\infty(\mathbb{R}^d))} \int_0^\tau \|\nabla u_j\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t. \end{aligned}$$

From this point on, we will need the smallness condition on f_η and f'_η. Because of

$$\|B^\eta_{ij} * u_j(t)\|_{L^\infty(\mathbb{R}^d)} \leq \|B^\eta_{ij}\|_{L^1(\mathbb{R}^d)} C_s \|u_j(t)\|_{H^s(\mathbb{R}^d)} \leq 2AC_s \|u_0\|_{H^s(\mathbb{R}^d)}, \quad (2.21)$$

where $C_s > 0$ is the constant of the embedding $H^s(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d)$, $(B^\eta_{ij} * u_j(t))(x)$ lies in the interval $I = [-2AC_s\|u_0\|_{H^s(\mathbb{R}^d)}, 2AC_s\|u_0\|_{H^s(\mathbb{R}^d)}]$ for $0 < t < T$ and $x \in \mathbb{R}^d$. On this interval, $f_\eta = f$ if $\eta > 0$ is sufficiently small. From now on, we use $f \le \varepsilon$ and $|f'| \le \varepsilon$ on I for a small $\varepsilon > 0$. Thus, we have

$$I_9 \le C\varepsilon \|u_0\|_{H^s(\mathbb{R}^d)} \sum_{j=1}^n \int_0^\tau \|\nabla u_j\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t.$$

Inserting these estimates into (2.19), we conclude that

$$\|u_i(\tau)\|^2_{L^2(\mathbb{R}^d)} + \big(\sigma_i - C\varepsilon\|u_0\|_{H^s(\mathbb{R}^d)}\big) \int_0^\tau \|\nabla u_i\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t \le \|u_{0,i}\|^2_{L^2(\mathbb{R}^d)}.$$

Choosing $\varepsilon > 0$ sufficiently small, this gives an estimate for u_i in $L^\infty(0,T;L^2(\mathbb{R}^d)) \cap L^2(0,T;H^1(\mathbb{R}^d))$.

Next, let $m \ge 1$. The estimate for I_7 is delicate since $\nabla U_i \notin L^2(\mathbb{R}^d)$, and the corresponding estimate for I_1 cannot be directly used. We split I_7 into two parts:

$$\begin{aligned} I_7 &= \int_0^\tau \int_{\mathbb{R}^d} D^\alpha u_i D^\alpha(\nabla u_i \cdot \nabla U_i + u_i \Delta U_i) \mathrm{d}x\mathrm{d}t \\ &= \int_0^\tau \int_{\mathbb{R}^d} D^\alpha u_i \big(D^\alpha(\nabla u_i \cdot \nabla U_i) - D^\alpha \nabla u_i \cdot \nabla U_i\big) \mathrm{d}x\mathrm{d}t \\ &\quad + \int_0^\tau \int_{\mathbb{R}^d} D^\alpha u_i \big(D^\alpha(u_i \Delta U_i) - D^\alpha u_i \Delta U_i\big) \mathrm{d}x\mathrm{d}t, \end{aligned} \tag{2.22}$$

noting that the second terms in both integrals are the same (with different signs) because of

$$-\int_{\mathbb{R}^d} D^\alpha u_i D^\alpha \nabla u_i \cdot \nabla U_i \mathrm{d}x = -\frac{1}{2}\int_{\mathbb{R}^d} \nabla (D^\alpha u_i)^2 \cdot \nabla U_i \mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}^d} (D^\alpha u_i)^2 \Delta U_i \mathrm{d}x.$$

Moreover, the last integral in (2.22) vanishes since $\Delta U_i = -d$. In the first integral of the right-hand side of (2.22), the first-order derivative of U_i cancels, while the second-order derivative equals $\partial^2 U_i/\partial x_j \partial x_k = -\delta_{jk}$ and all higher-order derivatives of U_i vanish. Then a straightforward computation leads to

$$I_7 = -d \int_0^\tau \int_{\mathbb{R}^d} (D^\alpha u_i)^2 \mathrm{d}x\mathrm{d}t \le 0.$$

For the estimates of I_8 and I_9, we need a smallness condition on f and its derivatives. We apply Young's inequality and Lemma 2.12 to estimate the (more delicate) term I_9:

$$\begin{aligned} I_9 &\le \frac{\sigma_i}{4} \int_0^\tau \|\nabla D^\alpha u_i\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t + C(\sigma_i) \int_0^\tau \|D^\alpha(u_i \nabla K_i(u))\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t \\ &\le \frac{\sigma_i}{4} \int_0^\tau \|\nabla D^\alpha u_i\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t + C \int_0^\tau \big(\|u_i\|_{L^\infty(\mathbb{R}^d)} \|D^m \nabla K_i(u)\|_{L^2(\mathbb{R}^d)} \\ &\quad + \|D^m u_i\|_{L^2(\mathbb{R}^d)} \|\nabla K_i(u)\|_{L^\infty(\mathbb{R}^d)}\big)^2 \mathrm{d}t. \end{aligned}$$

Estimate (2.21) shows that $f_\eta = f$ and $|f'| \le \varepsilon$ on I. Then, by similar arguments leading to (2.20),

$$\begin{aligned}\|\nabla K_i(u)\|_{L^\infty(\mathbb{R}^d)} &\le A\sum_{j=1}^n \|f_\eta'(B_{ij}^\eta * u_j)\|_{L^\infty(\mathbb{R}^d)}\|\nabla u_j\|_{L^\infty(\mathbb{R}^d)} \\ &\le A\varepsilon\|\nabla u\|_{L^\infty(\mathbb{R}^d)} \le \varepsilon A C_s\|\nabla u\|_{H^s(\mathbb{R}^d)}.\end{aligned}$$

Moreover, using Lemma 2.13, the embedding $H^s(\mathbb{R}^d) \hookrightarrow W^{1,\infty}(\mathbb{R}^d)$, and $m \le s$,

$$\begin{aligned}\|D^m\nabla K_i(u)\|_{L^2(\mathbb{R}^d)} &\le A\sum_{j=1}^n \|\nabla u_j\|_{L^\infty(\mathbb{R}^d)}\|D^m(f_\eta'(B_{ij}^\eta * u_j))\|_{L^2(\mathbb{R}^d)} \\ &\le C\sum_{j=1}^n \|\nabla u_j\|_{H^s(\mathbb{R}^d)}\|f''\|_{C^{m-1}(I)}\|B_{ij}^\eta * u_j\|_{L^\infty(\mathbb{R}^d)}^{m-1}\|B_{ij}^\eta * D^m u_j\|_{L^2(\mathbb{R}^d)} \\ &\le \varepsilon C\|\nabla u\|_{H^s(\mathbb{R}^d)}\|u\|_{L^\infty(\mathbb{R}^d)}^{m-1}\|D^m u\|_{L^2(\mathbb{R}^d)} \le \varepsilon C\|\nabla u\|_{H^s(\mathbb{R}^d)}\|u_0\|_{H^s(\mathbb{R}^d)}^s,\end{aligned}$$

recalling definition (2.11) of the interval I. Consequently, the estimate for I_9 becomes

$$I_9 \le \frac{\sigma_i}{4}\int_0^\tau \|\nabla D^\alpha u_i\|_{L^2(\mathbb{R}^d)}^2 \mathrm{d}t + C\varepsilon^2\|u_0\|_{H^s(\mathbb{R}^d)}^{2s}\int_0^\tau \|\nabla u\|_{H^s(\mathbb{R}^d)}^2 \mathrm{d}t.$$

The term I_8 is treated in a similar way, resulting in

$$I_8 \le \frac{\sigma_i}{4}\int_0^\tau \|\nabla D^\alpha u_i\|_{L^2(\mathbb{R}^d)}^2 \mathrm{d}t + C\varepsilon^2\|u_0\|_{H^s(\mathbb{R}^d)}^{2s}\int_0^\tau \|\nabla u\|_{H^s(\mathbb{R}^d)}^2 \mathrm{d}t.$$

Set $\sigma_{\min} = \min_{i=1,\ldots,n}\sigma_i > 0$. We conclude from (2.19) after summation over $|\alpha| \le s$ and $i = 1,\ldots,n$ that

$$\|u(\tau)\|_{H^s(\mathbb{R}^d)}^2 + \big(\sigma_{\min} - C\varepsilon^2\|u_0\|_{H^s(\mathbb{R}^d)}^s\big)\int_0^\tau \|\nabla u\|_{H^s(\mathbb{R}^d)}^2 \mathrm{d}t \le \|u_0\|_{H^s(\mathbb{R}^d)}^2.$$

Thus, for sufficiently small $\varepsilon > 0$, we arrive at the desired estimate uniform in η.
Step 3: Global existence and uniqueness. We have proved that $\|u(\tau)\|_{H^s(\mathbb{R}^d)} \le \|u_0\|_{H^s(\mathbb{R}^d)}$ for $0 < \tau \le T$ for some sufficiently small $T > 0$. The value for T does not depend on the solution. Thus, we can use $u(T)$ as an initial datum and solve the equation in $[T, 2T]$. Repeating this argument leads to a global solution. The uniqueness of a solution follows after standard estimates, based on the global Lipschitz continuity of f_η and f_η' (see the calculations for I_4, I_5, and I_6) and choosing $\varepsilon > 0$ sufficiently small.

2.4 Proof of Theorem 2.3

We show the global existence of smooth solutions to the local system (2.1) and an error estimate for the difference of the solutions to (2.1) and (2.7), respectively. First, we prove that a solution u_η to (2.7) converges to a solution u to (2.1) in a certain sense. Then we prove the error bound in Theorem 2.3 by estimating the difference $u_\eta - u$. The key of the proof is the estimate of the difference $f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})$.

Step 1. Existence and uniqueness of solutions. Let u_η be a smooth solution to (2.7) and let $\phi \in C_0^\infty(\mathbb{R}^d)$ with $\mathrm{supp}(\phi) \subset B_R$, $\zeta \in C^0([0,T])$ be test functions, where B_R is a ball around the origin with radius $R > 0$. Then the weak formulation of (2.7) reads as

$$\begin{aligned}\int_0^T \langle \partial_t u_{\eta.i}, \phi\rangle \zeta(t)\mathrm{d}t = &-\int_0^T \int_{\mathbb{R}^d} u_{\eta,i}\nabla U_i \cdot \nabla\phi\zeta(t)\mathrm{d}x\mathrm{d}t \\ &- \int_0^T \int_{\mathbb{R}^d} \big(\sigma_i \nabla u_{\eta,i} + \nabla(u_{\eta,i}K_i(u_\eta))\big)\cdot\nabla\phi\zeta(t)\mathrm{d}x\mathrm{d}t,\end{aligned} \tag{2.23}$$

where $\langle\cdot,\cdot\rangle$ is the duality pairing between $H^{-1}(\mathbb{R}^d)$ and $H^1(\mathbb{R}^d)$ and $K_i(u) = \sum_{j=1}^n f_\eta(B_{ij}^\eta * u_j)$. We want to perform the limit $\eta \to 0$. By the uniform estimate of Theorem 2.2, there exists a subsequence, which is not relabelled, such that $u_\eta \rightharpoonup u$ weakly in $L^2(0,T;H^{s+1}(\mathbb{R}^d))$ and weakly* in $L^\infty(0,T;H^s(\mathbb{R}^d)) \subset L^\infty(0,T;L^\infty(\mathbb{R}^d))$ as $\eta \to 0$. Our aim is to prove that u is a weak solution to (2.1).
It follows from the proof of Lemma 7 in [26] that

$$B_{ij}^\eta * \nabla u_{\eta,j} \rightharpoonup a_{ij}\nabla u_j \quad \text{weakly in } L^2(0,T;L^2(\mathbb{R}^d)).$$

We claim that $f_\eta(B_{ij}^\eta * u_{\eta,j}) \to f(a_{ij}u_j)$ strongly in $L^2(0,T;L^2(B_R))$. First, we observe that $u \in L^\infty(0,T;L^\infty(\mathbb{R}^d))$. The weak formulation (2.23) gives

$$\begin{aligned}\|\partial_t u_{\eta,i}\|_{L^2(0,T;H^{-1}(B_R))} \le\ & \|u_{\eta,i}\|_{L^2(0,T;L^2(\mathbb{R}^d))}\|\nabla U_i\|_{L^\infty(B_R)} + \sigma_i\|\nabla u_{\eta,i}\|_{L^2(0,T;L^2(\mathbb{R}^d))} \\ &+ \|\nabla u_{\eta,i}\|_{L^2(0,T;L^2(\mathbb{R}^d))}\|K_i(u_\eta)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\ &+ \|u_{\eta,i}\|_{L^2(0,T;L^2(\mathbb{R}^d))}\|\nabla K_i(u_\eta)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}.\end{aligned}$$

Because of

$$\begin{aligned}\|K_i(u_\eta)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} &\le \sum_{j=1}^n \|f_\eta(B_{ij}^\eta * u_{\eta,j})\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\|f\|_{L^\infty(I)}, \\ \|\nabla K_i(u_\eta)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} &\le \sum_{j=1}^n \|f_\eta'(B_{ij}^\eta * u_{\eta,j})\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\|B_{ij}^\eta * \nabla u_{\eta,j}\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\ &\le C\|f'\|_{L^\infty(I)}\|\nabla u_\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\|u_0\|_{H^s(\mathbb{R}^d)},\end{aligned}$$

we obtain a uniform bound for $\partial_t u_{\eta,i}$ in $L^2(0,T;H^{-1}(B_R))$ (the bound might depend on R). In particular, up to a subsequence, as $\eta \to 0$,

$$\partial_t u_{\eta,i} \rightharpoonup \partial_t u_i \quad \text{weakly in } L^2(0,T;H^{-1}(B_R)).$$

Since u_η is uniformly bounded in $L^2(0,T;H^1(B_R))$, the Aubin–Lions lemma implies the existence of a subsequence (not relabelled) such that

$$u_{\eta,i} \to u_i \quad \text{strongly in } L^2(0,T;L^2(B_R)).$$

We use the Lipschitz continuity of $f = f_\eta$ on I to infer that

$$\|f_\eta(B_{ij}^\eta * u_{\eta,j}) - f(a_{ij}u_j)\|_{L^2(0,T;L^2(B_R))}$$

$$\leq C\|B_{ij}^\eta * (u_{\eta,j} - u_j) + B_{ij}^\eta * u_j - a_{ij}u_j\|_{L^2(0,T;L^2(B_R))}$$
$$\leq C\|B_{ij}^\eta\|_{L^1(\mathbb{R}^d)}\|u_{\eta,j} - u_j\|_{L^2(0,T;L^2(B_R))} + \|B_{ij}^\eta * u_j - a_{ij}u_j\|_{L^2(0,T;L^2(B_R))} \to 0.$$

This shows the claim. In a similar way, it follows from the Lipschitz continuity of f_η' that $f_\eta'(B_{ij}^\eta * u_{\eta,j}) \to f'(a_{ij}u_j)$ strongly in $L^2(0,T;L^2(B_R))$.
The previous convergences allow us to perform the limit $\eta \to 0$ in (2.23), leading to

$$\int_0^T \langle \partial_t u_i, \phi\rangle \zeta(t)\mathrm{d}t = -\int_0^T \int_{\mathbb{R}^d} u_i \nabla U_i \cdot \nabla\phi\zeta(t)\mathrm{d}x\mathrm{d}t - \int_0^T \int_{\mathbb{R}^d} \nabla F_i(u)\cdot\nabla\phi\zeta(t)\mathrm{d}x\mathrm{d}t,$$

where $F_i(u) = u_i(\sigma_i + \sum_{j=1}^n f(a_{ij}u_j))$. Moreover, $u_i(0) = u_{0,i}$ in B_R for any $R > 0$. Thus, u is a weak solution to (2.1). Standard estimates show that u is the unique solution, again choosing $\varepsilon > 0$ sufficiently small.
Step 2: Convergence rate. We take the difference of (2.7) and (2.1), multiply the resulting equation by $u_{\eta,i} - u_i$, integrate over $(0,\tau)\times\mathbb{R}^d$ for any $\tau > 0$, and integrate by parts:

$$\frac{1}{2}\int_{\mathbb{R}^d}(u_{\eta,i} - u_i)(\tau)^2\mathrm{d}x + \sigma_i\int_0^\tau\int_{\mathbb{R}^d}|\nabla(u_{\eta,i} - u_i)|^2\mathrm{d}x\mathrm{d}t = \frac{1}{2}\int_0^\tau\int_{\mathbb{R}^d}\Delta U_i(u_{\eta,i} - u_i)^2\mathrm{d}x\mathrm{d}t$$
$$-\int_0^\tau\int_{\mathbb{R}^d}\nabla\sum_{j=1}^n\big(u_{\eta,i}f_\eta(B_{ij}^\eta * u_{\eta,j}) - u_i f(a_{ij}u_j)\big)\cdot\nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t. \quad (2.24)$$

The first integral on the right-hand side is nonpositive since $\Delta U_i = -d$. We split the second integral into three parts:

$$-\int_0^\tau\int_{\mathbb{R}^d}\sum_{j=1}^n\nabla\big(u_{\eta,i}f_\eta(B_{ij}^\eta * u_{\eta,j}) - u_i f(a_{ij}u_j)\big)\cdot\nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t = J_1 + J_2 + J_3, \quad (2.25)$$

where

$$J_1 = -\int_0^\tau\int_{\mathbb{R}^d}\sum_{j=1}^n\nabla\big((u_{\eta,i} - u_i)f_\eta(B_{ij}^\eta * u_{\eta,j})\big)\cdot\nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t,$$
$$J_2 = -\int_0^\tau\int_{\mathbb{R}^d}\sum_{j=1}^n\nabla\big(u_i\big(f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})\big)\big)\cdot\nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t,$$
$$J_3 = -\int_0^\tau\int_{\mathbb{R}^d}\sum_{j=1}^n\nabla\big(u_i\big(f_\eta(a_{ij}u_{\eta,j}) - f(a_{ij}u_j)\big)\big)\cdot\nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t.$$

We start with the estimate of J_1. The families $(B_{ij}^\eta * u_{\eta,j})$ and $(B_{ij}^\eta * \nabla u_{\eta,j})$ are bounded in $L^\infty(0,T;L^\infty(\mathbb{R}^d))$. Using $\|f_\eta\|_{L^\infty(I)} = \|f\|_{L^\infty(I)} \leq \varepsilon$ and Young's inequality, we have

$$J_1 \leq \|f_\eta(B_{ij}^\eta * u_{\eta,j})\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\int_0^\tau\|\nabla(u_{\eta,i} - u_i)\|_{L^2(\mathbb{R}^d)}^2\mathrm{d}t$$
$$+\int_0^\tau\|u_{\eta,i} - u_i\|_{L^2(\mathbb{R}^d)}\|f_\eta'(B_{ij}^\eta * u_{\eta,j})\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}$$
$$\times\|B_{ij}^\eta * \nabla u_{\eta,j}\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\|\nabla(u_{\eta,i} - u_i)\|_{L^2(\mathbb{R}^d)}\mathrm{d}t$$

$$\leq \left(\frac{\sigma_i}{4}+\varepsilon\right)\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C(\sigma_i)\int_0^\tau \|u_{\eta,i}-u_i\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t. \tag{2.26}$$

Next, we estimate $J_2 = J_{21} + J_{22}$, where

$$J_{21} = -\int_0^\tau\int_{\mathbb{R}^d}\nabla u_i \sum_{j=1}^n \big(f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})\big)\cdot\nabla(u_{\eta,i}-u_i)\mathrm{d}x\mathrm{d}t,$$

$$J_{22} = -\int_0^\tau\int_{\mathbb{R}^d} u_i \sum_{j=1}^n \big(f_\eta'(B_{ij}^\eta * u_{\eta,j})B_{ij}^\eta * \nabla u_{\eta,j} - f_\eta'(a_{ij}u_{\eta,j})a_{ij}\nabla u_{\eta,j}\big)\cdot\nabla(u_{\eta,i}-u_i)\mathrm{d}x\mathrm{d}t.$$

It follows that

$$\begin{aligned} J_{21} \leq & \|\nabla u_i\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\times \\ & \times \sum_{j=1}^n\int_0^\tau \|f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})\|_{L^2(\mathbb{R}^d)}\|\nabla(u_{\eta,i}-u_i)\|_{L^2(\mathbb{R}^d)}\mathrm{d}t \\ \leq & \frac{\sigma_i}{8}\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\sum_{j=1}^n\int_0^\tau \|f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t. \end{aligned}$$

Since both $B_{ij}^\eta * u_{\eta,j}$ and $u_{\eta,j}$ are uniformly bounded in $L^\infty(0,T;L^\infty(\mathbb{R}^d))$, we can choose $\eta > 0$ sufficiently small such that $f = f_\eta$ on I. On that interval, f is Lipschitz continuous uniformly in η. We use this information in

$$\left|\int_{\mathbb{R}^d}\big(f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})\big)g(x)\mathrm{d}x\right| \leq C\int_{\mathbb{R}^d}\big|B_{ij}^\eta * u_{\eta,j} - a_{ij}u_{\eta,j}\big||g(x)|\mathrm{d}x,$$

where $g \in L^2(\mathbb{R}^d)$. Recalling that $\operatorname{supp}(B_{ij}^\eta)\subset B_\eta(0)$ and $a_{ij} = \int_{B_\eta} B_{ij}^\eta \mathrm{d}x$, we obtain

$$\begin{aligned} &\left|\int_{\mathbb{R}^d}\big(f_\eta(B_{ij}^\eta * u_{\eta,j}) - f_\eta(a_{ij}u_{\eta,j})\big)g(x)\mathrm{d}x\right| \\ &\quad\leq C\int_{\mathbb{R}^d}\left|\int_{B_\eta} B_{ij}^\eta(y)\big(u_{\eta,j}(x-y) - u_{\eta,j}(x)\big)\mathrm{d}y\right||g(x)|\mathrm{d}x \\ &\quad\leq C\int_{\mathbb{R}^d}\int_{B_\eta}|B_{ij}^\eta(y)|\left(\int_0^1 |\nabla u_{\eta,j}(x-ry)|\eta\mathrm{d}r\right)\mathrm{d}y|g(x)|\mathrm{d}x \\ &\quad= C\eta\int_0^1\int_{B_\eta}|B_{ij}^\eta(y)|\left(\int_{\mathbb{R}^d}|\nabla u_{\eta,j}(x-ry)||g(x)|\mathrm{d}x\right)\mathrm{d}y\mathrm{d}r \\ &\quad\leq C\eta\int_0^1\int_{B_\eta}|B_{ij}^\eta(y)|\|\nabla u_{\eta,j}(\cdot - ry)\|_{L^2(\mathbb{R}^d)}\|g\|_{L^2(\mathbb{R}^d)}\mathrm{d}y\mathrm{d}r \\ &\quad\leq C\eta\int_{B_\eta}|B_{ij}^\eta(y)|\mathrm{d}y\|\nabla u_{\eta,j}\|_{L^2(\mathbb{R}^d)}\|g\|_{L^2(\mathbb{R}^d)} \leq C\eta\|g\|_{L^2(\mathbb{R}^d)}. \end{aligned}$$

By duality, we find that

$$J_{21} \leq \frac{\sigma_i}{8}\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2.$$

The integral J_{22} is split into $J_{22} = J_{221} + J_{222}$, where

$$J_{221} = -\int_0^\tau \int_{\mathbb{R}^d} u_i \sum_{j=1}^n f'_\eta(B^\eta_{ij} * u_{\eta,j})\big(B^\eta_{ij} * \nabla u_{\eta,j} - a_{ij}\nabla u_{\eta,j}\big) \cdot \nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t,$$
$$J_{222} = -\int_0^\tau \int_{\mathbb{R}^d} u_i \sum_{j=1}^n \big(f'_\eta(B^\eta_{ij} * u_{\eta,j}) - f'_\eta(a_{ij}u_{\eta,j})\big)a_{ij}\nabla u_{\eta,j} \cdot \nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t.$$

We infer from the uniform boundedness of $B^\eta_{ij} * u_{\eta,j}$ in $L^\infty(0,T;L^\infty(\mathbb{R}^d))$ and the fact that $f'_\eta = f'$ on I for sufficiently small $\eta > 0$ that

$$\begin{aligned} J_{221} &\le \frac{\sigma_i}{16}\int_0^\tau \|\nabla(u_{\eta,i} - u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\sum_{j=1}^n \int_0^T \|B^\eta_{ij} * \nabla u_{\eta,j} - a_{ij}\nabla u_{\eta,j}\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t \\ &\le \frac{\sigma_i}{16}\int_0^\tau \|\nabla(u_{\eta,i} - u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2\sum_{j=1}^n \int_0^\tau \|D^2 u_{\eta,j}\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t, \end{aligned}$$

where we estimated the difference $B^\eta_{ij} * \nabla u_{\eta,j} - a_{ij}\nabla u_{\eta,j}$ similarly as for J_{21}. Furthermore, the Lipschitz continuity of $f'_\eta = f'$ on I leads to

$$\begin{aligned} J_{222} &\le C\sum_{j=1}^n \int_0^\tau \|u_i\|_{L^\infty(\mathbb{R}^d)}\|B^\eta_{ij} * u_{\eta,j} - a_{ij}u_{\eta,j}\|_{L^2(\mathbb{R}^d)}\|\nabla u_{\eta,j}\|_{L^\infty(\mathbb{R}^d)}\|\nabla(u_{\eta,i} - u_i)\|_{L^2(\mathbb{R}^d)}\mathrm{d}t \\ &\le \frac{\sigma_i}{16}\int_0^\tau \|\nabla(u_{\eta,i} - u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2\sum_{j=1}^n \int_0^\tau \|\nabla u_{\eta,j}\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t. \end{aligned}$$

Summarizing these estimates, we infer that

$$J_{22} \le \frac{\sigma_i}{8}\int_0^\tau \|\nabla(u_{\eta,i} - u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2,$$

and combining the estimate for J_{21} and J_{22},

$$J_2 \le \frac{\sigma_i}{4}\int_0^\tau \|\nabla(u_{\eta,i} - u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2. \tag{2.27}$$

It remains to estimate $J_3 = J_{31} + J_{32}$, where

$$J_{31} = -\int_0^\tau \int_{\mathbb{R}^d} \sum_{j=1}^n \big(f_\eta(a_{ij}u_{\eta,j}) - f(a_{ij}u_j)\big)\nabla u_i \cdot \nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t,$$
$$J_{32} = -\int_0^\tau \int_{\mathbb{R}^d} u_i \sum_{j=1}^n \big(f'_\eta(a_{ij}u_{\eta,j})a_{ij}\nabla u_{\eta,j} - f'(a_{ij}u_j)a_{ij}\nabla u_j\big) \cdot \nabla(u_{\eta,i} - u_i)\mathrm{d}x\mathrm{d}t.$$

Similar arguments as above yield

$$J_{31} \le \frac{\sigma_i}{8}\int_0^\tau \|\nabla(u_{\eta,i} - u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\int_0^\tau \|\nabla u_i\|^2_{L^\infty(\mathbb{R}^d)}\|u_\eta - u\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t$$

$$\leq \frac{\sigma_i}{8}\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\int_0^\tau \|u_\eta - u\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t.$$

The second term J_{32} is again split into two parts, $J_{32} = J_{321} + J_{322}$, where

$$J_{321} = -\int_0^\tau\int_{\mathbb{R}^d} u_i \sum_{j=1}^n \big(f'_\eta(a_{ij}u_{\eta,j}) - f'_\eta(a_{ij}u_j)\big)a_{ij}\nabla u_{\eta,j}\cdot\nabla(u_{\eta,i}-u_i)\mathrm{d}x\mathrm{d}t,$$
$$J_{322} = -\int_0^\tau\int_{\mathbb{R}^d} u_i \sum_{j=1}^n a_{ij}\big(f'_\eta(a_{ij}u_j)\nabla u_{\eta,j} - f'(a_{ij}u_j)\nabla u_j\big)\cdot\nabla(u_{\eta,i}-u_i)\mathrm{d}x\mathrm{d}t.$$

Using the Lipschitz continuity again, $f'_\eta = f'$ on I, and $|f'|\leq\varepsilon$, we deduce that

$$J_{321} \leq C\|u_i\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\int_0^\tau\sum_{j=1}^n \|\nabla u_{\eta,j}\|_{L^\infty(\mathbb{R}^d)}\|u_{\eta,j}-u_j\|_{L^2(\mathbb{R}^d)}\|\nabla(u_{\eta,i}-u_i)\|_{L^2(\mathbb{R}^d)}\mathrm{d}t$$
$$\leq \frac{\sigma_i}{8}\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\int_0^\tau \|u_\eta - u\|^2_{L^2(\mathbb{R}^d)},$$
$$J_{322} \leq C\int_0^\tau\sum_{j=1}^n \|f'(a_{ij}u_j)\|_{L^\infty(\mathbb{R}^d)}\|\nabla(u_{\eta,j}-u_j)\|_{L^2(\mathbb{R}^d)}\|\nabla(u_{\eta,i}-u_i)\|_{L^2(\mathbb{R}^d)}\mathrm{d}t$$
$$\leq C\varepsilon\int_0^\tau \|\nabla(u_\eta-u)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t.$$

This shows that

$$J_{32} \leq \left(\frac{\sigma_i}{8}+C\varepsilon\right)\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\int_0^\tau \|u_\eta - u\|^2_{L^2(\mathbb{R}^d)}.$$

Summarizing the estimate for J_{31} and J_{32}, we arrive at

$$J_3 \leq \left(\frac{\sigma_i}{4}+C\varepsilon\right)\int_0^\tau \|\nabla(u_{\eta,i}-u_i)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\int_0^\tau \|u_\eta - u\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t. \tag{2.28}$$

Finally, putting together the estimates (2.26), (2.27), and (2.28), we infer from (2.25) that

$$\left|\int_0^\tau\int_{\mathbb{R}^d}\sum_{j=1}^n \nabla\big(u_{\eta,i}f_\eta(B^\eta_{ij}*u_{\eta,j}) - u_i f(a_{ij}u_j)\big)\cdot\nabla(u_{\eta,i}-u_i)\mathrm{d}x\mathrm{d}t\right|$$
$$\leq \left(\frac{3\sigma_i}{4}+C\varepsilon\right)\int_0^\tau \|\nabla(u_{\eta,j}-u_j)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\int_0^\tau \|u_\eta - u\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2.$$

This is the desired estimate for the last integral in (2.24). We conclude for sufficiently small $\varepsilon > 0$ and after summation over $i = 1,\ldots,n$ that

$$\|(u_\eta-u)(\tau)\|^2_{L^2(\mathbb{R}^d)} + \sigma_{\min}C\int_0^\tau \|\nabla(u_\eta-u)\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t \leq C\int_0^\tau \|u_\eta - u\|^2_{L^2(\mathbb{R}^d)}\mathrm{d}t + C\eta^2.$$

The proof ends after applying Gronwall's inequality.

2.5 Links between the SDEs and PDEs

We show that the density function $\widehat{u}$ from Proposition 2.4 coincides with the unique weak solution u to (2.1).

Theorem 2.7. *Let the assumptions of Theorem 2.3 hold. Let $\widehat{X}_i$ for $i = 1, \dots, n$ be the square-integrable process solving (2.8) with density function $\widehat{u}_i$ and let u_i be the unique weak solution to (2.1). Then $\widehat{u} = (\widehat{u}_1, \dots, \widehat{u}_n)$ solves the linear equation*

$$\partial_t \widehat{u}_i = \operatorname{div}(\widehat{u}_i \nabla U_i) + \Delta\bigg(\sigma_i \widehat{u}_i + \widehat{u}_i \sum_{j=1}^n f(a_{ij}u_j)\bigg) \quad \text{in } \mathbb{R}^d, \ i = 1, \dots, n, \tag{2.29}$$

in the weak integrable sense, i.e.

$$\begin{aligned} &\int_{\mathbb{R}^d} \widehat{u}_i(t)\phi(t)\mathrm{d}x - \int_{\mathbb{R}^d} u_{0,i}\phi(0)\mathrm{d}x - \int_0^t \int_{\mathbb{R}^d} \widehat{u}_i \partial_t \phi \mathrm{d}x\mathrm{d}s \\ &\quad = -\int_0^t \int_{\mathbb{R}^d} \widehat{u}_i \nabla U_i \cdot \nabla\phi \mathrm{d}x\mathrm{d}t + \int_0^t \int_{\mathbb{R}^d} \widehat{u}_i \bigg(\sigma_i + \sum_{j=1}^n f(a_{ij}u_j)\bigg)\Delta\phi \mathrm{d}x\mathrm{d}s \end{aligned}$$

for all $\phi \in C_0^\infty([0,\infty) \times \mathbb{R}^d)$ and $t > 0$, where we assume that the initial datum $\widehat{u}_i(0) = u_{0,i}$ fulfils

$$\int_{\mathbb{R}^d} u_{0,i}(x)\mathrm{d}x = 1, \quad \int_{\mathbb{R}^d} u_{0,i}(x)|x|^2 \mathrm{d}x < \infty. \tag{2.30}$$

Additionally, $\widehat{u} = u$ in $(0,\infty) \times \mathbb{R}^d$, $u_i \geq 0$, and (2.30) is fulfilled for u_i instead of $u_{0,i}$ for almost all $t > 0$ and all $i = 1, \dots, n$.

Proof. Since $\widehat{X}_{k,i}$ depends on k only via the initial data ξ_i^k with the same law $u_{0,i}$, we can omit the index k. Let $\phi \in C_0^\infty([0,\infty) \times \mathbb{R}^d)$ and set $F_i(u) = \sigma_i + \sum_{j=1}^n f(a_{ij}u_j)$. By Itô's lemma, we obtain

$$\begin{aligned} \phi(t, \widehat{X}_i(t)) &= \phi(0, \xi_i) + \int_0^t \partial_t \phi(s, \widehat{X}_i(s))\mathrm{d}s - \int_0^t \nabla U_i(s) \cdot \nabla\phi(s, \widehat{X}_i(s))\mathrm{d}s \\ &\quad + \int_0^t F_i\big(u(\widehat{X}_i(s))\big)\Delta\phi(s, \widehat{X}_i(s))\mathrm{d}s + \int_0^t F_i\big(u(\widehat{X}_i(s))\big)^{1/2}\nabla\phi(s, X(s)) \cdot \mathrm{d}W_i(s). \end{aligned} \tag{2.31}$$

We claim that the density function $\widehat{u}_i : [0,\infty) \to \mathcal{P}_2(\mathbb{R}^d)$, where $\mathcal{P}_2(\mathbb{R}^d)$ is the space of all density functions with finite second moment, is continuous with respect to the 2-Wasserstein distance W_2. Indeed, since $\widehat{X}_i$ is square-integrable, we have $\widehat{u}_i(t) \in \mathcal{P}_2(\mathbb{R}^d)$ for almost all $t > 0$ and the limit $s \to t$ in the Wasserstein distance leads to

$$\begin{aligned} W_2(\widehat{u}_i(t), \widehat{u}_i(s)) &= \inf\big\{\big(\mathbb{E}(|Y_t - Y_s|^2)\big)^{1/2} : \ \mathrm{Law}(Y_t) = \widehat{u}_i(t), \ \mathrm{Law}(Y_s) = \widehat{u}_i(s)\big\} \\ &\leq \big(\mathbb{E}(|\widehat{X}_i(t) - \widehat{X}_i(s)|^2)\big)^{1/2} \to 0, \end{aligned}$$

using the facts that $\widehat{X}_i$ is continuous in time and has bounded second moments. This shows the claim. We conclude that the point evaluation $\widehat{u}_i(t)$ is well defined.

The previous argumentation shows that we can apply the expectation to (2.31) to obtain

$$\int_{\mathbb{R}^d} \widehat{u}_i(t)\phi(t)\mathrm{d}x = \int_{\mathbb{R}^d} u_{0,i}\phi(0)\mathrm{d}x + \int_0^t \int_{\mathbb{R}^d} \widehat{u}_i(s)\partial_t\phi(s)\mathrm{d}x\mathrm{d}s \\ - \int_0^t \int_{\mathbb{R}^d} \widehat{u}_i(s)\nabla U_i \cdot \nabla\phi(s)\mathrm{d}x\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} \widehat{u}_i(s)F_i(u(s))\Delta\phi(s)\mathrm{d}x\mathrm{d}s.$$

This is the very weak formulation of (2.29), showing the first part of the theorem.
Next, we verify that the solution to (2.29) is unique. More precisely, we take $u_0 = 0$ and show that $\widehat{u}_i(t) = 0$ for almost all $t > 0$. The statement is usually proved by a duality argument. However, the coefficients of the dual problem associated to (2.29) are not regular enough such that we need to regularize it. As the proof is rather standard but tedious, we only sketch the arguments. Let χ_k be a family of mollifiers and consider the regularized dual backward problem on the ball B_R around the origin with radius $R > 0$:

$$\partial_t w_{k,R} - \nabla U_i \cdot \nabla w_{k,R} + (\chi_k * F_i(u))\Delta w_{k,R} = 0 \quad \text{in } B_R,\ 0 < s < t, \\ w_{k,R} = 0 \quad \text{on } \partial B_R, \quad w_{k,R}(t) = g \in C_0^\infty(B_R) \quad \text{in } B_R.$$

We extend the unique smooth solution $w_{k,R}$ to the whole space by setting $w_{k,R} = 0$ on $\mathbb{R}^d \backslash B_R$. Since the extension may be not smooth, we choose a cut-off function $\psi_R \in C^\infty(\mathbb{R}^d)$ and use $w_{k,R}\psi_R$ as an admissible test function in the very weak formulation of (2.29). Standard estimations give bounds for $w_{k,R}$ uniform in k and R. Then, passing to the limit $k \to \infty$, $R \to \infty$ in the weak formulation shows that $\int_{\mathbb{R}^d} g(x)\widehat{u}_i(s,x)\mathrm{d}x = 0$, and since g was arbitrary, we conclude that $\widehat{u}_i(s) = 0$ for $0 < s < t$.
The weak solution u to (2.1) is also a very weak solution to (2.29). Therefore, by the previous uniqueness result, $\widehat{u} = u$. □

Similar arguments lead to the following result that relates the solutions $\overline{u}_\eta$ and u_η.

Theorem 2.8. *Let the assumptions of Theorem 2.2 hold and let $\eta > 0$. Let $\overline{X}_{k,i}^\eta$ for $i = 1, \dots, n$ and $k = 1, \dots, N$ be the square-integrable process solving (2.6) with density function $\overline{u}_{\eta,i}$. Then $\overline{u}_\eta = (\overline{u}_{\eta,1}, \dots, \overline{u}_{\eta,n})$ solves the linear problem*

$$\partial_t \overline{u}_{\eta,i} = \operatorname{div}(\overline{u}_{\eta,i}\nabla U_i) + \Delta\Bigg(\sigma_i\overline{u}_{\eta,i} + \overline{u}_{\eta,i}\sum_{j=1}^n f_\eta(B_{ij}^\eta * u_{\eta,j})\Bigg) \quad \text{in } \mathbb{R}^d,\ i = 1, \dots, n,$$

with initial datum $\overline{u}_{\eta,i}(0) = u_{0,i}$, which fulfils (2.30), where $u_{\eta,i}$ is the unique weak solution to (2.7). Then $\overline{u}_\eta = u_\eta$ in $(0,\infty) \times \mathbb{R}^d$, $u_{\eta,i} \geq 0$, and

$$\int_{\mathbb{R}^d} u_{\eta,i}(x,t)\mathrm{d}x = 1, \qquad \int_{\mathbb{R}^d} u_{\eta,i}(x,t)|x|^2\mathrm{d}x < \infty$$

for almost all $t > 0$ and all $i = 1, \dots, n$.

2.6 Proof of Theorem 2.5

The proof is split into two parts. We estimate first the square mean error of the difference $X_{k,i}^{N,\eta} - \overline{X}_{k,i}^{\eta}$, where $\overline{X}_{k,i}^{\eta}$ is the solution to the intermediate system (2.6). In fact, this error bound is a generalization of a result due to [113]. Essential for this step are the facts that the Lipschitz constant of B_{ij}^{η} is of order η^{-d-1}, while the Lipschitz constant of f_η is of order $\eta^{-\alpha}$. Second, we estimate the square mean error of the difference $\overline{X}_{k,i}^{\eta} - \widehat{X}_{k,i}$, based on an estimate of $f_\eta(B_{ij}^{\eta} * u_j) - f_\eta(a_{ij}u_j)$ in L^2, which is of the order of $\eta^{1-\alpha}$.

Lemma 2.9. *Let $X_{k,i}^{N,\eta}$ and $\overline{X}_{k,i}^{\eta}$ be the solutions to (2.5) and (2.8), respectively, in the sense of Proposition 2.4. Under the assumptions of Theorem 2.5, there exists $\delta > 0$, depending on n, σ_{min}, and T, such that if $\eta^{-2(d+1+\alpha)} \le \delta \log N$, where $\alpha \ge 0$ is fixed in Assumption (A4), we have*

$$\sup_{k=1,\dots,N} \mathbb{E}\bigg(\sum_{i=1}^{n} \sup_{0<s<T} \big|(X_{k,i}^{N,\eta} - \overline{X}_{k,i}^{\eta})(s)\big|^2\bigg) \le C(T,n,\sigma_{min}) N^{-1+(T+1)C(n,T,\sigma_{min})\delta},$$

where $C(T,n,\sigma_{min}) > 0$ is a positive constant.

Proof. The process $D_{k,i}^{N,\eta} := X_{k,i}^{N,\eta} - \overline{X}_{k,i}^{\eta}$ solves

$$D_{k,i}^{N,\eta}(s) = E_{1,i}(s) + E_{2,i}(s), \quad 0 \le s \le T, \tag{2.32}$$

where

$$\begin{aligned}
E_{1,i}(s) &= -\int_0^s \big(\nabla U_i(X_{k,i}^{N,\eta}(t)) - \nabla U_i(\overline{X}_{k,i}^{\eta}(t))\big)\mathrm{d}t, \\
E_{2,i}(s) &= \int_0^s (E_{21}(t) - E_{22}(t))\mathrm{d}W_i^k(t), \\
E_{21}(t) &= \bigg(2\sigma_i + 2\sum_{j=1}^{n} f_\eta\bigg(\frac{1}{N}\sum_{\substack{\ell=1 \\ (\ell,j)\neq(k,i)}}^{N} B_{ij}^{\eta}(X_{k,i}^{N,\eta}(t) - X_{\ell,j}^{N,\eta}(t))\bigg)\bigg)^{1/2}, \\
E_{22}(t) &= \bigg(2\sigma_i + 2\sum_{j=1}^{n} f_\eta\big(B_{ij}^{\eta} * u_{\eta,j}(t,\overline{X}_{k,i}^{\eta}(t))\big)\bigg)^{1/2}.
\end{aligned}$$

We use the global Lipschitz continuity of ∇U_i and the Fubini theorem to estimate the first term:

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0<s<T} |E_{1,i}(s)|^2\Big) &\le CT\mathbb{E}\int_0^T \big|(X_{k,i}^{N,\eta} - \overline{X}_{k,i}^{\eta})(s)\big|^2 \mathrm{d}s \\
&\le CT\int_0^T \mathbb{E}\Big(\sup_{0<s<t} |(X_{k,i}^{N,\eta} - \overline{X}_{k,i}^{\eta})(s)|^2\Big)\mathrm{d}t.
\end{aligned}$$

Summing over $i = 1, \dots, n$ and taking the supremum over $k = 1, \dots, N$ leads to

$$\sup_{k=1,\dots,N} \mathbb{E}\bigg(\sum_{i=1}^{n} \sup_{0<s<T} |E_{1,i}(s)|^2\bigg) \le CT\int_0^T \sup_{k=1,\dots,N} \mathbb{E}\Big(\sup_{0<s<t} |(X_{k,i}^{N,\eta} - \overline{X}_{k,i}^{\eta})(s)|^2\Big)\mathrm{d}t. \tag{2.33}$$

Next, we apply the Burkholder–Davis–Gundy inequality [67, Theorem 3.28] to the second term $E_{2,i}$ and use the Lipschitz continuity of $x \mapsto (2\sigma_i + x)^{1/2}$ for $x \geq 0$:

$$\begin{aligned}
\mathbb{E}\Big(\sup_{0<s<T}|E_{2,i}(s)|^2\Big) &\leq C\mathbb{E}\int_0^T (E_{21}(t) - E_{22}(t))^2 \mathrm{d}t \\
&\leq C\mathbb{E}\int_0^T \Bigg[\sum_{j=1}^n f_\eta\Bigg(\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N B_{ij}^\eta(X_{k,i}^{N,\eta}(t) - X_{\ell,j}^{N,\eta}(t))\Bigg) \\
&\qquad - \sum_{j=1}^n f_\eta\big(B_{ij}^\eta * u_{\eta,j}(t,\overline{X}_{k,i}^\eta(t))\big)\Bigg]^2 \mathrm{d}t \\
&= C\mathbb{E}\int_0^T \Bigg[\sum_{j=1}^n (L_j^1(t) + L_j^2(t) + L_j^3(t))\Bigg]^2 \mathrm{d}t \\
&\leq C(n)\mathbb{E}\int_0^T \sum_{j=1}^n \big(L_j^1(t)^2 + L_j^2(t)^2 + L_j^3(t)^2\big)\mathrm{d}t,
\end{aligned} \tag{2.34}$$

where

$$\begin{aligned}
L_j^1(t) &= f_\eta\Bigg(\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N B_{ij}^\eta(X_{k,i}^{N,\eta}(t) - X_{\ell,j}^{N,\eta}(t))\Bigg) - f_\eta\Bigg(\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N B_{ij}^\eta(\overline{X}_{k,i}^\eta(t) - X_{\ell,j}^{N,\eta}(t))\Bigg), \\
L_j^2(t) &= f_\eta\Bigg(\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N B_{ij}^\eta(\overline{X}_{k,i}^\eta(t) - X_{\ell,j}^{N,\eta}(t))\Bigg) - f_\eta\Bigg(\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N B_{ij}^\eta(\overline{X}_{k,i}^\eta(t) - \overline{X}_{\ell,j}^\eta(t))\Bigg), \\
L_j^3(t) &= f_\eta\Bigg(\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N B_{ij}^\eta(\overline{X}_{k,i}^\eta(t) - \overline{X}_{\ell,j}^\eta(t))\Bigg) - f_\eta\big(B_{ij}^\eta * u_{\eta,j}(t,\overline{X}_{k,i}^\eta(t))\big).
\end{aligned}$$

We estimate these three terms separately. By construction, the Lipschitz constant of f_η can be estimated by $L_f \leq \eta^{-\alpha}$. Moreover, the Lipschitz constant of $B_{ij}^\eta(x) = \eta^{-d}B_{ij}(|x|/\eta)$ is computed by $L_B = \max_{i,j=1,\ldots,n}\|\nabla B_{ij}^\eta\|_{L^\infty(\mathbb{R}^d)} \leq C\eta^{-d-1}$. This shows that

$$\begin{aligned}
|L_j^1(t)| &\leq L_f\Bigg|\frac{1}{N}\sum_{\substack{\ell=1\\(\ell,j)\neq(k,i)}}^N \big(B_{ij}^\eta(X_{k,i}^{N,\eta}(t) - X_{\ell,j}^{N,\eta}(t)) - B_{ij}^\eta(\overline{X}_{k,i}^\eta(t) - X_{\ell,j}^{N,\eta}(t))\big)\Bigg| \\
&\leq L_f L_B\big|X_{k,i}^{N,\eta}(t) - \overline{X}_{k,i}^\eta(t)\big| \leq C\eta^{-d-1-\alpha}\big|X_{k,i}^{N,\eta}(t) - \overline{X}_{k,i}^\eta(t)\big|.
\end{aligned}$$

Therefore, by Fubini's theorem,

$$\mathbb{E}\int_0^T \sum_{j=1}^n |L_j^1(t)|^2 \mathrm{d}t \leq C(n)\eta^{-2(d+1+\alpha)}\mathbb{E}\int_0^T \big|X_{k,i}^{N,\eta}(t) - \overline{X}_{k,i}^\eta(t)\big|^2 \mathrm{d}t$$

$$\leq C(n)\eta^{-2(d+1+\alpha)} \int_0^T \sup_{k=1,\dots,N} \mathbb{E}\Big(\sup_{0<s<t} \big| X_{k,i}^{N,\eta}(t) - \overline{X}_{k,i}^{\eta}(t) \big|^2 \Big) \mathrm{d}t. \tag{2.35}$$

We can estimate the second term $L_j^2(t)$ in a similar way, leading to

$$\mathbb{E} \int_0^T \sum_{j=1}^n L_j^2(t)^2 \mathrm{d}t \leq C(n)\eta^{-2(d+1+\alpha)} \int_0^T \sup_{\ell=1,\dots,N} \mathbb{E}\Bigg(\sup_{0<s<t} \sum_{j=1}^n \big| X_{\ell,j}^{N,\eta}(t) - \overline{X}_{\ell,j}^{\eta}(t) \big|^2 \Bigg) \mathrm{d}t. \tag{2.36}$$

The third term $L_j^3(t)$ has to be treated in a different way. First, we use the Lipschitz continuity of f_η to find that

$$L_j^3(t) \leq \frac{C(n)}{N\eta^\alpha} \Bigg| \sum_{\ell=1}^N \big(B_{ij}^\eta(\overline{X}_{k,i}^\eta - \overline{X}_{\ell,j}^\eta) - B_{ij}^\eta * u_{\eta,j}(\overline{X}_{k,i}^\eta) \big) - \frac{1}{\eta^d} B_{ii}(0) \Bigg|.$$

This implies that

$$\begin{aligned} \mathbb{E} \int_0^T \sum_{j=1}^n L_j^3(t)^2 \mathrm{d}t &\leq \frac{C(n,T)}{N^2\eta^{2(d+\alpha)}} \\ &+ \frac{C(n)}{N^2\eta^{2\alpha}} \sum_{j=1}^n \int_0^T \mathbb{E}\Bigg(\sum_{\ell=1}^N \Big(B_{ij}^\eta\big(\overline{X}_{k,i}^\eta(t) - \overline{X}_{\ell,j}^\eta(t)\big) - B_{ij}^\eta * u_{\eta,j}(\overline{X}_{k,i}^\eta) \Big) \Bigg)^2 \mathrm{d}t. \end{aligned} \tag{2.37}$$

It remains to estimate the expectation. To this end, we introduce

$$D_{(k,i),(\ell,j)}(t) := B_{ij}^\eta(\overline{X}_{k,i}^\eta(t) - \overline{X}_{\ell,j}^\eta(t)) - B_{ij}^\eta * u_{\eta,j}(t, \overline{X}_{k,i}^\eta(t)), \quad (\ell,j) \neq (k,i).$$

The processes $\overline{X}_{k,i}^\eta$ and $\overline{X}_{\ell,j}^\eta$ are independent, since for $i = j$, we are considering N independent copies of the same process and for $i \neq j$, the equation fulfilled by $\overline{X}_{k,i}^\eta$ does not depend on the process $\overline{X}_{\ell,j}^\eta$. If $(k,i) \neq (\ell,j)$, $(k,i) \neq (m,j)$, and $\ell \neq m$, the processes $D_{(k,i),(\ell,j)}(t)$ and $D_{(k,i),(m,j)}(t)$ are orthogonal, since

$$\begin{aligned} \mathbb{E}\big(D_{(k,i),(\ell,j)}(t) D_{(k,i),(m,j)}(t)\big) = \int_{\mathbb{R}^d} \Bigg(&\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} B_{ij}^\eta(x-y) B_{ij}^\eta(x-z) u_{\eta,j}(t,y) u_{\eta,j}(t,z) \mathrm{d}y \mathrm{d}z \\ &- 2 \int_{\mathbb{R}^d} B_{ij}^\eta(x-y) u_{\eta,j}(t,y) (B_{ij}^\eta * u_{\eta,j})(t,y) \mathrm{d}y \\ &+ (B_{ij}^\eta * u_{\eta,j})(t,x)(B_{ij}^\eta * u_{\eta,j})(t,x) \Bigg) u_{\eta,i}(t,x) \mathrm{d}x = 0. \end{aligned}$$

Together with $\mathbb{E}(D_{(k,i),(\ell,j)}) = 0$, this shows that the processes $D_{(k,i),(\ell,j)}$ are uncorrelated. However, if $(k,i) \neq (\ell,j)$, $(k,i) \neq (m,j)$, and $\ell = m$, the expectation does not vanish:

$$\begin{aligned} \mathbb{E}\big(D_{(k,i),(\ell,j)}(t)^2\big) = \int_{\mathbb{R}^d} \Bigg(&(B_{ij}^\eta * u_{\eta,j})(t,x)(B_{ij}^\eta * u_{\eta,j})(t,x) + \int_{\mathbb{R}^d} \Big(B_{ij}^\eta(x-y)^2 u_{\eta,j}(t,y) \\ &- 2B_{ij}^\eta(x-y) u_{\eta,j}(t,y)(B_{ij}^\eta * u_{\eta,j}(t,x)) \Big) \mathrm{d}y \Bigg) u_{\eta,i}(t,x) \mathrm{d}x \end{aligned}$$

$$= \int_{\mathbb{R}^d} \big(((B^\eta_{ij})^2 * u_{\eta,j})(t,x) - (B^\eta_{ij} * u_{\eta,j})(t,x)^2\big) u_{\eta,i}(t,x)\mathrm{d}x.$$

This expression is independent of the particle index k and ℓ, it depends only on the species numbers i and j. The case $(k,i) = (\ell,j)$ can be treated in a similar way with the difference that, since $D_{(k,i),(k,i)}(t) = \eta^{-d}B_{ii}(0) - B^\eta_{ii} * u_{i,\eta}(\overline{X}^\eta_{k,i}(t))$, we obtain for $\mathbb{E}(D_{(k,i),(k,i)}(t)D_{(k,i),(m,j)}(t))$ an additional term of order η^{-2d}. Hence, we infer from (2.37) and the previous computation that

$$\begin{aligned}
\mathbb{E}\int_0^T \sum_{j=1}^n L^3_j(t)^2\mathrm{d}t - \frac{C(n,T)}{N^2\eta^{2(d+\alpha)}} &= \frac{C(n)}{N^2\eta^{2\alpha}} \sum_{j=1}^n \sum_{\ell=1}^N \int_0^T \mathbb{E}\big(D_{(k,i),(\ell,j)}(t)^2\big)\mathrm{d}t \\
&\le \frac{C(n)}{N\eta^{2\alpha}} \|u_{\eta,i}\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\
&\quad \times \sum_{j=1}^n \int_0^T \left(\|(B^\eta_{ij})^2 * u_{\eta,j}\|_{L^1(\mathbb{R}^d)} + \|B^\eta_{ij} * u_{\eta,j}\|^2_{L^2(\mathbb{R}^d)}\left(1 + \frac{1}{\eta^{2d}}\right)\right)\mathrm{d}t \\
&\le \frac{C(n)}{N\eta^{2\alpha}} \sum_{j=1}^n \int_0^T \left(\|B^\eta_{ij}\|^2_{L^2(\mathbb{R}^d)}\|u_{\eta,j}\|_{L^\infty(\mathbb{R}^d)} + \|B^\eta_{ij}\|^2_{L^1(\mathbb{R}^d)}\|u_{\eta,j}\|^2_{L^2(\mathbb{R}^d)}\left(1 + \frac{1}{\eta^{2d}}\right)\right)\mathrm{d}t \\
&\le \frac{C(T,n)}{N\eta^{2(d+\alpha)}}, \qquad (2.38)
\end{aligned}$$

recalling that $\|B^\eta_{ij}\|_{L^2(\mathbb{R}^d)} \le C\eta^{-d/2}$ and $\|B^\eta_{ij}\|_{L^1(\mathbb{R}^d)} = A_{ij} \le A$ and choosing $\eta < 1$.

Inserting estimates (2.35), (2.36), and (2.38) for $L^m_j(t)$ $(m = 1,2,3)$ into (2.34), we conclude that

$$\begin{aligned}
\sup_{k=1,\dots,N} \mathbb{E}\left(\sum_{i=1}^n \sup_{0<s<T} |E_{2,i}(s)|^2\right) &\le \frac{C(T,n)}{N\eta^{2(d+\alpha)}} \\
&\quad + C(n,\sigma_{\min})\eta^{-2(d+1+\alpha)} \int_0^T \sup_{k=1,\dots,N} \mathbb{E}\Big(\sup_{0<s<t} \big|X^{N,\eta}_{k,i}(t) - \overline{X}^\eta_{k,i}(t)\big|^2\Big)\mathrm{d}t.
\end{aligned}$$

We infer from (2.32), estimate (2.33), and the previous estimate for $E_{2,i}$ that

$$\begin{aligned}
S(T) := \sup_{k=1,\dots,N} \mathbb{E}\left(\sum_{i=1}^n \sup_{0<s<t} |D^{N,\eta}_{k,i}(s)|^2\right) \\
\le \frac{C(T,n)}{N\eta^{2(d+\alpha)}} + C(n,\sigma_{\min})(\eta^{-2(d+1+\alpha)} + T)\int_0^T S(t)\mathrm{d}t.
\end{aligned}$$

Note that the function S is continuous because of the continuity of the paths of $X^{N,\eta}_{k,i}$ and $\overline{X}^\eta_{k,i}$. Therefore, by Gronwall's inequality, we have

$$S(T) \le \frac{C(T,n)}{N\eta^{2(d+\alpha)}} \exp\big(C(n,T,\sigma_{\min})\eta^{-2(d+1+\alpha)}T\big).$$

We choose $\delta > 0$ such that $C(n, T, \sigma_{\min})T\delta < 1$ and $\eta > 0$ such that $\eta^{-2(d+1+\alpha)} \le \delta \log N$. Then

$$S(T) \le \frac{1}{N} C(T, n) \exp\big(C(n, T, \sigma_{\min})T\delta \log N\big) = C(T, n)N^{-1+C(n,T,\sigma_{\min})T\delta}.$$

This finishes the proof. □

Next, we prove an error estimate for the difference $\overline{X}^{\eta}_{k,i} - \widehat{X}_{k,i}$.

Lemma 2.10. *Let $\overline{X}^{\eta}_{k,i}$ and $\widehat{X}_{k,i}$ be the solutions to (2.6) and (2.8) in the sense of Proposition 2.4. Under the assumptions of Theorem 2.5, it holds for small $\eta > 0$ that*

$$\sup_{k=1,\ldots,N} \mathbb{E}\bigg(\sum_{i=1}^{n} \sup_{0<s<T} \big|(\overline{X}^{\eta}_{k,i} - \widehat{X}_{k,i})(s)\big|^2\bigg) \le C(T, \sigma_{\ min})\eta^{2(1-\alpha)}.$$

Proof. Since we are considering N independent copies, we can omit the particle index k. Set $D^{\eta}_i(s) := \overline{X}^{\eta}_{k,i}(s) - \widehat{X}_{k,i}(s)$. Then, similarly as in the proof of Lemma 2.9, $D^{\eta}_i(s) = D_1(s) + D_2(s)$, where

$$\begin{aligned} D_1(s) &= -\int_0^s \big(\nabla U_i(\overline{X}^{\eta}_i(t)) - \nabla U_i(\widehat{X}_i(t))\big)\mathrm{d}t, \\ D_2(s) &= \int_0^s \Bigg[\bigg(2\sigma_i + 2\sum_{j=1}^{n} f_\eta\big(B^{\eta}_{ij} * u_{\eta,j}(\overline{X}^{\eta}_i)\big)\bigg)^{1/2} \\ &\qquad - \bigg(2\sigma_i + 2\sum_{j=1}^{n} f\big(a_{ij}u_j(\widehat{X}_i)\big)\bigg)^{1/2}\Bigg]\mathrm{d}W_i(t). \end{aligned}$$

We infer from the Lipschitz continuity of ∇U_i and Fubini's theorem that

$$\mathbb{E}\Big(\sup_{0<s<T} |D_1(s)|^2\Big) \le CT\mathbb{E}\bigg(\int_0^T \big|\overline{X}^{\eta}_i(s) - \widehat{X}_i(s)\big|^2 \mathrm{d}s\bigg) \le CT\int_0^T \mathbb{E}\Big(\sup_{0<s<t} |D^{\eta}_i(s)|^2\Big)\mathrm{d}t. \tag{2.39}$$

Similarly as in the proof of Lemma 2.9, we use for D_2 the Burkholder–Davis–Gundy inequality and the Lipschitz continuity of $x \mapsto (2\sigma_i + x)^{1/2}$ on $[0, \infty)$ to obtain

$$\begin{aligned} \mathbb{E}\Big(\sup_{0<s<T} |D_2(s)|^2\Big) &\le C\mathbb{E}\int_0^T \bigg(\sum_{j=1}^{n} \big(f(a_{ij}u_j(\widehat{X}_i)) - f_\eta(B^{\eta}_{ij} * u_{\eta,j}(\overline{X}^{\eta}_i))\big)\bigg)^2 \mathrm{d}t \\ &\le C(n)(D_{21} + D_{22} + D_{23} + D_{24}), \end{aligned} \tag{2.40}$$

where

$$\begin{aligned} D_{21} &= \sum_{j=1}^{n} \mathbb{E}\int_0^T \big(f(a_{ij}u_j(\widehat{X}_i)) - f_\eta(a_{ij}u_j(\widehat{X}_i))\big)^2 \mathrm{d}t, \\ D_{22} &= \sum_{j=1}^{n} \mathbb{E}\int_0^T \big(f_\eta(a_{ij}u_j(\widehat{X}_i)) - f_\eta(B^{\eta}_{ij} * u_j(\widehat{X}_i))\big)^2 \mathrm{d}t, \end{aligned}$$

$$
\begin{aligned}
D_{23} &= \sum_{j=1}^{n} \mathbb{E}\int_0^T \big(f_\eta(B_{ij}^\eta * u_j(\widehat{X}_i)) - f_\eta(B_{ij}^\eta * u_j(\overline{X}_i^\eta))\big)^2 \mathrm{d}t, \\
D_{24} &= \sum_{j=1}^{n} \mathbb{E}\int_0^T \big(f_\eta(B_{ij}^\eta * u_j(\overline{X}_i^\eta)) - f_\eta(B_{ij}^\eta * u_{\eta,j}(\overline{X}_i^\eta))\big)^2 \mathrm{d}t.
\end{aligned}
$$

The first expression D_{21} vanishes if $\eta > 0$ is sufficiently small, since then $f = f_\eta$ on the range of $a_{ij}u_j(\widehat{X}_i)$. Using

$$
\|a_{ij}u_j - B_{ij}^\eta * u_j\|_{L^2(0,T;L^2(\mathbb{R}^d))} \le C\eta \|\nabla u_j\|_{L^2(0,T;L^2(\mathbb{R}^d))} \le C\eta,
$$

which was shown in the proof of Theorem 2.3, and the Lipschitz continuity of f_η with Lipschitz constant less or equal $\eta^{-\alpha}$, we find that

$$
\begin{aligned}
D_{22} &= \sum_{j=1}^{n} \int_0^T \int_{\mathbb{R}^d} \big(f_\eta(a_{ij}u_j) - f_\eta(B_{ij}^\eta * u_j)\big)^2 u_i \mathrm{d}x\mathrm{d}t \\
&\le \eta^{-2\alpha} \sum_{j=1}^{n} \|u_i\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \|a_{ij}u_j - B_{ij}^\eta * u_j\|^2_{L^2(0,T;L^2(\mathbb{R}^d))} \le C(n)\eta^{2(1-\alpha)}.
\end{aligned}
$$

Thanks to the uniform boundedness of the family $B_{ij}^\eta * u_j$, we can choose $\eta > 0$ sufficiently small, say $\eta \le \eta^*$ for some $\eta^* > 0$, such that $f(B_{ij}^\eta * u_j) = f_\eta(B_{ij}^\eta * u_j)$ for $0 < \eta \le \eta^*$. Then, using Young's convolution inequality and the uniform estimate $\|\nabla u_j\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\|u_0\|_{H^s(\mathbb{R}^d)}$ from Theorem 2.3, the third term D_{23} is estimated as

$$
\begin{aligned}
D_{23} &\le C(\eta^*) \sum_{j=1}^{n} \|\nabla(B_{ij}^\eta * u_j)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d)} \int_0^T \mathbb{E}\big(|\widehat{X}_i(t) - \overline{X}_i^\eta(t)|^2\big)\mathrm{d}t \\
&\le C \sum_{j=1}^{n} \|\nabla u_j\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d)} \int_0^T \mathbb{E}\big(|\widehat{X}_i(t) - \overline{X}_i^\eta(t)|^2\big)\mathrm{d}t \\
&\le C \int_0^T \mathbb{E}\Big(\sup_{0<s<t} |D_i^\eta(s)|^2\Big)\mathrm{d}t.
\end{aligned}
$$

Finally, it follows from the error estimate for $u - u_\eta$ from Theorem 2.3 that

$$
\begin{aligned}
D_{24} &\le C \sum_{j=1}^{n} \int_0^T \int_{\mathbb{R}^d} |B_{ij}^\eta * u_j - B_{ij}^\eta * u_{\eta,j}|^2 u_{\eta,i} \mathrm{d}x\mathrm{d}t \\
&\le C \sum_{j=1}^{n} \|u_{\eta,i}\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \int_0^T \|B_{ij}^\eta\|^2_{L^1(\mathbb{R}^d)} \|u_j - u_{\eta,j}\|^2_{L^2(\mathbb{R}^d)} \mathrm{d}t \\
&\le C(T)\eta^2.
\end{aligned}
$$

Inserting the estimates for $D_{21}, \ldots, D_{24}$ into (2.40), we conclude that

$$
\mathbb{E}\Big(\sup_{0<s<T} |D_2(s)|^2\Big) \le C(T,n)\eta^{2(1-\alpha)} + C(T)\int_0^T \mathbb{E}\Big(\sup_{0<s<t} |D_i^\eta(s)|^2\Big)\mathrm{d}t.
$$

Together with estimate (2.39) for $D_1(s)$ and recalling that $D_i^\eta = D_1 + D_2$, we arrive at

$$\mathbb{E}\Big(\sup_{0<s<T} |D_i^\eta(s)|^2\Big) \le C(T,n)\eta^{2(1-\alpha)} + C(T)\int_0^T \mathbb{E}\Big(\sup_{0<s<t} |D_i^\eta(s)|^2\Big)\mathrm{d}t.$$

The proof is finished after applying Gronwall's inequality and summing over $i = 1, \dots, n$. □

Theorem 2.5 now follows from Lemmas 2.9 and 2.10 and the triangle inequality:

$$\begin{aligned}
&\sup_{k=1,\dots,N} \mathbb{E}\bigg(\sum_{i=1}^n \sup_{0<s<t} \big|X_{\eta,i}^{k,N}(s) - \widehat{X}_i^k(s)\big|^2\bigg) \\
&\quad \le 2 \sup_{k=1,\dots,N} \mathbb{E}\bigg(\sum_{i=1}^n \sup_{0<s<t} \big|X_{\eta,i}^{k,N}(s) - \overline{X}_{\eta,i}^k(s)\big|^2\bigg) \\
&\qquad + 2 \sup_{k=1,\dots,N} \mathbb{E}\bigg(\sum_{i=1}^n \sup_{0<s<t} \big|\overline{X}_{\eta,i}^k(s) - \widehat{X}_i^k(s)\big|^2\bigg) \\
&\quad \le C_1 N^{-1+C_2\delta} + C_3\eta^{2(1-\alpha)}.
\end{aligned}$$

The condition $\log N \ge \delta^{-1}\eta^{-2(d+1+\alpha)}$ is equivalent to

$$N^{-1+C_2\delta} \le \exp((-\delta^{-1} + C_2)\eta^{-2(d+1+\alpha)}).$$

We choose $\delta > 0$ such that $-\delta^{-1} + C_2 < 0$ and observe that exponential decay is always faster than algebraic decay to conclude that $\exp((-\delta^{-1} + C_2)\eta^{-2(d+1+\alpha)}) \le \eta^{2(1-\alpha)}$. This yields

$$\sup_{k=1,\dots,N} \mathbb{E}\bigg(\sum_{i=1}^n \sup_{0<s<t} \big|X_{\eta,i}^{k,N}(s) - \widehat{X}_i^k(s)\big|^2\bigg) \le C_4\eta^{2(1-\alpha)},$$

finishing the proof.

2.7 Numerical tests

In this section, we perform some numerical simulations of the particle system (2.5) in one space dimension, without environmental potential, and with linear function $f(x) = x$. We are interested in the numerical comparison of the solutions to the particle systems (2.3) and (2.5) in terms of the segregation behavior. We explore the ability of both systems to model the segregation of the species. Numerical tests for the associated cross-diffusion systems (2.1) and (2.2) are work in progress.

We discretize the particle systems (2.3) and (2.5) by the Euler–Maruyama scheme. Let $M \in \mathbb{N}$ and introduce the time steps $0 < t_1 < \cdots < t_M = T$ with $\triangle t_m = t_{m+1} - t_m$. We approximate $X_{k,i}^{N,\eta}(t_m)$ by $x_m^{k,i}$ and $Y_{k,i}^{N,\eta}(t_m)$ by $y_m^{k,i}$, defined by, respectively,

$$x_{m+1}^{k,i} = x_m^{k,i} + \bigg(2\sigma_i + \frac{2}{N}\sum_{j=1}^n\sum_{\ell=1}^N B_{ij}^\eta(x_m^{k,i} - x_m^{\ell,j})\bigg)^{1/2}\sqrt{\triangle t_m} w_m,$$

$$y_{m+1}^{k,i} = y_m^{k,i} - \sum_{j=1}^{n} \frac{1}{N} \sum_{\ell=1}^{N} \nabla B_{ij}^{\eta}(y_m^{k,i} - y_\ell^{m,j}) \triangle t_m + \sqrt{2\sigma_i \triangle t_m} z_m,$$

with initial conditions $x_0^{i,k} = \xi_i^k$ and $y_0^{i,k} = \xi_i^k$, where ξ_i^k are iid random variables and w_m and z_m are normally distributed. It is well known that the solutions to the Euler–Maruyama scheme converge to the associated stochastic processes in the strong sense; see, e.g., [68, Theorem 9.6.2].
The numerical scheme is implemented in MATLAB using the parallel computing toolbox to accelerate the simulations. The interaction potential is given by $B(x) = \exp(-1/(1-x^2))$ for $|x| \leq 1$ and $B(x) = 0$ else. Then $B_{ij}^{\eta}(x) = \eta^{-1}B(x/\eta)$. The numerical parameters are $\triangle t = 1/100$, $\eta = 2$, $N = 5000$ particles, $n_{\text{sim}} = 500$ simulations.

2.7.1 Two species: nonsymmetric case

We consider a nonsymmetric diffusion matrix with $a_{11} = 0$, $a_{12} = 355$, $a_{21} = 25$, $a_{22} = 0$, and $\sigma_1 = 1$, $\sigma_2 = 2$. The initial data are Gaussian distributions with mean -1 (for species $i = 1$) and 1 (for species $i = 2$) and variance 2. Figure 2.1 shows the approximate densities of both species (histogram) for systems (2.5) and (2.3) at time $t = 2$. We observe a segregation of the densities in both models. In the population system (2.5), species 1 develops two clusters because of the very different "population pressure" parameters $a_{12} = 355$ and $a_{21} = 25$, while species 2 develops only one cluster around $x = 0$; see Figure 2.1 left. The segregation effect is stronger in the particle system (2.3) in the sense that both species avoid each other as far as possible; see Figure 2.1 right. This is not surprising since the diffusion of system (2.5) is generally larger than that one of system (2.3). The numerical results confirm the segregation property defined in [3]. Indeed, this work considers the cross-diffusion system (2.3) with $\sigma_1 = \sigma_2 = 0$ and $a_{11} = a_{12} = a_{21} = a_{22} = 1$. It was proved that the two species are segregated for all times if they do so initially. Here, segregation means that the intersection of the supports of the densities is empty.

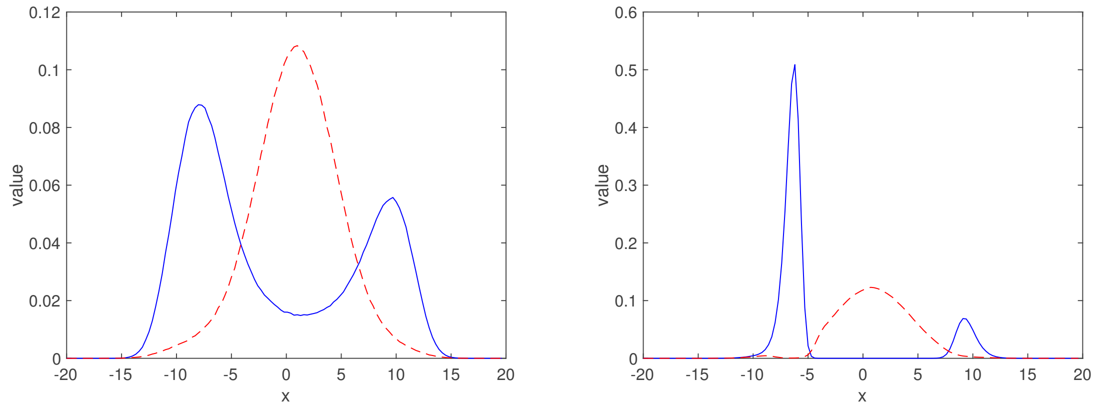

Figure 2.1: Nonsymmetric case: Densities of particle system (2.5) corresponding to the SKT population model (left) and particle system (2.3) (right) at time $t = 2$. Solid blue line: species 1; Dashed red line: species 2.

2.7.2 Two species: symmetric case

We investigate the symmetric case by choosing $a_{11} = a_{22} = 0$, $a_{12} = a_{21} = 355$, and, as before, $\sigma_1 = 1$, $\sigma_2 = 2$. The initial data are chosen as in the previous example. In this example, we expect that cross-diffusion dominates self-diffusion. We present the approximate densities for different times in Figure 2.2. In both models, the species have the tendency to segregate. As expected, the segregation in the particle system (2.3) is stronger than in system (2.5) corresponding to the SKT model.

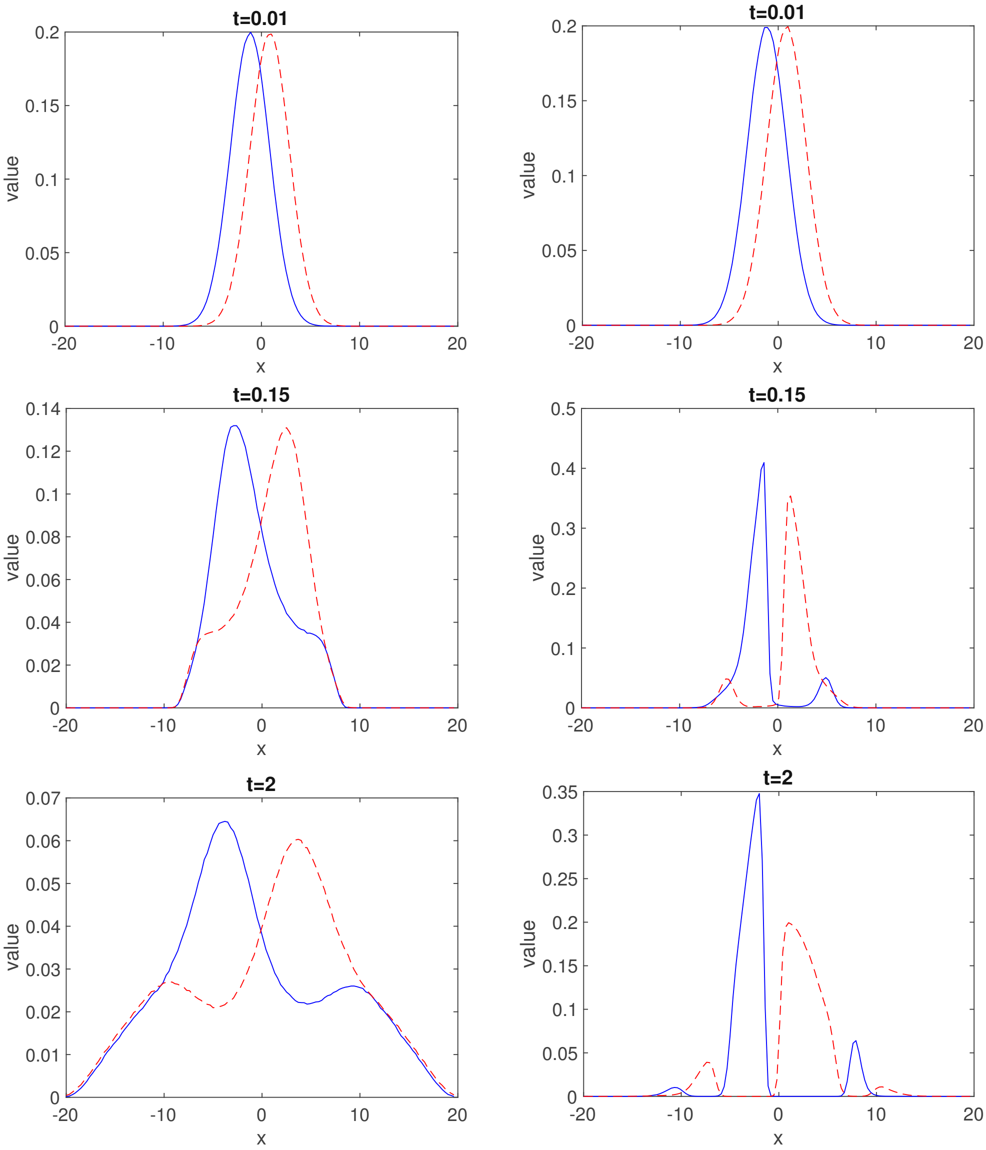

Figure 2.2: Symmetric case: Densities of particle system (2.5) corresponding to the SKT population model (left) and particle system (2.3) (right) for different times $t = 0.01, 0.15, 2$. Solid blue line: species 1; dashed red line: species 2.

2.7.3 Three species

Our third numerical experiment illustrates the segregation behaviour in case of three interacting species with coefficients $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 3$ and

$$(a_{ij}) = \begin{pmatrix} 0 & 355 & 355 \\ 25 & 0 & 25 \\ 355 & 0 & 0 \end{pmatrix}.$$

Similar as in the two-species case, the initial data are overlapping normal distributions with means -1, 2, and -3, respectively, and variance 2. The approximate densities at $t = 2$ are shown in Figure 2.3. We observe that the approximate densities of particle model (2.3) show a much clearer component-wise segregation behavior than the stochastic particle model (2.5), which corresponds to the SKT system, where the diffusion effects are much stronger. This may be explained by the fact that, on the PDE level, the gradient-flow structure of model (2.2) can be written species-wise, whereas the SKT model (2.1) (with $f(x) = x$) only possess a vector-valued gradient-flow structure.

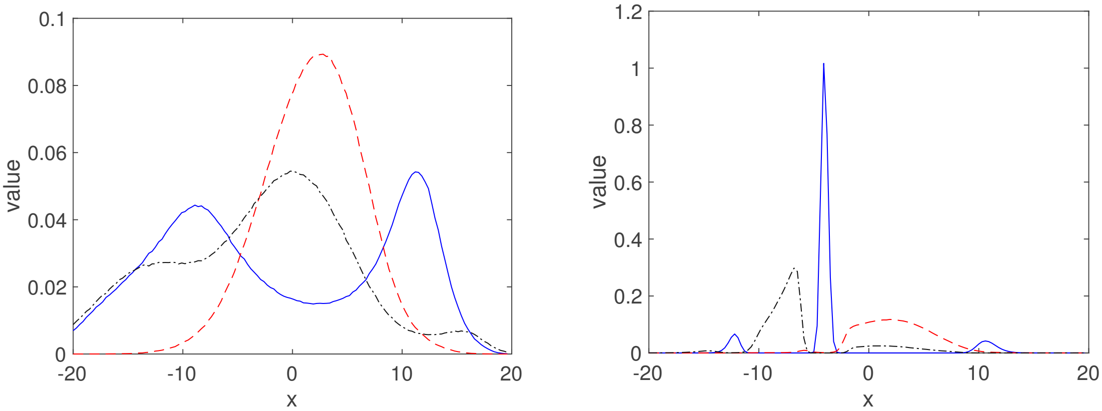

Figure 2.3: Three-species case: Densities of particle system (2.5) corresponding to the SKT population model (left) and particle system (2.3) (right) at time $t = 2$. Solid blue line: species 1; dashed red line: species 2; dash-dotted black line: species 3.

2.7.4 Cubic nonlinearity

For our last experiment, we compare the numerical results for the cubic nonlinearity $f(s) = s^3$ with the linear case imposed in the previous examples. The parameters are the same as in Section 2.7.2. The numerical simulations are performed without using approximating functions f_η. This may be justified by the fact that the simulations deal with the behavior for small time scales and with compactly supported initial data. We observe in Figure 2.4 that the cubic nonlinearity causes more clustering than the linear case $f(s) = s$. The simulations suggests that in the cubic case, diffusion happens on a faster time scale than segregation, while in the linear case, the particles diffuse slower and hence they form bigger but fewer clusters.

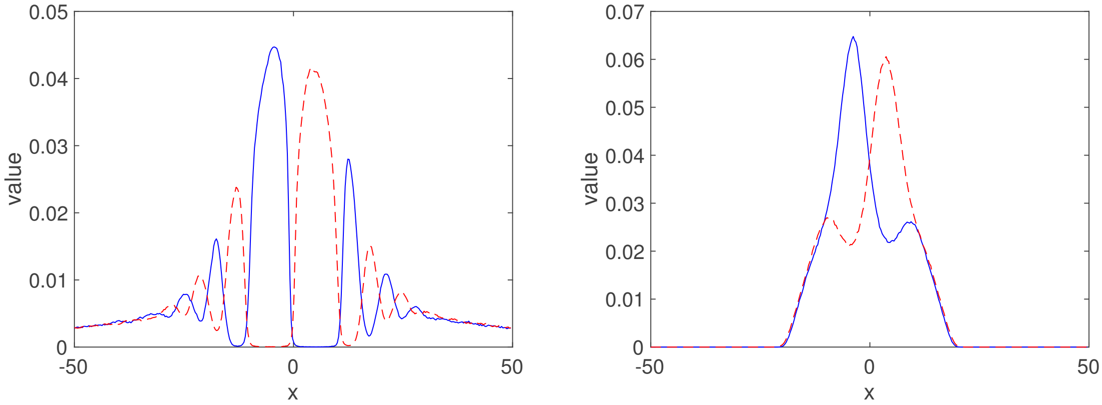

Figure 2.4: Densities of particle system (2.5) corresponding to the SKT population model with $f(s) = s^3$ (left) and $f(s) = s$ (right) at time $t = 2$. Solid blue line: species 1; dashed red line: species 2. The right figure is the same as in Figure 2.2 but with the range $x = -50, \dots, 50$.

2.A Auxiliary results

For the convenience of the reader, we recall some well-known estimates used in this chapter of the thesis:

Lemma 2.11 (Young's convolution inequality, [75, Formula (7), page 107])**.** *Let* $1 \le p, q, r \le \infty$ *be such that* $1/p + 1/q = 1 + 1/r$ *and let* $f \in L^p(\mathbb{R}^d)$, $g \in L^q(\mathbb{R}^d)$. *Then* $f * g \in L^r(\mathbb{R}^d)$ *and*

$$\|f * g\|_{L^r(\mathbb{R}^d)} \le \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^q(\mathbb{R}^d)}.$$

Lemma 2.12 (Moser-type estimate I, [80, Prop. 2.1(A)])**.** *Let* $s \in \mathbb{N}$ *and* $\alpha \in \mathbb{N}_0^n$ *with* $|\alpha| = s$. *Then there exists a constant* $C > 0$ *such that for all* f, $g \in H^s(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$,

$$\|D^\alpha(fg)\|_{L^2(\mathbb{R}^d)} \le C\big(\|f\|_{L^\infty(\mathbb{R}^d)} \|D^s g\|_{L^2(\mathbb{R}^d)} + \|D^s f\|_{L^2(\mathbb{R}^d)} \|g\|_{L^\infty(\mathbb{R}^d)}\big).$$

Lemma 2.13 (Moser-type estimate II, [80, Prop. 2.1(C)])**.** *Let* $s \in \mathbb{N}$ *and* $\alpha \in \mathbb{N}_0^n$ *with* $|\alpha| = s$. *Then there exists a constant* $C > 0$ *such that for smooth* $g : \mathbb{R} \to \mathbb{R}$ *and* $u \in H^s(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$,

$$\|D^\alpha g(u)\|_{L^2(\mathbb{R}^d)} \le C\|g'\|_{C^{s-1}(\mathbb{R})} \|u\|^{s-1}_{L^\infty(\mathbb{R}^d)} \|D^\alpha u\|_{L^2(\mathbb{R}^d)}.$$

Lemma 2.14 (Moser-type commutator inequality, [80, Prop. 2.1(B)])**.** *Let* $s \in \mathbb{N}$ *and* $\alpha \in \mathbb{N}_0^n$ *with* $|\alpha| = s$. *Then there exists* $C > 0$ *such that for all* $f \in H^s(\mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d)$ *and* $g \in H^{s-1}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$,

$$\|D^\alpha(fg) - fD^\alpha(g)\|_{L^2(\mathbb{R}^d)} \le C\big(\|Df\|_{L^\infty(\mathbb{R}^d)} \|D^{s-1}g\|_{L^2(\mathbb{R}^d)} + \|D^s f\|_{L^2(\mathbb{R}^d)} \|g\|_{L^\infty(\mathbb{R}^d)}\big),$$

where $D^s = \sum_{|\alpha|=s} D^\alpha$.

A physical law must possess mathematical beauty
— Paul Dirac[1]

3 Mean-field derivation of a porous-medium equation with fractional diffusion

Contents

This chapter is taken from the article

[30] Li Chen, Alexandra Holzinger, Ansgar Jüngel, and Nicola Zamponi. Analysis and mean-field derivation of a porous-medium equation with fractional diffusion. *Communications in Partial Differential Equations*, 1-53, 2022.

3.1 Introduction and problem setting

The aim of this chapter is to derive and analyze the following nonlocal porous-medium equation:

$$\partial_t \rho = \operatorname{div}(\rho \nabla P), \quad P = (-\Delta)^{-s} f(\rho), \quad \rho(0) = \rho^0 \quad \text{in } \mathbb{R}^d, \tag{3.1}$$

where $0 < s < 1$, $d \geq 2$, and $f \in C^1([0, \infty))$ is a nondecreasing function satisfying $f(0) = 0$. This model describes a particle system that evolves according to a continuity equation for the density $\rho(x,t)$ with velocity $v = -\nabla P$. The velocity is assumed to be the gradient of a potential, which expresses Darcy's law. The pressure P is related to the density in a nonlinear and nonlocal way through $P = (-\Delta)^{-s} f(\rho)$. The nonlocal operator $(-\Delta)^{-s}$ can be written as a convolution operator with a singular kernel,

$$(-\Delta)^{-s} u = \mathcal{K} * u, \quad \mathcal{K}(x) = c_{d,-s}|x|^{2s-d}, \quad x \in \mathbb{R}^d, \tag{3.2}$$

where $c_{d,-s} = \Gamma(d/2 - s)/(4^s \pi^{d/2} \Gamma(s))$ and Γ denotes the Gamma function [110, Theorem 5].

[1] Moscow University, 1956.

If $s = 0$, we recover the porous-medium equation (for nonnegative solutions), while the case $s = 1$ was investigated in [24, 118] with $f(u) = u$ for the evolution of the vortex density in a superconductor. Related models (with $f(u) = u$) appear in the dynamics of dislocations (line defects) in crystals [5, (1.5)]. Other applications include particle systems with long-range interactions [116, Sec. 6.2]. The case $0 < s < 1$ corresponds to long-range repulsive interactions. This model, still with $f(u) = u$, was investigated in [5], but a mathematical justification is missing. In this chapter, we provide a rigorous derivation from an interacting particle system for general functions $f(u)$. In this way, we aim to contribute to the understanding of mean-field limits involving nonquadratic nonlinearities.
Equation (3.1) was first analyzed in [17] with $f(u) = u$ for nonnegative solutions and in [4] with $f(u) = |u|^{m-2}u$ ($m > 1$) for sign-changing solutions. The nonnegative solutions have the interesting property that they propagate with finite speed, which is not common in other fractional diffusion models [17, 107]. Equation (3.1) was probabilistically interpreted in [99], and it was shown that the probability density of a so-called random flight process is given by a Barenblatt-type profile. Previous mean-field limits leading to (3.1) were concerned with the linear case $f(u) = u$ only; see [42] (using the technique of [103]) and [96] (including additional diffusion as in (3.7) below). In [34], equation (3.1) (with $f(u) = u$) was derived in the high-force regime from the Euler–Riesz equations, which can be derived in the mean-field limit from interacting particle systems [43]. A direct derivation from particle systems with Lévy noise was proved in [38] for cross-diffusion systems, but still with $f(u) = u$. Up to our knowledge, a rigorous derivation of (3.1) from stochastic interacting particle systems for general nonlinearities $f(u)$ like power functions is missing in the literature. With the main result of this chapter, we fill this gap.

3.1.1 Problem setting

Equation (3.1) is derived from an interacting particle system with N particles, moving in the whole space $\mathbb{R}^d$. Because of the singularity of the integral kernel and the degeneracy of the nonlinearity, we approximate (3.1) using three levels. First, we introduce a parabolic regularization adding a Brownian motion to the particle system with diffusivity $\sigma \in (0, 1)$ and replacing f by a smooth approximation f_σ. Second, we replace the interaction kernel $\mathcal{K}$ by a smooth kernel $\mathcal{K}_\zeta$ with compact support, where $\zeta > 0$. Third, we consider interaction functions $\mathcal{W}_\beta$ with $\beta \in (0, 1)$, which approximate the delta distribution. We refer to Subsection 3.1.3 for the precise definitions.
The particle positions are represented on the *microscopic level* by the stochastic processes $X_i^N(t)$ evolving according to

$$\begin{aligned} \mathrm{d}X_i^N(t) &= -\nabla\mathcal{K}_\zeta * f_\sigma\bigg(\frac{1}{N}\sum_{j=1,\, j\neq i}^N \mathcal{W}_\beta(X_j^N(t) - X_i^N(t))\bigg)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ X_i^N(0) &= \xi_i, \quad i = 1, \ldots, N, \end{aligned} \tag{3.3}$$

where the convolution has to be understood with respect to x_i[2], $(W_i(t))_{t\geq 0}$ are independent d-dimensional Brownian motions defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, and

[2]This means that the drift term becomes $-\int_{\mathbb{R}^d} \nabla\mathcal{K}_\zeta(y) f_\sigma\big(\frac{1}{N}\sum_{j=1,\, j\neq i}^N \mathcal{W}_\beta(X_j^N(t) - X_i^N(t) + y)\big)\mathrm{d}y$

ξ_i are independent identically distributed random variables in $\mathbb{R}^d$ with the same probability density function ρ_σ^0 (defined in (3.12) below).
We remark that in comparison to the classical moderate regime (see Section 1.2.2 for an introduction), the *strength of interaction* now depends on ζ and β, which are scaled differently with respect to the number of particles N (see below), which makes the study more involved.
The mean-field-type limit is performed in three steps. First, for fixed (σ, β, ζ), system (3.3) is approximated for $N \to \infty$ on the *intermediate level* by

$$\begin{aligned} &\mathrm{d}\bar{X}_i^N(t) = -\nabla\mathcal{K}_\zeta * f_\sigma\big(\mathcal{W}_\beta * \rho_{\sigma,\beta,\zeta}(\bar{X}_i^N(t),t)\big)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ &\bar{X}_i^N(0) = \xi_i, \quad i = 1,\dots,N, \end{aligned} \tag{3.4}$$

where $\rho_{\sigma,\beta,\zeta}$ is the probability density function of $\bar{X}_i^N$ and a strong solution to

$$\partial_t\rho_{\sigma,\beta,\zeta} - \sigma\Delta\rho_{\sigma,\beta,\zeta} = \operatorname{div}\big(\rho_{\sigma,\beta,\zeta}\nabla\mathcal{K}_\zeta * f_\sigma(\mathcal{W}_\beta * \rho_{\sigma,\beta,\zeta})\big), \quad \rho_{\sigma,\beta,\zeta}(0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d. \tag{3.5}$$

System (3.4) is uncoupled, since $\bar{X}_i^N$ depends on N only through the initial datum.
Second, passing to the limit $(\beta, \zeta) \to 0$ in the intermediate system leads on the *macroscopic level* to

$$\begin{aligned} &\mathrm{d}\widehat{X}_i^N(t) = -\nabla\mathcal{K} * f_\sigma(\rho_\sigma(\widehat{X}_i^N(t),t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ &\widehat{X}_i^N(0) = \xi_i, \quad i = 1,\dots,N, \end{aligned} \tag{3.6}$$

where ρ_σ is the density function of $\widehat{X}_i^N$ and a weak solution to

$$\partial_t\rho_\sigma = \sigma\Delta\rho_\sigma + \operatorname{div}(\rho_\sigma\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)), \quad \rho_\sigma(0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d. \tag{3.7}$$

We perform the limits $N \to \infty$ and $(\beta, \zeta) \to 0$ simultaneously. In this limit, we use the logarithmic scaling $\beta \sim (\log N)^{-\mu}$ for some $\mu > 0$ between the strength of interaction β and the number of particles N. This can be viewed as a moderately interacting particle system. For the smoothing parameter ζ of the singularity from $\mathcal{K}$, we can even allow an algebraic dependence on N, i.e. $\zeta \sim N^{-\nu}$ for some $\nu > 0$; see Theorem 3.2 for details. Our approach also implies the two-step limit but leading to weak convergence only, compared to the convergence in expectation obtained in Theorem 3.3.
Third, the limit $\sigma \to 0$ is performed on the level of the diffusion equation, based on a priori estimates uniform in σ and the div-curl lemma.
The main result of this chapter is that the particles of system (3.3) become independent in the limit with a common density function that is a weak solution to (3.1)–(3.2).

3.1.2 State of the art

We already mentioned that the existence of weak solutions to (3.1) with $f(u) = u$ was proved first in [17]. The convergence of the weak solution to a self-similar profile was shown by the same authors in [16]. The convergence becomes exponential, at least in one space dimension, when adding a confinement potential [20]. Equation (3.1) with $f(u) = u$ was identified as the Wasserstein gradient flow of a square fractional Sobolev norm [78], implying time decay as well as energy and entropy estimates. The Hölder regularity of solutions to (3.1) was proved in [15] for $f(u) = u$ and in [60] for $f(u) = u^{m-1}$ and $m \geq 2$.

In the literature, related equations have been analyzed too. Equation (3.1) for $f(u) = u$ and the limit case $s = 1$ was shown in [1] to be the Wasserstein gradient flow on the space of probability measures, leading to the well-posedness of the equation and energy-dissipation inequalities. The existence of local smooth solutions to the regularized equation (3.7) are proved in [33]. The solutions $\partial_t \rho = \operatorname{div}(\rho^{m-1}\nabla P)$ with $P = (-\Delta)^{-s}\rho$ in $\mathbb{R}^d$ propagate with finite speed if and only if $m \geq 2$ [107]. The existence of weak solutions to this equation with $P = (-\Delta)^{-s}(\rho^n)$ and $n > 0$ is proved in [88] (in bounded domains). While (3.1) has a parabolic-elliptic structure, parabolic-parabolic systems have been also investigated. For instance, the global existence of weak solutions to $\partial_t \rho = \operatorname{div}(\rho\nabla P)$ and $\partial_t P + (-\Delta)^s P = \rho^\beta$, where $\beta > 1$, was shown in [14]. In [37], the algebraic decay towards the steady state was proved in the case $\beta = 2$. We also mention that fractional porous-medium equations of the type $\partial_t \rho + (-\Delta)^{s/2} f(\rho) = 0$ in $\mathbb{R}^d$ have been studied in the literature; see, e.g., [97]. Compared to (3.1), this problem has infinite speed of propagation. For a review and comparison of this model and (3.1), we refer to [115].

For an introduction to the general concept of mean-field limit we refer to the introduction of this thesis, especially Sections 1.2.1 and 1.2.2 for an introduction of coupling techniques and *moderately interacting particles* which are used in this chapter.

There is a huge literature concerning mean-field limits leading to diffusion equations. In the following, we shortly summarise articles which are relevant for this chapter and refer to the introduction of this thesis as well as reviews [54, 62] and the classical works of Sznitman [112, 113] for more information. Oelschläger proved the mean-field limit in weakly interacting particle systems [92], leading to deterministic nonlinear processes, and moderately interacting particle systems [93], giving porous-medium-type equations with quadratic diffusion. First investigations of moderate interactions in stochastic particle systems with nonlinear diffusion coefficients were performed in [65]. The approach of moderate interactions was extended in [25, 26] to multi-species systems, deriving population cross-diffusion systems. Reaction-diffusion equations with nonlocal terms were derived in the mean-field limit in [59]. The large population limit of point measure-valued Markov processes leads to nonlocal Lotka–Volterra systems with cross diffusion [51]. Further references can be found in [96, Sec. 1.3].

Compared to previous works, we consider a singular kernel $\mathcal{K}$ and derive a partial differential equation without Laplace diffusion by taking the limit $\sigma \to 0$. The authors of [47] derived the viscous porous-medium equation by starting from a stochastic particle system with a double convolution structure in the drift term, similar to (3.4). The main difference to our work is that (besides different techniques for the existence and regularity of solutions to the parabolic problems) we consider a singular kernel in one part of the convolution and a different scaling for the approximating regularized kernel $\mathcal{K}_\zeta = \mathcal{K}\omega_\zeta * \mathcal{W}_\zeta$, where ω_ζ is a $W^{1,\infty}(\mathbb{R}^d)$ cut-off function (see Section 3.1.3 and definition (3.11) for the exact approximating sequence), in comparison to the interaction scaling $\mathcal{W}_\beta * \rho_{\sigma,\beta,\zeta}$. The two different scalings β and ζ allow us to establish a result, for which the kernel regularization on the particle level does not need to be of logarithmic type but of power-law type only.

3.1.3 Main results and key ideas

We impose the following hypotheses:

(H1) Data: Let $0 < s < 1$, $d \geq 2$.

(H2) $\rho^0 \in L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ satisfies $\rho^0 \geq 0$ in $\mathbb{R}^d$ and $\int_{\mathbb{R}^d} \rho^0(x)|x|^{2d/(d-2s)} \mathrm{d}x < \infty$.

(H3) Nonlinearity: $f \in C^1([0,\infty))$ is nondecreasing, $f(0) = 0$, and $u \mapsto uf(u)$ for $u > 0$ is strictly convex.

Let us discuss these assumptions. We assume that $d \geq 2$; the case $d = 1$ can be treated if $s < 1/2$; see [17]. Extending the range of s to $s < 0$ leads to the fractional (higher-order) thin-film equation, which is studied in [77]. The case $1 < s < d/2$ may be considered too, since it yields better regularity results; we leave the details to the reader. On the other hand, the case $s \geq d/2$ is more delicate since the multiplier in the definition of $(-\Delta)^{-s}$ using Fourier transforms does not define a tempered distribution. The case $s = d/2$ for $d \leq 2$ (with a logarithmic Riesz kernel) was analyzed in [42]. We need the moment bound for the initial datum ρ^0 to prove the same moment bound for ρ_σ, which in turn is used several times, for instance to show the entropy balance and the convergence $\rho_\sigma \to \rho$ as $\sigma \to 0$ in the sense of $C^0_{\text{weak}}([0,T]; L^1(\mathbb{R}^d))$. The monotonicity of f and the strict convexity of $u \mapsto uf(u)$ are needed to prove the strong convergence of (ρ_σ), which then allows us to identify the limit of $(f_\sigma(\rho_\sigma))$. An example of a function satisfying Hypothesis (H3) is $f(u) = u^\beta$ with $\beta \geq 1$.
Our first result is concerned with the existence analysis of (3.1). This result is needed to prove the main theorem below. We write $\|\cdot\|_p$ for the $L^p(\mathbb{R}^d)$ norm and define the so-called entropy density $h : [0,\infty) \to \mathbb{R}$ by

$$h(u) = \int_0^u \int_1^v \frac{f'(w)}{w} \mathrm{d}w\mathrm{d}v \quad \text{for } u \geq 0.$$

Theorem 3.1 (Existence of weak solutions to (3.1))**.** *Let Hypotheses (H1)–(H3) hold. Then there exists a weak solution $\rho \geq 0$ to (3.1) satisfying* (i) *the regularity*

$$\rho \in L^\infty(0,\infty; L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)), \quad \nabla(-\Delta)^{-s/2} f(\rho) \in L^2(0,\infty; L^2(\mathbb{R}^d)),$$
$$\partial_t \rho \in L^2(0,\infty; H^{-1}(\mathbb{R}^d)),$$

(ii) *the weak formulation*

$$\int_0^T \langle \partial_t \rho, \phi \rangle \mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} \rho \nabla(-\Delta)^{-s} f(\rho) \cdot \nabla\phi \mathrm{d}x\mathrm{d}t = 0 \tag{3.8}$$

for all $\phi \in L^2(0,T; H^1(\mathbb{R}^d))$ and $T > 0$, (iii) *the initial datum $\rho(0) = \rho^0$ in the sense of $H^{-1}(\mathbb{R}^d)$, and* (iv) *the following properties for $t > 0$:*

- *Mass conservation: $\|\rho(t)\|_1 = \|\rho^0\|_1$,*
- *Dissipation of the L^∞ norm: $\|\rho(t)\|_\infty \leq \|\rho^0\|_\infty$,*

- *Moment estimate:* $\sup_{0<t<T} \int_{\mathbb{R}^d} \rho(x,t)|x|^{2d/(d-2s)}\mathrm{d}x \le C(T)$,
- *Entropy inequality:*

$$\int_{\mathbb{R}^d} h(\rho(t))\mathrm{d}x + \int_0^t \int_{\mathbb{R}^d} |\nabla(-\Delta)^{-s/2} f(\rho)|^2 \mathrm{d}x\mathrm{d}s \le \int_{\mathbb{R}^d} h(\rho^0)\mathrm{d}x.$$

Note that the Hardy–Littlewood–Sobolev-type inequality (3.68) (see Appendix 3.B) implies that

$$\|\rho\nabla(-\Delta)^{-s} f(\rho)\|_2 = \|\rho(-\Delta)^{-s/2}[\nabla(-\Delta)^{-s/2} f(\rho)]\|_2 \le C\|\rho\|_{d/(2s)}\|\nabla(-\Delta)^{-s/2} f(\rho)\|_2,$$

such that $\rho\nabla(-\Delta)^{-s} f(\rho) \in L^2(\mathbb{R}^d)$, and the weak formulation (3.8) is well defined.
The main ideas of the proof of Theorem 3.1 are as follows. A priori estimates for strong solutions ρ_σ to the regularized equation (3.7) are derived from mass conservation, the entropy inequality, and energy-type bounds. The energy-type bound allows us to show, for sufficiently small $\sigma > 0$, that the L^∞ norm of ρ_σ is bounded by the L^∞ norm of ρ^0, up to some factor depending on the moment bound for ρ^0. The existence of a strong solution ρ_σ is proved by regularizing (3.7) in a careful way to deal with the singular kernel. The regularized equation is solved locally in time by Banach's fixed-point theorem. Entropy estimates allow us to extend this solution globally in time and to pass to the de-regularization limit. The second step is the limit $\sigma \to 0$ in (3.7). Since the bounds only provide weak convergence of (a subsequence of) ρ_σ, the main difficulty is the identification of the nonlinear limit $f_\sigma(\rho_\sigma)$. This is done by applying the div-curl lemma and exploiting the monotonicity of f and the strict convexity of $u \mapsto uf(u)$ [45].
We already mentioned that the existence of local smooth solutions ρ_σ to (3.7) has been proven in [34]. However, we provide an independent proof that allows for global strong solutions and that yields a priori estimates needed in the mean-field limit.
Our second and main result is the propagation of chaos, which shows a mean-field-type convergence of the particle system (3.3) to a solution of (3.1). To define our particle system properly, we need some definitions. Introduce the smooth approximation

$$f_\sigma(u) = \int_0^u (\Gamma_\sigma * (f'1_{[0,\infty)}))(w)\widetilde{\Xi}(\sigma w)\mathrm{d}w \quad u \in \mathbb{R}, \tag{3.9}$$

where the mollifier Γ_σ for $\sigma > 0$ is given by $\Gamma_\sigma(x) = \sigma^{-1}\Gamma_1(x/\sigma)$, and $\Gamma_1 \in C_0^\infty(\mathbb{R})$ satisfies $\Gamma_1 \ge 0$, $\|\Gamma_1\|_1 = 1$, while the cutoff function $\widetilde{\Xi} \in C_0^\infty(\mathbb{R})$ satisfies $0 \le \widetilde{\Xi} \le 1$ in $\mathbb{R}$ and $\widetilde{\Xi}(x) = 1$ for $|x| \le 1$. Then, thanks to Γ_σ, we have $f_\sigma \in C^\infty(\mathbb{R})$. The cut-off function guarantees that the derivatives $\mathrm{D}^k f_\sigma$ are bounded and compactly supported for all $k \ge 1$. Furthermore, it holds that $f'_\sigma \ge 0$, $f_\sigma(0) = 0$.
In a similar way, we introduce the mollifier function $\mathcal{W}_\beta$ for $\beta > 0$ and $x \in \mathbb{R}^d$ by

$$\mathcal{W}_\beta(x) = \beta^{-d}\mathcal{W}_1(\beta^{-1}x), \quad \mathcal{W}_1 \in C_0^\infty(\mathbb{R}^d) \text{ is symmetric}, \ \mathcal{W}_1 \ge 0, \ \|\mathcal{W}_1\|_1 = 1. \tag{3.10}$$

Let us define the cut-off version of the singular kernel $\mathcal{K}$ by

$$\widetilde{\mathcal{K}}_\zeta := \mathcal{K}\omega_\zeta, \text{ where the cut-off function } \omega_\zeta \in W^{1,\infty}(\mathbb{R}^d) \text{ is such that}$$

$$0 \le \omega_\zeta(x) \le 1 \text{ for } x \in \mathbb{R}^d, \quad \|\nabla \omega_\zeta\|_\infty \le 2\zeta, \tag{3.11}$$
$$\omega_\zeta(x) = 1 \text{ for all } |x| \le \zeta^{-1}, \quad \omega_\zeta(x) = 0 \text{ for all } |x| \ge 2\zeta^{-1}.$$

Then the regularized kernel $\mathcal{K}_\zeta$ is given by

$$\mathcal{K}_\zeta(x) := \widetilde{\mathcal{K}}_\zeta * \mathcal{W}_\zeta(x) \quad \text{for all } x \in \mathbb{R}^d,$$

where $\zeta > 0$. Let the cutoff function $\Xi \in C_0^\infty(\mathbb{R}^d)$ satisfy $0 \le \Xi \le 1$ in $\mathbb{R}^d$ and $\Xi(x) = 1$ for $|x| \le 1$. Then we define the regularized initial datum for $x \in \mathbb{R}^d$ by

$$\rho_\sigma^0(x) = \kappa_\sigma(\mathcal{W}_\sigma * \rho^0)(x)\Xi(\sigma x), \quad \text{where } \kappa_\sigma = \frac{\int_{\mathbb{R}^d} \rho^0(y)\mathrm{d}y}{\int_{\mathbb{R}^d} (\mathcal{W}_\sigma * \rho^0)(y)\Xi(\sigma y)\mathrm{d}y}. \tag{3.12}$$

This definition guarantees the mass conservation since $\|\rho_\sigma^0\|_1 = \|\rho^0\|_1$; see Section 3.2.1.
Note that our particle system (3.3) depends on 4 parameters: $N \in \mathbb{N}$ denotes the number of particles, $\beta > 0$ models the strength of interaction between the particles, $\zeta > 0$ describes the regularization of the singular kernel $\mathcal{K}$, and $\sigma > 0$ is a measure of the additional diffusion. The quantities (β, ζ, σ) are regularization parameters needed to overcome the singularity of the kernel $\mathcal{K}$ and the (possible) degeneracy of the nonlinearity f.
In the limit $N \to \infty$, $(\beta, \zeta, \sigma) \to 0$, we prove the following propagation-of-chaos result.

Theorem 3.2 (Propagation of chaos)**.** *Let $\zeta^{-2s-1} \le C_1 N^{1/4}$ and $\beta^{-3d-7} \le \varepsilon \log N$ for some constants C_1, $\varepsilon > 0$, and let $\mathrm{P}^k_{N,\sigma,\beta,\zeta}(t)$ be the joint distribution of $(X_1^N(t), \ldots, X_k^N(t))$ for $k \ge 1$ and $t \in (0, T)$. Then there exists a subsequence in σ such that*

$$\lim_{\sigma \to 0} \lim_{N\to\infty,\, (\beta,\zeta)\to 0} \mathrm{P}^k_{N,\sigma,\beta,\zeta}(t) = \mathrm{P}^{\otimes k}(t) \quad \textit{in the sense of distributions},$$

where the limit is locally uniform in t, and the measure $\mathrm{P}(t)$ is absolutely continuous with respect to the Lebesgue measure with the probability density function $\rho(t)$ that is a weak solution to (3.1).

It is well known (see, e.g., Proposition 1.1.2 in the introduction of this thesis) that the result of Theorem 3.2 implies the weak convergence of the empirical measure associated to the particle system (3.3) towards the deterministic measure $\rho(t)$, i.e.

$$\mu_{N,\sigma,\beta,\zeta}(t) = \sum_{i=1}^N \delta_{X_i^N(t)} \rightharpoonup \rho(t),$$

for a subsequence in σ. Furthermore, Theorem 3.2 shows that at any time $t > 0$, in the limit $N \to \infty$, $(\beta, \zeta, \sigma) \to 0$, any finite selection of k particles in (3.3) becomes independent with limiting distribution $\rho^{\otimes k}(t)$.
If equation (3.1) was uniquely solvable, we would obtain the convergence of the whole sequence in σ. Unfortunately, the regularity of the solution ρ to (3.1) is too weak to conclude the uniqueness of weak solutions. Up to our knowledge, none of the known methods, such as [8, 35], seem to be applicable to equation (3.1).
Theorem 3.2 is proved in two steps: First, we show (strong) error estimates between particles of systems (3.3) and (3.6), respectively; see Proposition 3.3 below. Second, we show the weak convergence of (a subsequence of) ρ_σ to a solution ρ to (3.1); see Corollary 3.13.

Proposition 3.3 (Error estimate for the stochastic system)**.** *Let X_i^N and $\widehat{X}_i^N$ be the solutions to* (3.3) *and* (3.6)*, respectively. We assume that $\zeta^{-2s-1} \le C_1 N^{1/4}$ for some constant $C_1 > 0$. Let $\delta \in (0, 1/4)$ and $a := \min\{1, d-2s\} > 0$. Then there exist constants $\varepsilon > 0$, depending on σ and δ, and $C_2 > 0$, depending on σ and T, such that if $\beta^{-3d-7} \le \varepsilon \log N$ then*

$$\mathbb{E}\bigg(\sup_{0<s<T}\max_{i=1,\dots,N}|(X_i^N - \widehat{X}_i^N)(s)|\bigg) \le C_2(\beta+\zeta^a) \to 0 \text{ as } (N,\zeta,\beta) \to (\infty,0,0).$$

The proposition is proved by estimating the differences

$$E_1(t) := \mathbb{E}\bigg(\sup_{0<s<t}\max_{i=1,\dots,N}|(X_i^N - \bar{X}_i^N)(s)|\bigg),$$
$$E_2(t) := \mathbb{E}\bigg(\sup_{0<s<t}\max_{i=1,\dots,N}|(\bar{X}_i^N - \widehat{X}_i^N)(s)|\bigg),$$

and applying the triangle inequality. For the first difference, we estimate expressions like $\|\mathrm{D}^k\mathcal{K}_\zeta * u\|_\infty$ for appropriate functions u and $\|\mathrm{D}^k\mathcal{W}_\beta\|_\infty$ for $k \in \mathbb{N}$ in terms of negative powers of β (here, D^k denotes the kth-order partial derivatives). Using properties of Riesz potentials, in particular Hardy–Littlewood–Sobolev-type inequalities (see Lemmas 3.22 and 3.23), we show that for some $\mu_i > 0$ ($i = 1, 2, 3$),

$$E_1(t) \le C(\sigma)\beta^{-\mu_1}\int_0^t E_1(s)\mathrm{d}s + C(\sigma)\beta^{-\mu_2}\zeta^{-\mu_3}N^{-1/2}.$$

By applying the Gronwall lemma and choosing a logarithmic scaling for β and an algebraic scaling for ζ with respect to N, we infer that $E_1(t) \le C(\sigma)N^{-\mu_4}$ for some $\mu_4 \in (0, 1/4)$. For the second difference E_2, we need the estimates $\|\mathcal{W}_\beta * u - u\|_\infty \le C(\sigma)\beta$ (Lemma 3.21), and $\|(\mathcal{K}_\zeta - \mathcal{K}) * \rho_\sigma\|_\infty \le C(\sigma)\zeta^a$, $\|\rho_{\sigma,\beta,\zeta} - \rho_\sigma\|_\infty \le C(\sigma)(\beta+\zeta^a)$ (Proposition 3.14), recalling that $a = \min\{1, d-2s\}$. The proof of these estimates is very technical. The idea is to apply several times fractional Gagliardo–Nirenberg inequalities that are proved in Appendix 3.B and Hardy–Littlewood–Sobolev inequalities that are recalled in Lemmas 3.22–3.23. Then, after suitable computations,

$$E_2(t) \le C(\sigma)(\beta+\zeta^a) + C(\sigma)\int_0^t E_2(s)\mathrm{d}s,$$

and we conclude with Gronwall's lemma that $E_2(t) \le C(\sigma)(\beta+\zeta^a)$.

The chapter is organized as follows. The existence of global nonnegative weak solutions to (3.1) is proved in Section 3.2 by establishing an existence analysis for (3.7) and performing the limit $\sigma \to 0$. Some uniform estimates for the solution $\rho_{\rho,\beta,\zeta}$ to (3.5) and for the difference $\rho_{\sigma,\beta,\zeta} - \rho_\sigma$ are shown in Section 3.3. Section 3.4 is devoted to the proof of the error estimate in Theorem 3.3 and the propagation of chaos in Theorem 3.2. In Appendices 3.A–3.C we recall some auxiliary results and Hardy–Littlewood–Sobolev-type inequalities, prove new variants of fractional Gagliardo–Nirenberg inequalities, and formulate a result on parabolic regularity.

Notation

We write $\|\cdot\|_p$ for the $L^p(\mathbb{R}^d)$ or $L^p(\mathbb{R})$ norm with $1 \le p \le \infty$. The ball around the origin with radius $R > 0$ is denoted by B_R. The partial derivative $\partial/\partial x_i$ is abbreviated as ∂_i for $i = 1, \ldots, d$, and D^α denotes a partial derivative of order $|\alpha|$, where $\alpha \in \mathbb{N}_0^d$ is a multiindex. The notation D^k refers to the kth-order tensor of partial derivatives of order $k \in \mathbb{N}$. In this situation, the norm $\|\mathrm{D}^k u\|_p$ is the sum of all L^p norms of partial derivatives of u of order k. Finally, $C > 0$, $C_1 > 0$, etc. denote generic constants with values changing from line to line.

3.2 Analysis of the fractional porous media equation

In this section, we prove the existence of global nonnegative weak solutions to (3.1). We first prove the existence of a solution ρ_σ to (3.7) by a fixed-point argument and then perform the limit $\sigma \to 0$. In Section 3.2.1, we prove some basic estimates for a strong solution ρ_σ to (3.7). Entropy and moment estimates as well as higher-order estimates are derived in Sections 3.2.2 and 3.2.3, respectively. The existence of a unique strong solution to (3.7) is proved in Section 3.2.4 using a regularized version of (3.7) and Banach's fixed-point theorem. The strong $L^1(\mathbb{R}^d)$ limit $\sigma \to 0$ is performed in Section 3.2.5 using the div-curl lemma. Finally, Section 3.2.6 is concerned with the proof of a time-uniform weak $L^1(\mathbb{R}^d)$ limit of (ρ_σ), which is needed in the proof of Proposition 3.3. Recall definition (3.12) of the number κ_σ, which is stated in (iv) below.

Proposition 3.4. *Let Hypotheses (H1)–(H3) hold. Then for all $\sigma > 0$, there exists a unique weak solution $\rho_\sigma \ge 0$ to* (3.7) *satisfying* (i) *the regularity*

$$\rho_\sigma \in L^\infty(0,\infty; L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)) \cap C^0([0,\infty); L^2(\mathbb{R}^d)),$$
$$\nabla\rho_\sigma \in L^2(0,\infty; L^2(\mathbb{R}^d)), \quad \partial_t\rho_\sigma \in L^2(0,\infty; H^{-1}(\mathbb{R}^d)),$$

(ii) *the weak formulation of* (3.7) *with test functions $\phi \in L^2(0,T; H^1(\mathbb{R}^d))$,* (iii) *the inital datum $\rho_\sigma(0) = \rho_\sigma^0$ in $L^2(\mathbb{R}^d)$, and* (vi) *the following properties for $t > 0$, which are uniform in σ for sufficiently small $\sigma > 0$:*

- *Mass conservation:* $\|\rho_\sigma(t)\|_1 = \|\rho^0\|_1$.
- *Dissipation of the L^∞ norm:* $\|\rho_\sigma\|_{L^\infty(0,\infty;L^\infty(\mathbb{R}^d))} \le \kappa_\sigma \|\rho^0\|_{L^\infty(\mathbb{R}^d)} \le C\|\rho^0\|_{L^\infty(\mathbb{R}^d)}$.
- *Moment estimate:* $\sup_{t\in[0,\infty)} \int_{\mathbb{R}^d} \rho_\sigma(x,t)|x|^{\frac{2d}{d-2s}}\mathrm{d}x \le C_T$.
- *Entropy inequality:*

$$\int_{\mathbb{R}^d} h(\rho_\sigma(T))\mathrm{d}x + 4\sigma\int_0^T\int_{\mathbb{R}^d} f_\sigma'(\rho_\sigma)|\nabla\sqrt{\rho_\sigma}|^2\mathrm{d}x\mathrm{d}t + \int_0^T\int_{\mathbb{R}^d} |\nabla(-\Delta)^{-s/2} f_\sigma(\rho_\sigma)|^2\mathrm{d}x\mathrm{d}t \le \int_{\mathbb{R}^d} h(\rho_\sigma^0)\mathrm{d}x \quad \textit{for all } T > 0.$$

Additionally, for any $T > 0$, $1 < p < \infty$, and $2 \le q < \infty$, there exists $C > 0$, depending on T, σ, p, and q, such that

$$\|\rho_\sigma\|_{L^p(0,T;W^{3,p}(\mathbb{R}^d))} + \|\partial_t\rho_\sigma\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + \|\rho_\sigma\|_{C^0([0,T];W^{2,1}(\mathbb{R}^d)\cap W^{3,q}(\mathbb{R}^d))} \le C,$$

i.e., ρ_σ is even a strong solution to (3.7) and $\rho_\sigma \in C^0([0,T];W^{2,1}(\mathbb{R}^d)\cap W^{3,q}(\mathbb{R}^d))$ for $q \ge 2$.

3.2.1 Basic estimates for ρ_σ

We prove a priori estimates in L^p spaces and an energy-type estimate. Let $\sigma \in (0,1)$ and let ρ_σ be a nonnegative strong solution to (3.7). Integration of (3.7) in $\mathbb{R}^d$ and the definition of ρ_σ^0 immediately yield the mass conservation

$$\|\rho_\sigma(t)\|_1 = \|\rho_\sigma^0\|_1 = \|\rho^0\|_1 \quad \text{for } t > 0. \tag{3.13}$$

Lemma 3.5 (Energy-type estimate)**.** *Let $F \in C^2([0,\infty))$ be convex and let $F(\rho_\sigma^0) \in L^1(\mathbb{R}^d)$. Then*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} F(\rho_\sigma)\mathrm{d}x &= -\sigma\int_{\mathbb{R}^d} F''(\rho_\sigma)|\nabla\rho_\sigma|^2\mathrm{d}x \\ &\quad - \frac{c_{d,1-s}}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{(G(\rho_\sigma(x)) - G(\rho_\sigma(y)))(f_\sigma(\rho_\sigma(x)) - f_\sigma(\rho_\sigma(y)))}{|x-y|^{d+2(1-s)}}\mathrm{d}x\mathrm{d}y \le 0,\end{aligned} \tag{3.14}$$

where $G(u) := \int_0^u vF''(v)\mathrm{d}v$ for $u \ge 0$ and $c_{d,1-s}$ is defined after (3.2).

Proof. First, we assume that F'' is additionally bounded. Then $F'(\rho_\sigma) - F'(0)$ is an admissible test function in the weak formulation of (3.7), since $|F'(\rho_\sigma) - F'(0)| \le \|F''\|_\infty|\rho_\sigma|$. It follows from definition (3.66) of the fractional Laplacian and integration by parts that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} F(\rho_\sigma)\mathrm{d}x + \sigma\int_{\mathbb{R}^d} F''(\rho_\sigma)|\nabla\rho_\sigma|^2\mathrm{d}x &= -\int_{\mathbb{R}^d} F''(\rho_\sigma)\rho_\sigma\nabla\rho_\sigma\cdot\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\mathrm{d}x \\ = -\int_{\mathbb{R}^d}\nabla G(\rho_\sigma)\cdot\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\mathrm{d}x &= -\int_{\mathbb{R}^d} G(\rho_\sigma)(-\Delta)^{1-s}f_\sigma(\rho_\sigma)\mathrm{d}x \\ = -c_{d,1-s}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} G(\rho_\sigma(x))\frac{f_\sigma(\rho_\sigma(x)) - f_\sigma(\rho_\sigma(y))}{|x-y|^{d+2(1-s)}}\mathrm{d}x\mathrm{d}y.\end{aligned}$$

A symmetrization of the last integral yields (3.14).
In the general case, we introduce $F_k(u) = F(0) + F'(0)u + \int_0^u\int_0^v \min\{F''(w), k\}\mathrm{d}w\mathrm{d}v$ for $k > 0$. Then $F_k''(u)$ is bounded and (3.14) follows for F replaced by F_k. The result follows after taking the limit $k \to \infty$ using monotone convergence. □

We need a bound on κ_σ, defined in (3.12), to derive uniform $L^\infty(\mathbb{R}^d)$ bounds for ρ_σ.

Lemma 3.6 (Bound for κ_σ)**.** *There exists $C > 0$ such that, for sufficiently small $\sigma > 0$,*

$$1 \le \kappa_\sigma \le \frac{1}{1 - C\sigma E}, \quad \textit{where } E := \frac{1}{\|\rho^0\|_1}\int_{\mathbb{R}^d}(1 + |x|^{2d/(d-2s)})\rho^0(x)\mathrm{d}x.$$

Proof. By Young's convolution inequality (Lemma 3.19), we have

$$\int_{\mathbb{R}^d}(\mathcal{W}_\sigma * \rho^0)(x)\Xi(\sigma x)\mathrm{d}x \leq \|\mathcal{W}_\sigma * \rho^0\|_1 \leq \|\mathcal{W}_\sigma\|_1\|\rho^0\|_1 = \|\rho^0\|_1,$$

which shows that $\kappa_\sigma \geq 1$. To prove the upper bound, we use the triangle inequality $|x| \leq |x-y| + |y|$:

$$\begin{aligned}
\int_{\mathbb{R}^d}(\mathcal{W}_\sigma * \rho^0)(x)\Xi(\sigma x)\mathrm{d}x &\geq \int_{\{|x|\leq 1/\sigma\}}\int_{\mathbb{R}^d}\mathcal{W}_\sigma(x-y)\rho^0(y)\mathrm{d}y\mathrm{d}x \\
&= \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d}\mathcal{W}_\sigma(x-y)\mathrm{d}x\Big)\rho^0(y)\mathrm{d}y - \int_{\{|x|>1/\sigma\}}\int_{\mathbb{R}^d}\mathcal{W}_\sigma(x-y)\rho^0(y)\mathrm{d}y\mathrm{d}x \\
&\geq \int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y - \sigma^{2d/(d-2s)}\int_{\{|x|>1/\sigma\}}\int_{\mathbb{R}^d}|x|^{2d/(d-2s)}\mathcal{W}_\sigma(x-y)\rho^0(y)\mathrm{d}y\mathrm{d}x \\
&\geq \int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y - \sigma^{2d/(d-2s)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|x-y|^{2d/(d-2s)}\mathcal{W}_\sigma(x-y)\rho^0(y)\mathrm{d}y\mathrm{d}x \\
&\quad - \sigma^{2d/(d-2s)}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|y|^{2d/(d-2s)}\mathcal{W}_\sigma(x-y)\rho^0(y)\mathrm{d}y\mathrm{d}x.
\end{aligned}$$

Using the property $\int_{\mathbb{R}^d}|z|^{2d/(d-2s)}\mathcal{W}_\sigma(z)\mathrm{d}z \leq C\sigma^{2d/(d-2s)}$ for the second term on the right-hand side and $\|\mathcal{W}_\beta\|_{L^1(\mathbb{R}^d)} = 1$ for the third term, we find that

$$\begin{aligned}
\int_{\mathbb{R}^d}(\mathcal{W}_\sigma * \rho^0)(x)\Xi(\sigma x)\mathrm{d}x \geq \int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y - C\sigma^{4d/(d-2s)}\int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y \\
- \sigma^{2d/(d-2s)}\int_{\mathbb{R}^d}|y|^{2d/(d-2s)}\rho^0(y)\mathrm{d}y.
\end{aligned}$$

Because of $\sigma^{2d/(d-2s)} \leq \sigma$ for $\sigma \leq 1$, we obtain

$$\begin{aligned}
\frac{\|\rho^0\|_1}{\kappa_\sigma} &= \int_{\mathbb{R}^d}(\mathcal{W}_\sigma * \rho^0)(x)\Xi(\sigma x)\mathrm{d}x \geq \int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y - C\sigma\int_{\mathbb{R}^d}(1+|y|^{2d/(d-2s)})\rho^0(y)\mathrm{d}y \\
&\geq \int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y - C\sigma\int_{\mathbb{R}^d}\rho^0(y)\mathrm{d}y \cdot E = \|\rho^0\|_1(1 - C\sigma E),
\end{aligned}$$

which proves the lemma. □

Lemma 3.7 (Bounds for ρ_σ)**.** *The following bounds hold:*

$$\|\rho_\sigma(t)\|_\infty \leq \kappa_\sigma\|\rho^0\|_\infty \leq C\|\rho^0\|_\infty, \quad t > 0, \tag{3.15}$$

$$\sqrt{\sigma}\|\rho_\sigma\|_{L^2(0,T;H^1(\mathbb{R}^d))} \leq \|\rho^0\|_2, \tag{3.16}$$

where (3.15) *holds for sufficiently small* $\sigma > 0$.

Lemma 3.7 and mass conservation imply that $\|\rho_\sigma(t)\|_p$ is bounded for all $t > 0$ and $1 \leq p \leq \infty$. Observe that $\kappa_\sigma \to 1$ as $\sigma \to 0$. So, if $\rho_\sigma(t) \to \rho(t)$ a.e., the dissipation of the L^∞ norm follows, as stated in Theorem 3.1 (iv).

Proof. The convexity of F shows that G, defined in Lemma 3.5, is nondecreasing. Therefore, $(\mathrm{d}/\mathrm{d}t)\int_{\mathbb{R}^d} F(\rho_\sigma)\mathrm{d}x \le 0$ and

$$\sup_{t>0}\int_{\mathbb{R}^d} F(\rho_\sigma(t))\mathrm{d}x \le \int_{\mathbb{R}^d} F(\rho_\sigma^0)\mathrm{d}x.$$

We choose a convex function $F \in C^2([0,\infty))$ such that $F(u) = 0$ for $u \le \|\rho_\sigma^0\|_\infty$, $F(u) > 0$ for $u > \|\rho_\sigma^0\|_\infty$ and satisfying $F(u) \le Cu$ for $u \to \infty$. Then

$$0 \le \int_{\mathbb{R}^d} F(\rho_\sigma(t))\mathrm{d}x \le \int_{\mathbb{R}^d} F(\rho_\sigma^0)\mathrm{d}x = 0 \quad \text{for } t > 0.$$

Consequently, $\rho_\sigma(x,t) \le \|\rho_\sigma^0\|_\infty \le \kappa_\sigma\|\rho^0\|_\infty$ for $t > 0$, showing the $L^\infty(\mathbb{R}^d)$ bound. Finally, choosing $F(u) = u^2$ in Lemma 3.5, the $L^2(0,T;H^1(\mathbb{R}^d))$ estimate follows. □

3.2.2 Entropy and moment estimates

We need a fractional derivative estimate for $f_\sigma(\rho_\sigma)$, which is not an immediate consequence of Lemma 3.5. To this end, we define the entropy density

$$h_\sigma(u) = \int_0^u\int_1^v \frac{f_\sigma'(w)}{w}\mathrm{d}w\mathrm{d}v, \quad u \ge 0.$$

Lemma 3.8 (Entropy balance)**.** *It holds for all $t > 0$ that*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} h_\sigma(\rho_\sigma)\mathrm{d}x + 4\sigma\int_{\mathbb{R}^d} f_\sigma'(\rho_\sigma)|\nabla\rho_\sigma^{1/2}|^2\mathrm{d}x + \int_{\mathbb{R}^d}|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)|^2\mathrm{d}x = 0.$$

In particular, for all $T > 0$, there exists $C > 0$ such that

$$\|f_\sigma(\rho_\sigma)\|_{L^2(0,T;H^{1-s}(\mathbb{R}^d))} \le C. \tag{3.17}$$

Proof. The idea is to apply Lemma 3.5. Since $h_\sigma \notin C^2([0,\infty))$, we cannot use the lemma directly. Instead, we apply it to the regularized function

$$h_\sigma^\delta(u) = \int_0^u\int_1^v \frac{f_\sigma'(w)}{w+\delta}\mathrm{d}w\mathrm{d}v, \quad u \ge 0,$$

where $\delta > 0$. Choosing $F = h_\sigma^\delta$ in Lemma 3.5 gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} h_\sigma^\delta(\rho_\sigma)\mathrm{d}x &+ 4\sigma\int_{\mathbb{R}^d} f_\sigma'(\rho_\sigma)\frac{\rho_\sigma}{\rho_\sigma+\delta}|\nabla\rho_\sigma^{1/2}|^2\mathrm{d}x \\ &= -\frac{c_{d,1-s}}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{(f_\sigma^\delta(\rho_\sigma(x)) - f_\sigma^\delta(\rho_\sigma(y)))(f_\sigma(\rho_\sigma(x)) - f_\sigma(\rho_\sigma(y)))}{|x-y|^{d+2(1-s)}}\mathrm{d}x\mathrm{d}y,\end{aligned} \tag{3.18}$$

where $f_\sigma^\delta(u) := \int_0^u (v/(v+\delta))f_\sigma'(v)\mathrm{d}v$ for $u \ge 0$.
Step 1: Estimate of h_σ^δ. The pointwise limit $h_\sigma^\delta(\rho_\sigma) \to h_\sigma(\rho_\sigma)$ holds a.e. in $\mathbb{R}^d \times (0,T)$ as $\delta \to 0$. We observe that for all $0 < u \le 1$,

$$|h_\sigma^\delta(u)| \le \sup_{0<v<1} f'(v)\int_0^u\int_v^1 \frac{\mathrm{d}w}{w}\mathrm{d}v \le Cu(|\log u| + 1),$$

while for all $u > 1$, since $f_\sigma' \geq 0$ in $[0, \infty)$,

$$|h_\sigma^\delta(u)| \leq \int_0^1 \int_v^1 \frac{f_\sigma'(w)}{w+\delta} \mathrm{d}w\mathrm{d}v + \int_1^u \int_1^v \frac{f_\sigma'(w)}{w+\delta} \mathrm{d}w\mathrm{d}v$$
$$\leq C + \int_1^u \int_1^v f_\sigma'(w)\mathrm{d}w\mathrm{d}v \leq C + \int_0^u f_\sigma(v)\mathrm{d}v \leq C + uf_\sigma(u).$$

The last inequality follows after integration of $f_\sigma(v) \leq f_\sigma(v) + vf_\sigma'(v) = (vf_\sigma(v))'$ in $(0, u)$. Therefore, since $\rho_\sigma \leq \|\rho_\sigma^0\|_\infty$ a.e. in $\mathbb{R}^d \times (0, \infty)$, we find that

$$|h_\sigma^\delta(\rho_\sigma)| \leq C\rho_\sigma(|\log \rho_\sigma| + 1)1_{\{\rho_\sigma \leq 1\}} + C1_{\{\rho_\sigma > 1\}} \leq C(\rho_\sigma^\theta + \rho_\sigma),$$

where $\theta \in (0, 1)$ is arbitrary, and consequently, because of mass conservation,

$$\int_{\mathbb{R}^d} |h_\sigma^\delta(\rho_\sigma)|\mathrm{d}x \leq C + C\int_{\mathbb{R}^d} \rho_\sigma^\theta \mathrm{d}x. \tag{3.19}$$

Step 2: Estimate of $\int_{\mathbb{R}^d} \rho_\sigma^\theta \mathrm{d}x$. Let $0 < \alpha < 1$ and $d/(d+\alpha) < \theta < 1$. Then, by Young's inequality,

$$\int_{\mathbb{R}^d} \rho_\sigma^\theta \mathrm{d}x = \int_{\mathbb{R}^d} (1 + |x|^2)^{\alpha\theta/2}\rho_\sigma^\theta(1 + |x|^2)^{-\alpha\theta/2}\mathrm{d}x$$
$$\leq \int_{\mathbb{R}^d} (1 + |x|^2)^{\alpha/2}\rho_\sigma \mathrm{d}x + C\int_{\mathbb{R}^d} (1 + |x|^2)^{-\alpha\theta/(2(1-\theta))}\mathrm{d}x$$
$$\leq \int_{\mathbb{R}^d} (1 + |x|^2)^{\alpha/2}\rho_\sigma \mathrm{d}x + C,$$

since the choice of θ guarantees that $-\alpha\theta/(2(1-\theta)) < -d/2$, so $\int_{\mathbb{R}^d}(1+|x|^2)^{-\alpha\theta/(2(1-\theta))}\mathrm{d}x < \infty$. To control the right-hand side, we need to bound a suitable moment of ρ_σ.
For this, we use the test function $(1 + |x|^2)^{\alpha/2}\xi_k$ in the weak formulation of (3.7), where $\xi_k \in C_0^2(\mathbb{R}^d)$ is a cut-off function with the properties

$$\xi_k(x) = 1 \quad \text{for } |x| \leq k, \quad \xi_k(x) = 0 \quad \text{for } |x| \geq 2k,$$
$$k|\nabla\xi_k(x)| + k^2|\Delta\xi_k(x)| \leq C, \quad 0 \leq \xi_k(x) \leq 1 \quad \text{for } x \in \mathbb{R}^d,$$

and $k > 1$ is arbitrary. We find that

$$\int_{\mathbb{R}^d} (1 + |x|^2)^{\alpha/2}\xi_k\rho_\sigma(t)\mathrm{d}x = \int_{\mathbb{R}^d} (1 + |x|^2)^{\alpha/2}\xi_k\rho_\sigma^0 \mathrm{d}x + \sigma\int_0^t \int_{\mathbb{R}^d} \rho_\sigma\xi_k\Delta(1 + |x|^2)^{\alpha/2}\mathrm{d}x\mathrm{d}s$$
$$+ \sigma\int_0^t \int_{\mathbb{R}^d} \rho_\sigma\big(2\nabla[(1 + |x|^2)^{\alpha/2}] \cdot \nabla\xi_k + (1 + |x|^2)^{\alpha/2}\Delta\xi_k\big)\mathrm{d}x\mathrm{d}s$$
$$- \alpha\int_0^t \int_{\mathbb{R}^d} \rho_\sigma\xi_k(1 + |x|^2)^{\alpha/2-1}x \cdot \nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\mathrm{d}x\mathrm{d}s$$
$$- \int_0^t \int_{\mathbb{R}^d} \rho_\sigma(1 + |x|^2)^{\alpha/2}\nabla\xi_k \cdot \nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\mathrm{d}x\mathrm{d}s.$$

Since $\alpha < 1$ and $0 \le \xi_k \le 1$ in $\mathbb{R}^d$, the terms involving $\Delta(1+|x|^2)^{\alpha/2}$ and $x(1+|x|^2)^{\alpha/2-1}$ are bounded in $\mathbb{R}^d$. It follows from the choice of ξ_k that

$$|\nabla[(1+|x|^2)^{\alpha/2}]\cdot\nabla\xi_k| + |(1+|x|^2)^{\alpha/2}\Delta\xi_k| \le Ck^{\alpha-2}, \quad (1+|x|^2)^{\alpha/2}|\nabla\xi_k| \le Ck^{\alpha-1}.$$

Thus, taking into account the assumption on ρ^0 and mass conservation,

$$\sup_{0<t<T}\int_{\mathbb{R}^d}(1+|x|^2)^{\alpha/2}\xi_k\rho_\sigma(t)\mathrm{d}x \le C + C\int_0^T\int_{\mathbb{R}^d}\rho_\sigma(-\Delta)^{-s/2}|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)|\mathrm{d}x\mathrm{d}t.$$

Next, we apply the Hardy–Littlewood–Sobolev inequality (see Lemma 3.22) and the Hölder inequality and use the fact that $\rho_\sigma(t)$ is bounded in any $L^p(\mathbb{R}^d)$:

$$\begin{aligned}
\sup_{0<t<T}&\int_{\mathbb{R}^d}(1+|x|^2)^{\alpha/2}\rho_\sigma(x,t)\xi_k(x)\mathrm{d}x \\
&\le C + C\int_0^T\|\rho_\sigma\|_{2d/(d+2s)}\|(-\Delta)^{-s/2}[\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)]\|_{2d/(d-2s)}\mathrm{d}t \\
&\le C + \int_0^T\|\rho_\sigma\|_{2d/(d+2s)}\|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)\|_2\mathrm{d}t \\
&\le C(\eta) + \eta\int_0^T\|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)\|_2^2\mathrm{d}t
\end{aligned}$$

for all $\eta > 0$. Since $\xi_k(x) \le \xi_{k+1}(x)$ for $x \in \mathbb{R}^d$, $k > 1$, and $\xi_k \to 1$ a.e. in $\mathbb{R}^d$ as $k \to \infty$, we deduce from monotone convergence that in the limit $k \to \infty$,

$$\sup_{0<t<T}\int_{\mathbb{R}^d}(1+|x|^2)^{\alpha/2}\rho_\sigma(x,t)\mathrm{d}x \le C(\eta) + \eta\int_0^T\|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)\|_2^2\mathrm{d}t$$

for all $\eta > 0$. This proves that

$$\int_{\mathbb{R}^d}\rho_\sigma^\theta\mathrm{d}x \le C(\eta) + \eta\int_0^T\|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)\|_2^2\mathrm{d}t.$$

Step 3: A priori estimate. Inserting the previous estimate into (3.19) leads to

$$\sup_{0<t<T}\int_{\mathbb{R}^d}|h_\sigma^\delta(\rho_\sigma(x,t))|\mathrm{d}x \le C(\eta) + \eta\int_0^T\|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)\|_2^2\mathrm{d}t.$$

We integrate (3.18) in time and use the previous estimate:

$$\begin{aligned}
4\sigma&\int_0^T\int_{\mathbb{R}^d}f_\sigma'(\rho_\sigma)\frac{\rho_\sigma}{\rho_\sigma+\delta}|\nabla\rho_\sigma^{1/2}|^2\mathrm{d}x\mathrm{d}t \\
&\quad + \frac{c_{d,1-s}}{2}\int_0^T\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{(f_\sigma^\delta(\rho_\sigma(x)) - f_\sigma^\delta(\rho_\sigma(y)))(f_\sigma(\rho_\sigma(x)) - f_\sigma(\rho_\sigma(y))}{|x-y|^{d+2(1-s)}}\mathrm{d}x\mathrm{d}y\mathrm{d}t \\
&\le \int_{\mathbb{R}^d}|h_\sigma^\delta(\rho_\sigma(T))|\mathrm{d}x + \int_{\mathbb{R}^d}|h_\sigma^\delta(\rho_\sigma^0)|\mathrm{d}x \le C(\eta) + \eta\int_0^T\|\nabla(-\Delta)^{-s/2}f_\sigma(\rho_\sigma)\|_2^2\mathrm{d}t.
\end{aligned}$$

We wish to pass to the limit $\delta \to 0$ in the previous inequality. We deduce from dominated convergence that $f_\sigma^\delta(\rho_\sigma) \to f_\sigma(\rho_\sigma)$ a.e. in $\mathbb{R}^d \times [0,\infty)$. The integrand of the second term on the left-hand side is nonnegative, and we obtain from Fatou's lemma that

$$4\sigma \int_0^T \int_{\mathbb{R}^d} f_\sigma'(\rho_\sigma)|\nabla \rho_\sigma^{1/2}|^2 \mathrm{d}x\mathrm{d}t + \frac{c_{d,1-s}}{2} \int_0^T \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{(f_\sigma(\rho_\sigma(x)) - f_\sigma(\rho_\sigma(y)))^2}{|x-y|^{d+2(1-s)}} \mathrm{d}x\mathrm{d}y\mathrm{d}t \tag{3.20}$$
$$\leq C(\eta) + \eta \int_0^T \|\nabla(-\Delta)^{-s/2} f_\sigma(\rho_\sigma)\|_2^2 \mathrm{d}t.$$

By the integral representation of the fractional Laplacian,

$$\frac{c_{d,1-s}}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{(f_\sigma(\rho_\sigma(x)) - f_\sigma(\rho_\sigma(y)))^2}{|x-y|^{d+2(1-s)}} \mathrm{d}x\mathrm{d}y = \|\nabla(-\Delta)^{-s/2} f_\sigma(\rho_\sigma)\|_2^2,$$

the last term in (3.20) can be absorbed for sufficiently small $\eta > 0$ by the second term on the left-hand side. This leads to the estimate

$$4\sigma \int_0^T \int_{\mathbb{R}^d} f_\sigma'(\rho_\sigma)|\nabla \rho_\sigma^{1/2}|^2 \mathrm{d}x\mathrm{d}t + \int_0^T \int_{\mathbb{R}^d} |\nabla(-\Delta)^{-s/2} f_\sigma(\rho_\sigma)|^2 \mathrm{d}x\mathrm{d}t \leq C.$$

Thus, we can pass to the limit $\delta \to 0$ in (3.18) giving the desired entropy balance. Finally, bound (3.17) follows from the definition of the $H^{1-s}(\mathbb{R}^d)$ norm and the facts that $f_\sigma(\rho_\sigma) \in L^2(\mathbb{R}^d)$ since f_σ is locally Lipschitz continuous, $f_\sigma(0) = 0$, and ρ_σ is bounded both in $L^\infty(\mathbb{R}^d)$ and $L^2(\mathbb{R}^d)$ independently of σ. □

Lemma 3.9 (Moment estimate)**.** *It holds that*

$$\sup_{0<t<T} \int_{\mathbb{R}^d} \rho_\sigma(x,t)|x|^{2d/(d-2s)} \mathrm{d}x \leq C,$$

where $C > 0$ depends on T and the $L^1(\mathbb{R}^d)$ norms of ρ^0 and $|\cdot|^{2d/(d-2s)}\rho^0$.

Proof. For the following computations, we would need to use cut-off functions to make the calculations rigorous. We leave the details to the reader, as we wish to simplify the presentation. Let $m = 2d/(d-2s)$. Since $|\cdot|^m \rho^0 \in L^1(\mathbb{R}^d)$ by assumption, we can compute

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \rho_\sigma(t) \frac{|x|^m}{m} \mathrm{d}x = \sigma(m-2+d) \int_{\mathbb{R}^d} |x|^{m-2} \rho_\sigma \mathrm{d}x - \int_{\mathbb{R}^d} \rho_\sigma |x|^{m-2} x \cdot \nabla(-\Delta)^{-s} f_\sigma(\rho_\sigma) \mathrm{d}x$$
$$\leq C\||\cdot|^{m-2}\rho_\sigma\|_1 + \||\cdot|^{m-1}\rho_\sigma\|_{2d/(d+2s)} \|\nabla(-\Delta)^{-s} f_\sigma(\rho_\sigma)\|_{2d/(d-2s)}. \tag{3.21}$$

By Young's inequality and mass conservation, we have

$$\||\cdot|^{m-2}\rho_\sigma\|_1 \leq C \int_{\mathbb{R}^d} (1+|x|^m)\rho_\sigma \mathrm{d}x \leq C + C \int_{\mathbb{R}^d} |x|^m \rho_\sigma \mathrm{d}x.$$

It follows from (3.17) that $\nabla(-\Delta)^{-s} f_\sigma(\rho_\sigma)$ is bounded in $L^2(0,T;H^s(\mathbb{R}^d))$. In particular, because of the Sobolev embedding $H^s(\mathbb{R}^d) \hookrightarrow L^m(\mathbb{R}^d)$,

$$\|\nabla(-\Delta)^{-s} f_\sigma(\rho_\sigma)\|_{L^2(0,T;L^m(\mathbb{R}^d))} \leq C.$$

Furthermore, using $\rho_\sigma \in L^\infty(0,\infty;L^\infty(\mathbb{R}^d))$, Young's inequality, and the property $2d/(d+2s) \geq 1$ (recall that $d \geq 2$)

$$\||\cdot|^{m-1}\rho_\sigma\|_{2d/(d+2s)}^{2d/(d+2s)} = \int_{\mathbb{R}^d} \rho_\sigma^{2d/(d+2s)}|x|^{2d(m-1)/(d+2s)}\mathrm{d}x \\ \leq C + C\int_{\mathbb{R}^d} \rho_\sigma |x|^{2d(m-1)/(d+2s)}\mathrm{d}x.$$

Thus, we infer from (3.21) and the identity $2d(m-1)/(d+2s) = m$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} \rho_\sigma(t)\frac{|x|^m}{m}\mathrm{d}x \leq C + C\int_{\mathbb{R}^d}\rho_\sigma(t)|x|^m\mathrm{d}x,$$

and Gronwall's lemma concludes the proof. □

3.2.3 Higher-order estimate

We need some estimates in higher-order Sobolev spaces.

Proposition 3.10 (Higher-order regularity). *Let $T > 0$, $1 < p < \infty$ and $2 \leq q < \infty$. Then there exists $C > 0$, depending on T, σ, p, and q, such that*

$$\|\rho_\sigma\|_{L^p(0,T;W^{3,p}(\mathbb{R}^d))} + \|\partial_t\rho_\sigma\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + \|\rho_\sigma\|_{C^0([0,T];W^{2,q}(\mathbb{R}^d))} \leq C.$$

Proof. Step 1: Case $s > 1/2$. If $s > 1/2$ then $w := \rho_\sigma\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)$ does not involve any derivative of ρ_σ. Thus $w \in L^p(0,T;L^p(\mathbb{R}^d))$ for $p < \infty$ and Lemma 3.26 in Appendix 3.C implies that $\rho_\sigma \in L^p(0,T;W^{1,p}(\mathbb{R}^d))$. Iterating the argument leads to the conclusion. Thus, in the following, we can assume that $0 < s \leq 1/2$.

Step 2: Estimate of $\operatorname{div} w$ *in* $L^p(0,T;W^{-1,p}(\mathbb{R}^d))$. We claim that w can be estimated in $L^p(0,T;L^p(\mathbb{R}^d))$ for any $p < \infty$. Then, by Lemma 3.26, $\nabla\rho_\sigma \in L^p(0,T;L^p(\mathbb{R}^d))$. We use the L^∞ bound for ρ_σ, the fractional Gagliardo–Nirenberg inequality (Lemma 3.24), and Young's inequality to find that

$$\|w\|_p \leq C\|\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_p \leq C\|f_\sigma(\rho_\sigma)\|_p^{2s}\|\nabla f_\sigma(\rho_\sigma)\|_p^{1-2s} \leq C(\eta) + \eta\|\nabla\rho_\sigma\|_p,$$

where $\eta > 0$ is arbitrary. By estimate (3.72) in Lemma 3.26,

$$\|\rho_\sigma\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_p = \|w\|_p \leq C(\eta) + \eta\big(\|\rho_\sigma\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_p + T^{1/p}\|\nabla\rho^0\|_p\big).$$

Choosing $\eta > 0$ sufficiently small shows the claim.

Step 3: Estimate of $\operatorname{div} w$ *in* $L^p(0,T;L^p(\mathbb{R}^d))$. We use Hölder's inequality with $1/p = 2s/(d+p) + 1/q$ to obtain

$$\|\operatorname{div} w\|_p \leq \|\nabla\rho_\sigma\cdot\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_p + \|\rho_\sigma(-\Delta)^{1-s}f_\sigma(\rho_\sigma)\|_p \\ \leq \|\nabla\rho_\sigma\|_{(d+p)/(2s)}\|\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_q + C\|(-\Delta)^{1-s}f_\sigma(\rho_\sigma)\|_p.$$

By the fractional Gagliardo–Nirenberg inequality (Lemma 3.25 with $\theta = 1 + d/p - d/q - 2s$ and Lemma 3.24 with s replaced by $1-s$) and Young's inequality, it follows that

$$\|\operatorname{div} w\|_p \leq C\|\nabla\rho_\sigma\|_{(d+p)/(2s)}\|f_\sigma(\rho_\sigma)\|_p^{1-\theta}\|\nabla f_\sigma(\rho_\sigma)\|_p^{\theta} + C\|f_\sigma(\rho_\sigma)\|_p^{s}\|\mathrm{D}^2 f_\sigma(\rho_\sigma)\|_p^{1-s}$$

$$\leq C\|\nabla\rho_\sigma\|_{(d+p)/(2s)}\|\nabla\rho_\sigma\|_p^\theta + C\|f_\sigma'(\rho_\sigma)\mathrm{D}^2\rho_\sigma + f_\sigma''(\rho_\sigma)\nabla\rho_\sigma\otimes\nabla\rho_\sigma\|_p^{1-s}$$
$$\leq C(\eta) + C\|\nabla\rho_\sigma\|_{(d+p)/(2s)}^{1/(1-\theta)} + C\|\nabla\rho_\sigma\|_p + C\|\nabla\rho_\sigma\|_{2p}^2 + \eta\|\mathrm{D}^2\rho_\sigma\|_p,$$

where $\eta > 0$ is arbitrary. Taking the $L^p(0,T)$ norm of the previous inequality and observing that $p/(1-\theta) = (d+p)/(2s)$ (because of $\theta = d(1/p - 1/q) + 1 - 2s$), it follows that

$$\begin{aligned}\|\operatorname{div} w\|_{L^p(0,T;L^p(\mathbb{R}^d))} &\leq C + C\|\nabla\rho_\sigma\|_{L^{(d+p)/(2s)}(0,T;L^{(d+p)/(2s)}(\mathbb{R}^d))}^{1/(1-\theta)} + C\|\nabla\rho_\sigma\|_{L^p(0,T;L^p(\mathbb{R}^d))} \\ &\quad + C\|\nabla\rho_\sigma\|_{L^{2p}(0,T;L^{2p}(\mathbb{R}^d))}^2 + \eta\|\mathrm{D}^2\rho_\sigma\|_{L^p(0,T;L^p(\mathbb{R}^d))}.\end{aligned}$$

Lemma 3.26 and Step 2 ($\nabla\rho_\sigma \in L^p(0,T;L^p(\mathbb{R}^d))$) show that

$$\|\partial_t\rho_\sigma\|_{L^p(0,T;L^p(\mathbb{R}^d))} + (1 - C\eta)\|\mathrm{D}^2\rho_\sigma\|_{L^p(0,T;L^p(\mathbb{R}^d))} \leq C.$$

Choosing $\eta > 0$ sufficiently small, this yields $\partial_t\rho_\sigma \in L^p(0,T;L^p(\mathbb{R}^d))$ and $\rho_\sigma \in L^p(0,T; W^{2,p}(\mathbb{R}^d))$. We deduce from Lemma 3.20, applied to $\nabla\rho_\sigma$, that $\nabla\rho_\sigma \in L^\infty(0,T;L^q(\mathbb{R}^d))$ for any $2 \leq q < \infty$. (At this point, we need the restriction $q \geq 2$.)
Step 4: Higher-order regularity. To improve the regularity of ρ_σ, we differentiate (3.7) in space. Recall that $\partial_i = \partial/\partial x_i$, $i = 1,\ldots,d$. Then

$$\begin{aligned}\partial_t\partial_i\rho_\sigma - \sigma\Delta\partial_i\rho_\sigma &= \sum_{j=1}^d \partial_i\partial_j\big(\rho_\sigma\partial_j(-\Delta)^{-s}f_\sigma(\rho_\sigma)\big) = \sum_{j=1}^d \big(\partial_{ij}^2\rho_\sigma\partial_j(-\Delta)^s f_\sigma(\rho_\sigma) \\ &+ \partial_i\rho_\sigma\partial_{jj}^2(-\Delta)^{-s}f_\sigma(\rho_\sigma) + \partial_j\rho_\sigma\partial_{ij}^2(-\Delta)^{-s}f_\sigma(\rho_\sigma) + \rho_\sigma\partial_{ijj}^3(-\Delta)^{-s}f_\sigma(\rho_\sigma)\big).\end{aligned} \tag{3.22}$$

We estimate the right-hand side term by term. Let $0 < s \leq 1/2$. First, by Hölder's inequality with $1/p = 1/q + 1/r$, $1 < p < q < \infty$, $\max\{2,p\} < r < \infty$ and the fractional Gagliardo–Nirenberg inequality (Lemma 3.24),

$$\begin{aligned}\|\partial_{ij}^2\rho_\sigma\partial_j(-\Delta)^s f_\sigma(\rho_\sigma)\|_{L^p(0,T;L^p(\mathbb{R}^d))}^p &\leq \int_0^T \|\partial_{ij}^2\rho_\sigma\|_q^p\|\partial_j(-\Delta)^s f_\sigma(\rho_\sigma)\|_r^p\mathrm{d}t \\ &\leq C\int_0^T \|\partial_{ij}^2\rho_\sigma\|_q^p\|f_\sigma(\rho_\sigma)\|_r^{(1-2s)p}\|\nabla f_\sigma(\rho_\sigma)\|_r^{2sp}\mathrm{d}t \\ &\leq C\|f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;L^r(\mathbb{R}^d))}^{(1-2s)p}\|\nabla f_\sigma(\rho_\sigma)\|_{L^\infty(0,L^r(\mathbb{R}^d))}^{2sp}\int_0^T \|\partial_{ij}^2\rho_\sigma\|_q^p\mathrm{d}t \leq C.\end{aligned}$$

The second and third term on the right-hand side of (3.22) can be treated in a similar way, observing that $\partial_{ij}^2(-\Delta)^{-s} = \partial_j(-\Delta)^{-s}\partial_i$. The last term is estimated according to

$$\begin{aligned}\|\rho_\sigma\partial_{ijj}^3(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_p &\leq C\|\partial_{ijj}^3(-\Delta)^{-s}f_\sigma(\rho_\sigma)\|_p \leq C\|\partial_{jj}^2 f_\sigma(\rho_\sigma)\|_p^{2s}\|\nabla\partial_{jj}^2 f_\sigma(\rho_\sigma)\|_p^{1-2s} \\ &\leq C(\eta)\|\partial_{jj}^2 f_\sigma(\rho_\sigma)\|_p + \eta\|\nabla\partial_{jj}^2 f_\sigma(\rho_\sigma)\|_p,\end{aligned}$$

and the last expression can be absorbed by the corresponding estimate of $\Delta\partial_i\rho_\sigma$ from the left-hand side of (3.22). Then we deduce from Lemma 3.26 that $\partial_t\partial_i\rho_\sigma$, $\partial_{ijj}^3\rho_\sigma \in L^p(0,T; L^p(\mathbb{R}^d))$ for all $p > 1$ and Lemma 3.20, applied to $\partial_{ij}^2\rho_\sigma$, yields $\partial_{ij}^2\rho_\sigma \in C^0([0,T];L^q(\mathbb{R}^d))$ for all $q \geq 2$.
Next, if $1/2 < s < 1$, we use the second inequality in Lemma 3.24 and argue similarly as before. This finishes the proof. □

Lemma 3.11. *Under the assumptions of Proposition 3.10, for every $q \geq 2$, there exists a constant $C = C(q) > 0$, depending on σ, such that*

$$\|\rho_\sigma\|_{C^0([0,T];W^{2,1}(\mathbb{R}^d)\cap W^{3,q}(\mathbb{R}^d))} \leq C.$$

The embedding $W^{3,q}(\mathbb{R}^d) \hookrightarrow W^{2,\infty}(\mathbb{R}^d)$ for $q > d$ yields a bound for ρ_σ in $C^0([0,T];W^{2,\infty}(\mathbb{R}^d))$.

Proof. We first prove the bound in $C^0([0,T];W^{3,q}(\mathbb{R}^d))$. By differentiating (3.7) twice in space, estimating similarly as in Step 4 of the previous proof, and using the regularity results of Proposition 3.10, we can show that ρ_σ is bounded in $L^\infty(0,T;W^{3,q}(\mathbb{R}^d))$ for any $q \geq 2$.
It remains to show the $C^0([0,T];W^{2,1}(\mathbb{R}^d))$ bound for ρ_σ. In view of mass conservation and Gagliardo–Nirenberg–Sobolev's inequality, it suffices to show a bound for $\mathrm{D}^2\rho_\sigma$ in $L^\infty(0,T;L^1(\mathbb{R}^d))$. To this end, we define the weights $\gamma_n = (1+|x|^2)^{n/2}$ for $n \geq 0$ and test equation (3.7) for ρ_σ with $v_n := \gamma_n\rho_\sigma$. Then

$$\begin{aligned}&\partial_t v_n - \sigma\Delta v_n = \operatorname{div}\big(v_n\nabla\mathcal{K} * f_\sigma(\rho_\sigma)\big) + I_n, \quad v_n(0) = \gamma_n\rho_\sigma^0 \quad \text{in } \mathbb{R}^d,\\ &\text{where } I_n = -2\sigma\nabla\gamma_n\cdot\nabla\rho_\sigma - \sigma\rho_\sigma\Delta\gamma_n - \rho_\sigma\nabla\gamma_n\cdot\nabla\mathcal{K} * f_\sigma(\rho_\sigma).\end{aligned}$$

Arguing as in Step 4 of the previous proof, we can find a bound in $L^\infty(0,T;W^{2,p}(\mathbb{R}^d))$ for v_n. Indeed, we can proceed by induction over n, since the additional terms in I_n can be controlled by Sobolev norms of $v_0,\ldots,v_{n-1}$. The definition of ρ_σ^0 implies that $\gamma_n\rho_\sigma^0$, $\gamma_n\nabla\rho_\sigma^0 \in L^\infty(\mathbb{R}^d)\cap L^1(\mathbb{R}^d)$ for every $n \geq 0$. Then choosing $n > d$ yields, for $0 \leq t \leq T$, that

$$\|\gamma_n\mathrm{D}^2\rho_\sigma\|_p \leq \|\mathrm{D}^2(\gamma_n\rho_\sigma)\|_p + 2\|\nabla\gamma_n\cdot\nabla\rho_\sigma\|_p + \|\rho_\sigma\mathrm{D}^2\gamma_n\|_p \leq C(T).$$

We conclude from $\gamma_n^{-1} \in L^\infty(\mathbb{R}^d)\cap L^1(\mathbb{R}^d)$ that

$$\|\mathrm{D}^2\rho_\sigma\|_1 \leq \|\gamma_n^{-1}\|_{p/(p-1)}\|\gamma_n\mathrm{D}^2\rho_\sigma\|_p \leq C(T).$$

This proves the desired bound. □

3.2.4 Existence of solutions to (3.7)

We show that the regularized equation (3.7) possesses a unique strong solution ρ_σ.

Step 1: Existence for an approximated system. Let $T > 0$ arbitrary, define the spaces

$$\begin{aligned}&X_T := L^2(0,T;H^1(\mathbb{R}^d))\cap H^1(0,T;H^{-1}(\mathbb{R}^d)) \hookrightarrow Y_T := C^0([0,T];L^2(\mathbb{R}^d)),\\ &Y_{T,R} := \{u \in Y_T : \|u-\rho_\sigma^0\|_{L^\infty(0,T;L^2(\mathbb{R}^d))} \leq R\},\end{aligned}$$

and consider the mapping $S : v \in Y_T \mapsto u \in Y_T$,

$$\begin{aligned}&\partial_t u - \sigma\Delta u = \operatorname{div}(u\nabla\mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v)) \quad \text{in } \mathbb{R}^d\times(0,T),\\ &u(0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d,\end{aligned} \tag{3.23}$$

where $\mathcal{K}_s^{(\delta)} : \mathbb{R}^d \to \mathbb{R}_+$ is a regularized version of $\mathcal{K}_s$, defined by

$$\mathcal{K}_s^{(\delta)} = \widetilde{\mathcal{K}}_{s/2}^{(\delta)} * \widetilde{\mathcal{K}}_{s/2}^{(\delta)},$$

$$\widetilde{\mathcal{K}}_{s/2}^{(\delta)}(x) = c_{d,-s/2} \begin{cases} \delta^{s-d} + (s-d)\delta^{s-d-1}(|x| - \delta) & \text{for } |x| < \delta, \\ |x|^{s-d} & \text{for } \delta \le |x| \le \delta^{-1}, \\ [\delta^{d-s} + (s-d)\delta^{d+1-s}(|x| - \delta^{-1})]_+ & \text{for } |x| > \delta^{-1}, \end{cases}$$

and $f_\sigma^{(\eta)}$ is given by

$$f_\sigma^{(\eta)}(\rho) = \int_0^{|\rho|} f'_\sigma(u) \min(1, u\eta^{-1}) \mathrm{d}u + \frac{\eta}{2}\rho^2, \quad \rho \in \mathbb{R}.$$

The regularization with parameter η is needed for the entropy estimates.

Lemma 3.12. *For any $0 < s < 1$ and a.e. $x \in \mathbb{R}^d$, the function $\delta \mapsto \widetilde{\mathcal{K}}_{s/2}^{(\delta)}(x)$ is nonincreasing for $\delta \in (0,1)$.*

Proof. Let $r_\delta^* = (d-s+1)/((d-s)\delta)$. We can write $\widetilde{\mathcal{K}}_{s/2}^{(\delta)}(x) = c_{d,-s/2}\Phi_\delta(|x|)$ with

$$\Phi_\delta(r) = \begin{cases} \delta^{s-d} + (s-d)\delta^{s-d-1}(r-\delta) & \text{for } r < \delta, \\ r^{s-d} & \text{for } \delta \le r \le \delta^{-1}, \\ \delta^{d-s} + (s-d)\delta^{d+1-s}(r - \delta^{-1}) & \text{for } \delta^{-1} < r < r_\delta^*, \\ 0 & \text{for } r \ge r_\delta^*. \end{cases}$$

Then $\Phi_\delta \in C^0([0,\infty)) \cap C^1(0, r_\delta^*)$, and its derivative equals

$$\Phi'_\delta(r) = \begin{cases} -(d-s)\max\{r,\delta\}^{s-d-1} & 0 \le r \le 1, \\ -(d-s)\min\{r,\delta^{-1}\}^{s-d-1} & 1 \le r < r_\delta^*. \end{cases}$$

We show that $\Phi_\delta(r)$ is nonincreasing in $\delta \in (0,1)$ for $r \ge 1$. We have for $1 \le r < r_\delta^*$,

$$\Phi_\delta(r) = \Phi_\delta(1) + \int_1^r \Phi'_\delta(u)\mathrm{d}u = 1 - (d-s)\int_1^r \min\{u, \delta^{-1}\}^{s-d-1}\mathrm{d}u.$$

Furthermore, we have $\Phi_\delta(r_\delta^*) = 0$, while $\min(u, \delta^{-1}) = \delta^{-1} > 0$ for $u > r_\delta^*$, so it holds that

$$\Phi_\delta(r) = \left(1 - (d-s)\int_1^r \min\{u,\delta^{-1}\}^{s-d-1}\mathrm{d}u\right)_+ \quad \text{for } r \ge 1.$$

At this point, the above representation formula together with elementary monotonicity considerations show that $\Phi_\delta(r)$ is nonincreasing in $\delta \in (0,1)$ for $r \ge 1$. It remains to show that $\Phi_\delta(r)$ is nonincreasing in $\delta \in (0,1)$ for $0 \le r < 1$. It holds that

$$\Phi_\delta(r) = \Phi_\delta(1) - \int_r^1 \Phi'_\delta(u)\mathrm{d}u = 1 + (d-s)\int_r^1 \max\{u,\delta\}^{s-d-1}\mathrm{d}u \quad \text{for } 0 \le r < 1.$$

Once again, we conclude from the above representation formula together with elementary monotonicity considerations that $\Phi_\delta(r)$ is nonincreasing in $\delta \in (0,1)$ for $0 \le r < 1$. This finishes the proof. ☐

We derive some estimates for $f_\sigma^{(\eta)}$. First, we have $0 \le f_\sigma^{(\eta)}(\rho) \le C_\eta \rho^2$ for $\rho \in \mathbb{R}$, since

$$f_\sigma^{(\eta)}(\rho) \le \Big(\eta + \eta^{-1} \max_{[0,\eta]} f_\sigma'\Big) \frac{\rho^2}{2} \quad \text{for } |\rho| \le \eta,$$
$$f_\sigma^{(\eta)}(\rho) \le f_\sigma(|\rho|) + \frac{\eta}{2}\rho^2 \le \Big(\|f_\sigma\|_\infty \eta^{-2} + \frac{\eta}{2}\Big)\rho^2 \quad \text{for } |\rho| > \eta.$$

Furthermore,

$$|\mathrm{D} f_\sigma^{(\eta)}(\rho)| = \left| \frac{\rho}{|\rho|} f_\sigma'(|\rho|) \min(1, |\rho|\eta^{-1}) + \eta\rho \right| \le (\eta + \eta^{-1}\|f_\sigma'\|_\infty)|\rho|,$$

which implies that $|\mathrm{D} f_\sigma^{(\eta)}(\rho)| \le C_\eta |\rho|$ for $\rho \in \mathbb{R}$. This shows that there exists $C(\eta) > 0$ such that for any ρ_1, $\rho_2 \in \mathbb{R}$,

$$|f_\sigma^{(\eta)}(\rho_1) - f_\sigma^{(\eta)}(\rho_2)| \le C(\eta)(|\rho_1| + |\rho_2|)|\rho_1 - \rho_2|.$$

It follows that $f_\sigma^{(\eta)}(v) \in L^\infty(0,T; L^1(\mathbb{R}^d))$ for $v \in Y_T$.
Since $\nabla \mathcal{K}_s^{(\delta)} \in L^\infty(\mathbb{R}^d)$, a standard argument shows that (3.23) has a unique solution $u \in X_T \hookrightarrow Y_T$. Therefore, the mapping S is well-defined. Additionally, the nonnegativity of u follows immediately after by testing (3.23) with $\min(0,u)$.
We show now that S is a contraction on $Y_{T,R}$ for sufficiently small $T > 0$. We start with a preparation. By testing (3.23) with u and taking into account the L^∞ bound for $\nabla \mathcal{K}_s^{(\delta)}$, we deduce from Young's inequality for products and convolutions that

$$\int_{\mathbb{R}^d} u(t)^2 \mathrm{d}x + \frac{\sigma}{2} \int_0^t \int_{\mathbb{R}^d} |\nabla u|^2 \mathrm{d}x\mathrm{d}\tau \le \int_{\mathbb{R}^d} |\rho_\sigma^0|^2 \mathrm{d}x + C(\delta,\eta,\sigma) \int_0^t \|u\|_2^2 \|v\|_2^4 \mathrm{d}\tau,$$

since $\|f_\sigma^{(\eta)}(v)\|_1 \le C_\eta \|v\|_2^2$ for $v \in Y_T$. Then, if $v \in Y_{T,R}$, we infer from Gronwall's lemma that

$$\int_{\mathbb{R}^d} u(t)^2 \mathrm{d}x + \sigma \int_0^t \int_{\mathbb{R}^d} |\nabla u|^2 \mathrm{d}x\mathrm{d}\tau \le e^{C(\sigma,\delta,\eta)R^4 t} \int_{\mathbb{R}^d} |\rho_\sigma^0|^2 \mathrm{d}x \quad \text{for } 0 \le t \le T. \tag{3.24}$$

Let $v_i \in Y_{T,R}$ and set $u_i = S(v_i)$, $i = 1,2$. We compute

$$\begin{aligned}
&\|u_1 \nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v_1) - u_2 \nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v_2)\|_2 \\
&\quad \le \|(u_1 - u_2) \nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v_1)\|_2 + \|u_2 \nabla \mathcal{K}_s^{(\delta)} * (f_\sigma^{(\eta)}(v_1) - f_\sigma^{(\eta)}(v_2))\|_2 \\
&\quad \le \|u_1 - u_2\|_2 \|\nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v_1)\|_\infty + \|u_2\|_2 \|\nabla \mathcal{K}_s^{(\delta)} * (f_\sigma^{(\eta)}(v_1) - f_\sigma^{(\eta)}(v_2))\|_\infty \\
&\quad \le \|u_1 - u_2\|_2 \|\nabla \mathcal{K}_s^{(\delta)}\|_\infty \|f_\sigma^{(\eta)}(v_1)\|_1 + \|u_2\|_2 \|\nabla \mathcal{K}_s^{(\delta)}\|_\infty \|f_\sigma^{(\eta)}(v_1) - f_\sigma^{(\eta)}(v_2)\|_1 \\
&\quad \le C(\delta,\eta)\big(\|u_1 - u_2\|_2 \|v_1\|_2^2 + \|u_2\|_2(\|v_1\|_2 + \|v_2\|_2)\|v_1 - v_2\|_2\big).
\end{aligned}$$

Therefore, using (3.24), for $v_1, v_2 \in Y_{T,R}$,

$$\|u_1 \nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v_1) - u_2 \nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(v_2)\|_2 \le C(\delta,\eta,R,T)(\|u_1 - u_2\|_2 + \|v_1 - v_2\|_2). \tag{3.25}$$

Next, we write (3.23) for (u_i, v_i) in place of (u, v), $i = 1, 2$, take the difference between the two equations, and test the resulting equation with $u_1 - u_2$:

$$\begin{aligned}
&\frac{1}{2}\|(u_1-u_2)(t)\|_2^2 + \sigma\int_0^t\int_{\mathbb{R}^d}|\nabla(u_1-u_2)|^2\mathrm{d}x\mathrm{d}\tau \\
&\quad = -\int_0^t\int_{\mathbb{R}^d}\nabla(u_1-u_2)\cdot(u_1\nabla\mathcal{K}_s^{(\delta)}*f_\sigma^{(\eta)}(v_1) - u_2\nabla\mathcal{K}_s^{(\delta)}*f_\sigma^{(\eta)}(v_2))\mathrm{d}x\mathrm{d}\tau \\
&\quad \le \frac{\sigma}{2}\int_0^t\int_{\mathbb{R}^d}|\nabla(u_1-u_2)|^2\mathrm{d}x\mathrm{d}\tau + \frac{1}{2\sigma}\int_0^t\|u_1\nabla\mathcal{K}_s^{(\delta)}*f_\sigma^{(\eta)}(v_1) - u_2\nabla\mathcal{K}_s^{(\delta)}*f_\sigma^{(\eta)}(v_2)\|_2^2\mathrm{d}\tau.
\end{aligned}$$

It follows from (3.25) that

$$\begin{aligned}
&\|(u_1-u_2)(t)\|_2^2 + \sigma\int_0^t\int_{\mathbb{R}^d}|\nabla(u_1-u_2)|^2\mathrm{d}x\mathrm{d}\tau \\
&\quad \le C(\delta,\eta,R,T,\sigma)\int_0^t(\|u_1-u_2\|_2^2 + \|v_1-v_2\|_2^2)\mathrm{d}\tau,
\end{aligned}$$

and we conclude from Gronwall's lemma that

$$\|(u_1-u_2)(t)\|_2^2 \le e^{C(\delta,\eta,R,T,\sigma)t}\int_0^T\|v_1-v_2\|_2^2\mathrm{d}\tau \quad \text{for } 0\le t\le T.$$

This inequality implies that S is a contraction in $Y_{T,R}$, provided that T is sufficiently small. Therefore, by Banach's theorem, S admits a unique fixed point $u \in Y_{T,R} \subset Y_T$ for $T > 0$ sufficiently small.

It remains to show that the local solution can be extended to a global one. To this end, we note that the function $u \in X_T$ satisfies (3.23) with $v = u$:

$$\begin{aligned}
&\partial_t u - \sigma\Delta u = \operatorname{div}(u\nabla\mathcal{K}_s^{(\delta)}*f_\sigma^{(\eta)}(u)) \quad \text{in } \mathbb{R}^d\times(0,T), \\
&u(\cdot,0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d.
\end{aligned} \tag{3.26}$$

Then, defining the truncated entropy density

$$h^{(\eta)}(\rho) = \int_0^\rho\int_0^u \mathrm{D}f_\sigma^{(\eta)}(v)v^{-1}\mathrm{d}v\mathrm{d}u, \quad \rho\ge 0,$$

and testing (3.26) with $\mathrm{D}h^{(\eta)}(u)$ yields, in view of the definition of $\mathcal{K}_s^{(\delta)}$, that

$$\begin{aligned}
&\int_{\mathbb{R}^d}h^{(\eta)}(u(t))\mathrm{d}x + \sigma\int_0^t\int_{\mathbb{R}^d}\mathrm{D}f_\sigma^{(\eta)}(u)u^{-1}|\nabla u|^2\mathrm{d}x\mathrm{d}\tau \\
&\qquad + \int_0^t\int_{\mathbb{R}^d}|\nabla\widetilde{\mathcal{K}}_{s/2}^{(\delta)}*f_\sigma^{(\eta)}(u)|^2\mathrm{d}x\mathrm{d}\tau = \int_{\mathbb{R}^d}h^{(\eta)}(\rho_\sigma^0)\mathrm{d}x
\end{aligned} \tag{3.27}$$

for $0 \le t \le T$. This inequality and the definitions of $f_\sigma^{(\eta)}$ and $h^{(\eta)}$ yield a (δ, T)-uniform bound for u in $L^2(0,T;H^1(\mathbb{R}^d))$, which in turn (together with (3.26)) implies a (δ, T)-uniform bound for u in X_T, and a fortiori in Y_T. This means that the solution u can be prolonged to the whole time interval $[0,\infty)$ and exists for all times.

Finally, we point out that, since $\nabla\mathcal{K}_s^{(\delta)} \in L^2(\mathbb{R}^d)$, then $\nabla\mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(u) \in L^\infty(0,T;L^2(\mathbb{R}^d))$ and so $u\nabla\mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(u) \in L^\infty(0,T;L^1(\mathbb{R}^d))$. This fact yields the conservation of mass for u, i.e. $\int_{\mathbb{R}^d} u(t)\mathrm{d}x = \int_{\mathbb{R}^d} \rho_\sigma^0 \mathrm{d}x$ for $t > 0$. Indeed, it is sufficient to test (3.26) with a cut-off $\psi_R \in C_0^1(\mathbb{R}^d)$ satisfying $\psi_R(x) = 1$ for $|x| < R$, $\psi_R(x) = 0$ for $|x| > 2R$, $|\nabla\psi_R(x)| \le CR^{-1}$ for $x \in \mathbb{R}^d$, and then to take the limit $R \to \infty$.

Step 2: Limit $\delta \to 0$. Let $u^{(\delta)}$ be the solution to (3.26). An adaption of the proof of [14, Lemma 1] shows that the embedding $H^1(\mathbb{R}^d) \cap L^1(\mathbb{R}^d;(1+|x|^2)^{\kappa/2}) \hookrightarrow L^2(\mathbb{R}^d)$ is compact. Thus, because of the δ-uniform bounds for $u^{(\delta)}$, the Aubin–Lions Lemma implies that (up to a subsequence) $u^{(\delta)} \to u$ strongly in $L^2(0,T;L^2(\mathbb{R}^d))$ for every $T > 0$. We wish now to study the convergence of the nonlinear and nonlocal terms in (3.26)–(3.27) as $\delta \to 0$.
It follows from (3.27) that (up to a subsequence)

$$\nabla\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) \rightharpoonup U \quad \text{weakly in } L^2(\mathbb{R}^d \times (0,T)) \quad \text{as } \delta \to 0. \tag{3.28}$$

In order to identify the limit U, we first notice that, by construction, $0 \le \widetilde{\mathcal{K}}_{s/2}^{(\delta)} \nearrow \mathcal{K}_{s/2}$ a.e. in $\mathbb{R}^d$. Furthermore, the Hardy–Littlewood–Sobolev inequality, the bound for $f_\sigma^{(\eta)}$, and then the Gagliardo-Nirenberg-Sobolev inequality yield that

$$\begin{aligned}\|\mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u)\|_{(d+2)/(d-s)} &\le C\|f_\sigma^{(\eta)}(u)\|_{(d+2)/(d+2s/d)} \le C(\eta)\|u\|^2_{(2d+4)/(d+2s/d)} \\ &\le C(\eta)\|u\|_2^{2(s+2)/(d+2)}\|\nabla u\|_2^{2(d-s)/(d+2)}.\end{aligned}$$

Therefore, since $u \in L^\infty(0,T;L^2(\mathbb{R}^d)) \cap L^2(0,T;H^1(\mathbb{R}^d))$,

$$\int_0^T \|\mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u)\|_{(d+2)/(d-s)}^{(d+2)/(d-s)}\mathrm{d}t \le C(\eta)\|u\|_{L^\infty(0,T;L^2(\mathbb{R}^d))}^{2(s+2)/(d-s)} \int_0^T \|\nabla u\|_2^2 \mathrm{d}t \le C(\eta,T),$$

meaning that $\mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u) \in L^{(d+2)/(d-s)}(\mathbb{R}^d \times (0,T))$. Taking into account that $f_\sigma^{(\eta)}(u) \ge 0$ and that $\delta \mapsto \widetilde{\mathcal{K}}_{s/2}^{(\delta)}(x) \in \mathbb{R}$ is nonincreasing (see Lemma 3.12), we deduce from monotone convergence that

$$\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u) \to \mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u) \quad \text{strongly in } L^{(d+2)/(d-s)}(\mathbb{R}^d \times (0,T)). \tag{3.29}$$

Furthermore, arguing as before and using the estimates for $\mathrm{D}f_\sigma^{(\eta)}$ leads to

$$\begin{aligned}\|\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * (f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u))\|_{(d+2)/(d-s)} &\le \|\widetilde{\mathcal{K}}_{s/2} * |f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u)|\|_{(d+2)/(d-s)} \\ &\le C\|f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u)\|_{(d+2)/(d+2s/d)} \\ &\le C(\eta)\big\||u| + |u^{(\delta)}|\big\|_{(2d+4)/(d+2s/d)}\|u - u^{(\delta)}\|_{(2d+4)/(d+2s/d)} \\ &\le C(\eta)\big(\|u\|_2^{(s+2)/(d+2)}\|\nabla u\|_2^{(d-s)/(d+2)} + \|u^{(\delta)}\|_2^{(s+2)/(d+2)}\|\nabla u^{(\delta)}\|_2^{(d-s)/(d+2)}\big) \\ &\quad \times \|u - u^{(\delta)}\|_2^{(s+2)/(d+2)}\|\nabla(u - u^{(\delta)})\|_2^{(d-s)/(d+2)}.\end{aligned}$$

Since $u^{(\delta)}$ is bounded in $L^\infty(0,T;L^2(\mathbb{R}^d)) \cap L^2(0,T;H^1(\mathbb{R}^d))$, it follows that (up to a subsequence) $\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * (f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u))$ converges weakly to some limit in $L^{(d+2)/(d-s)}(\mathbb{R}^d \times$

$(0,T))$. However, Hölder's inequality and the fact that $u^{(\delta)} \to u$ strongly in $L^p(0,T;L^2(\mathbb{R}^d))$ for every $2 \le p < \infty$, which follows from

$$\int_0^T \|u^{(\delta)} - u\|_2^p \mathrm{d}t \le \sup_{0<t<T} \|(u^{(\delta)} - u)(t)\|_2^{p-2} \int_0^T \|u^{(\delta)} - u\|_2^2 \mathrm{d}t \to 0 \quad \text{as } \delta \to 0,$$

imply that

$$\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * (f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u)) \to 0 \quad \text{strongly in } L^p(0,T;L^{(d+2)/(d-s)}(\mathbb{R}^d)),\ p < \frac{d+2}{d-s}.$$

We conclude that

$$\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * (f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u)) \rightharpoonup 0 \quad \text{weakly in } L^{(d+2)/(d-s)}(\mathbb{R}^d \times (0,T)). \tag{3.30}$$

We deduce from (3.29)–(3.30) that

$$\begin{aligned}
&\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) - \mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u) \\
&\quad = \big(\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u) - \mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u)\big) + \widetilde{\mathcal{K}}_{s/2}^{(\delta)} * (f_\sigma^{(\eta)}(u^{(\delta)}) - f_\sigma^{(\eta)}(u)) \\
&\quad \rightharpoonup 0 \quad \text{weakly in } L^{(d+2)/(d-s)}(\mathbb{R}^d \times (0,T)),
\end{aligned}$$

which, together with (3.28), implies that $U = \nabla\mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u)$, that is,

$$\nabla\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) \rightharpoonup \nabla\mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u) \quad \text{weakly in } L^2(\mathbb{R}^d \times (0,T)). \tag{3.31}$$

Let $\psi \in C_0^\infty(\mathbb{R}^d \times (0,T))$. Because of

$$\nabla\mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) = \widetilde{\mathcal{K}}_{s/2}^{(\delta)} * \big(\nabla\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)})\big),$$

we find that

$$\int_0^T \int_{\mathbb{R}^d} \psi \cdot \nabla\mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) \mathrm{d}x\mathrm{d}t = \int_0^T \int_{\mathbb{R}^d} \big(\nabla\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)})\big) \cdot \big(\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * \psi\big) \mathrm{d}x\mathrm{d}t.$$

Our goal is to show that $\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * \psi \to \mathcal{K}_{s/2} * \psi$ strongly in $L^2(\mathbb{R}^d \times (0,T))$ as $\delta \to 0$. We can assume without loss of generality that $\psi \ge 0$ a.e. in $\mathbb{R}^d \times (0,T)$. Indeed, for general functions ψ, we may write $\psi = \psi_+ + \psi_-$, where $\psi_+ = \max\{0,\psi\}$ and $\psi_- = \min\{0,\psi\}$, and we have $\widetilde{\mathcal{K}}_{s/2}^{(\delta)} * \psi = \widetilde{\mathcal{K}}_{s/2}^{(\delta)} * \psi_+ - \widetilde{\mathcal{K}}_{s/2}^{(\delta)} * (-\psi_-)$. Once again, since $\widetilde{\mathcal{K}}_{s/2}^{(\delta)} \nearrow \widetilde{\mathcal{K}}_{s/2}$ a.e. in $\mathbb{R}^d$, it is sufficient to show that $\mathcal{K}_{s/2} * \psi \in L^2(\mathbb{R}^d \times (0,T))$. The Hardy–Littlewood–Sobolev inequality (see Appendix 3.B) yields

$$\int_0^T \|\mathcal{K}_{s/2} * \psi\|_2^2 \mathrm{d}t \le C \int_0^T \|\psi\|_{2d/(d+2s)}^2 \mathrm{d}t.$$

It follows from (3.31), the previous argument, and the fact that $\mathcal{K}_s * u = (-\Delta)^{-s} u = \mathcal{K}_{s/2} * \mathcal{K}_{s/2} * u$ that

$$\int_0^T \int_{\mathbb{R}^d} \psi \cdot \nabla\mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) \mathrm{d}x\mathrm{d}t \to \int_0^T \int_{\mathbb{R}^d} \big(\nabla\mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u)\big) \cdot (\mathcal{K}_{s/2} * \psi) \mathrm{d}x\mathrm{d}t$$

$$= \int_0^T \int_{\mathbb{R}^d} \psi \cdot \nabla \mathcal{K}_s * f_\sigma^{(\eta)}(u) \mathrm{d}x \mathrm{d}t$$

for every $\psi \in L^2(0,T; L^{2d/(d+2s)}(\mathbb{R}^d))$, which means that

$$\nabla \mathcal{K}_s^{(\delta)} * f_\sigma^{(\eta)}(u^{(\delta)}) \rightharpoonup \nabla \mathcal{K}_s * f_\sigma^{(\eta)}(u) \quad \text{weakly in } L^2(0,T; L^{2d/(d-2s)}(\mathbb{R}^d)). \tag{3.32}$$

Since $u^{(\delta)} \to u$ strongly in $L^2(0,T;L^2(\mathbb{R}^d))$ and $(u^{(\delta)})$ is bounded in $L^\infty(0,T;L^1(\mathbb{R}^d))$ (via mass conservation), it also holds that $u^{(\delta)} \to u$ strongly in $L^2(0,T;L^{2d/(d+2s)}(\mathbb{R}^d))$. Therefore, the convergence (3.32) is sufficient to pass to the limit $\delta \to 0$ in (3.26).
Step 3: Limit $\eta \to 0$ and conclusion. The limit $\delta \to 0$ in (3.26) shows that the limit u solves

$$\begin{aligned} &\partial_t u - \sigma \Delta u = \operatorname{div}(u \nabla \mathcal{K}_s * f_\sigma^{(\eta)}(u)) \quad \text{in } \mathbb{R}^d \times (0,T), \\ &u(\cdot, 0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d. \end{aligned} \tag{3.33}$$

Fatou's Lemma and the weakly lower semicontinuity of the L^2 norm allow us to infer from (3.27) that for $t > 0$,

$$\begin{aligned} \int_{\mathbb{R}^d} h^{(\eta)}(u(t)) \mathrm{d}x + \sigma \int_0^t \int_{\mathbb{R}^d} \mathrm{D} f_\sigma^{(\eta)}(u) u^{-1} |\nabla u|^2 \mathrm{d}x \mathrm{d}\tau \\ + \int_0^t \int_{\mathbb{R}^d} |\nabla \mathcal{K}_{s/2} * f_\sigma^{(\eta)}(u)|^2 \mathrm{d}x \mathrm{d}\tau \le \int_{\mathbb{R}^d} h^{(\eta)}(\rho_\sigma^0) \mathrm{d}x. \end{aligned} \tag{3.34}$$

At this point, all the bounds for u, derived in the previous subsections, and the moment estimate, contained in Lemma 3.9, can be proved like in Sections 3.2.1–3.2.2. All these estimates are uniform in η. It is rather straightforward to perform the limit $\eta \to 0$ in (3.33)–(3.34) to obtain a weak solution to (3.7). However, the higher regularity bounds obtained in Section 3.2.3 imply that u is actually a strong solution to (3.7), which in turn yields the uniqueness of u as a weak solution to (3.7). This finishes the proof of Theorem 3.4.

3.2.5 Limit $\sigma \to 0$

We prove that there exists a subsequence of (ρ_σ) that converges strongly in $L^1(\mathbb{R}^d \times (0,T))$ to a weak solution ρ to (3.1).
The uniform $L^\infty(\mathbb{R}^d \times (0,T))$ bound for ρ_σ in Lemma 3.7 implies that, up to a subsequence, $\rho_\sigma \rightharpoonup^* \rho$ weakly* in $L^\infty(\mathbb{R}^d \times (0,T))$ as $\sigma \to 0$. We deduce from the uniform $L^\infty(0,T;L^1(\mathbb{R}^d))$ bound (3.13) and the moment bound for ρ_σ in Lemma 3.9 that (ρ_σ) is equi-integrable. Thus, by the Dunford–Pettis theorem, again up to a subsequence, $\rho_\sigma \rightharpoonup \rho$ weakly in $L^1(\mathbb{R}^d \times (0,T))$. It follows from the $L^2(0,T;H^1(\mathbb{R}^d))$ estimate (3.16) that $\sigma \Delta \rho_\sigma \to 0$ strongly in $L^2(0,T;H^{-1}(\mathbb{R}^d))$. The estimates in (3.17) and Lemma 3.7 show that $(\partial_t \rho_\sigma)$ is bounded in $L^2(0,T;H^{-1}(\mathbb{R}^d))$ and consequently, up to a subsequence, $\partial_t \rho_\sigma \rightharpoonup \partial_t \rho$ weakly in $L^2(0,T;H^{-1}(\mathbb{R}^d))$. Therefore, the limit $\sigma \to 0$ in (3.7) leads to

$$\partial_t \rho = \operatorname{div}(\overline{\rho_\sigma \nabla (-\Delta)^{-s} f_\sigma(\rho_\sigma)}) \quad \text{in } L^2(0,T;H^{-1}(\mathbb{R}^d)), \tag{3.35}$$

where the overline denotes the weak limit of the corresponding sequence.

We need to identify the weak limit on the right-hand side. The idea is to use the div-curl lemma [45, Theorem 10.21]. For this, we define the vector fields with $d+1$ components

$$U_\sigma := \big(\rho_\sigma, -\rho_\sigma\nabla(-\Delta)^{-s}f_\sigma(\rho_\sigma)\big), \quad V_\sigma := \big(f_\sigma(\rho_\sigma), 0, \ldots, 0\big).$$

Let $R>0$ be arbitrary and write B_R for the ball around the origin with radius R. The $L^\infty(\mathbb{R}^d)$ bound (3.15) for ρ_σ and the $L^2(0,T;H^{1-s}(\mathbb{R}^d))$ bound (3.17) for $f_\sigma(\rho_\sigma)$ show that (U_σ) is bounded in $L^p(B_R\times(0,T))$ for some $p>1$, while (V_σ) is bounded in $L^\infty(B_R\times(0,T))$. Furthermore, by (3.17),

$$\begin{aligned}&\operatorname{div}_{(t,x)} U_\sigma = \sigma\Delta\rho_\sigma \to 0 \quad \text{strongly in } L^2(0,T;H^{-1}(B_R)) \hookrightarrow H^{-1}(B_R\times(0,T)),\\ &\|\operatorname{curl}_{(t,x)} V_\sigma\|_{L^2(0,T;H^{-s}(B_R))} \le C\|\nabla f_\sigma(\rho_\sigma)\|_{L^2(0,T;H^{-s}(B_R))} \le C,\end{aligned}$$

where $\operatorname{curl}_{(t,x)} V_\sigma$ is the antisymmetric part of the Jacobian matrix of V_σ. Hence, by the compact embedding $H^{-s}(B_R\times(0,T)) \hookrightarrow W^{-1,r}(B_R\times(0,T))$ (since $L^2(0,T;H^{-s}(B_R))\subset H^{-s}(B_R\times(0,T))$), the sequence $(\operatorname{curl}_{(t,x)} V_\sigma)$ is relatively compact in $W^{-1,r}(B_R\times(0,T))$ for some $r>1$. Therefore, we can apply the div-curl lemma giving $\overline{U_\sigma\cdot V_\sigma} = \overline{U_\sigma}\cdot\overline{V_\sigma}$ or

$$\overline{\rho_\sigma f_\sigma(\rho_\sigma)} = \rho\overline{f_\sigma(\rho_\sigma)} \quad \text{a.e. in } B_R\times(0,T).$$

By definition (3.9) of $f_\sigma(\rho_\sigma)$, it follows for arbitrary $\rho_\sigma\in[0,L]$ and sufficiently large $L>0$, that

$$\begin{aligned}f_\sigma(\rho_\sigma) &= \int_0^{\rho_\sigma}(\Gamma_\sigma * (f'1_{[0,\infty)}))(u)\widetilde{\Xi}(\sigma u)\mathrm{d}u = \int_0^{\rho_\sigma}\int_0^\infty \Gamma_\sigma(u-w)f'(w)\mathrm{d}w\widetilde{\Xi}(\sigma u)\mathrm{d}u\\ &= \int_0^{\rho_\sigma}\int_0^\infty \Gamma'_\sigma(u-w)f(w)\mathrm{d}w\widetilde{\Xi}(\sigma u)\mathrm{d}u = \int_0^\infty\bigg(\int_0^{\rho_\sigma}\Gamma'_\sigma(u-w)\widetilde{\Xi}(\sigma u)\mathrm{d}u\bigg)f(w)\mathrm{d}w.\end{aligned}$$

We use the properties that (ρ_σ) is uniformly bounded and $\widetilde{\Xi}=1$ in $[-1,1]$. Then, choosing $\sigma>0$ sufficiently small,

$$\begin{aligned}f_\sigma(\rho_\sigma) &= \int_0^\infty\bigg(\int_0^{\rho_\sigma}\Gamma'_\sigma(u-w)\mathrm{d}u\bigg)f(w)\mathrm{d}w\\ &= \int_0^\infty \Gamma_\sigma(\rho_\sigma-w)f(w)\mathrm{d}w - \int_0^\infty\Gamma_\sigma(-w)f(w)\mathrm{d}w\\ &= \int_{\mathbb{R}}\Gamma_\sigma(\rho_\sigma-w)\tilde{f}(w)\mathrm{d}w - \int_{\mathbb{R}}\Gamma_\sigma(-w)\tilde{f}(w)\mathrm{d}w,\end{aligned}$$

setting $\tilde{f} := f1_{[0,\infty)}$. Hence, using $f(0)=0$, we find that

$$f_\sigma(\rho_\sigma) - f(\rho_\sigma) = \int_{\mathbb{R}}\Gamma_\sigma(u)(\tilde{f}(u+\rho_\sigma)-\tilde{f}(\rho_\sigma))\mathrm{d}u - \int_{\mathbb{R}}\Gamma_\sigma(-w)(\tilde{f}(w)-\tilde{f}(0))\mathrm{d}w.$$

Taking into account the fundamental theorem of calculus for the function $\tilde{f}\in C^0\cap W^{1,1}(\mathbb{R})$, we can estimate as follows:

$$|f_\sigma(\rho_\sigma)-f(\rho_\sigma)| \le \operatorname*{ess\,sup}_{u\in\operatorname{supp}(\Gamma_\sigma)\backslash\{0\}}\bigg(\frac{|\tilde{f}(u+\rho_\sigma)-\tilde{f}(\rho_\sigma)|}{|u|} + \frac{|\tilde{f}(u)-\tilde{f}(0)|}{|u|}\bigg)\int_{\mathbb{R}}\Gamma_\sigma(w)|w|\mathrm{d}w$$

$$\leq \left(\max_{\xi \in \operatorname{supp}(\Gamma_\sigma) \cap [0,\infty)} (f'(\xi + \rho_\sigma) + f'(\xi)) \right) \int_{\mathbb{R}} \Gamma_\sigma(w)|w|\mathrm{d}w.$$

Then, since $\Gamma_\sigma(u) = \sigma^{-1}\Gamma_1(\sigma^{-1}u)$, $\operatorname{supp}(\Gamma_\sigma) \subset B_\sigma(0)$ is compact, $f \in C^1([0,\infty))$, and (ρ_σ) is uniformly bounded, we conclude that

$$|f_\sigma(\rho_\sigma) - f(\rho_\sigma)| \leq C\sigma.$$

This means that $f_\sigma(\rho_\sigma) - f(\rho_\sigma) \to 0$ strongly in $L^\infty(B_R \times (0,T))$, and it shows that $\overline{\rho_\sigma f(\rho_\sigma)} = \rho\overline{f(\rho_\sigma)}$ a.e. in $B_R \times (0,T)$. As f is nondecreasing, we can apply [45, Theorem 10.19] to infer that $\overline{f(\rho_\sigma)} = f(\rho)$ a.e. in $B_R \times (0,T)$. Consequently, $\overline{\rho_\sigma f(\rho_\sigma)} = \rho f(\rho)$. As $u \mapsto uf(u)$ is assumed to be strictly convex, we conclude from [45, Theorem 10.20] that (ρ_σ) converges a.e. in $B_R \times (0,T)$. Since (ρ_σ) is bounded in $L^\infty(\mathbb{R}^d \times (0,T))$, it follows that $\rho_\sigma \to \rho$ strongly in $L^p(B_R \times (0,T))$ for all $p < \infty$. Using the moment estimate from Lemma 3.9, we infer from

$$\begin{aligned} \limsup_{\sigma\to 0} \int_0^T \int_{\mathbb{R}^d} |\rho_\sigma - \rho| \mathrm{d}x\mathrm{d}t &= \limsup_{\sigma\to 0} \int_0^T \int_{\mathbb{R}^d\setminus B_R} |\rho_\sigma - \rho|\mathrm{d}x\mathrm{d}t \\ &\leq R^{-2d/(d-2s)} \limsup_{\sigma\to 0} \int_0^T \int_{\mathbb{R}^d\setminus B_R} \rho_\sigma(t,x)|x|^{2d/(d-2s)}\mathrm{d}x \\ &\leq R^{-2d/(d-2s)}C \to 0 \quad \text{as } R \to \infty \end{aligned}$$

that $\rho_\sigma \to \rho$ strongly in $L^p(\mathbb{R}^d \times (0,T))$ for all $p < \infty$. The strong convergences of ρ_σ and $f_\sigma(\rho_\sigma)$ in $L^p(\mathbb{R}^d \times (0,T))$ for all $p < \infty$ allow us to identify the weak limit in (3.35), proving the weak formulation (3.8).

Finally, we deduce from the uniform $L^2(0,T;H^{-1}(\mathbb{R}^d))$ bound for $\partial_t\rho_\sigma$ and the fact that $\rho_\sigma \to \rho$ strongly in $L^p(\mathbb{R}^d)$ for any $p < \infty$ that $\rho(0) = \rho^0$ in the sense of $H^{-1}(\mathbb{R}^d)$. Properties (iv) of Theorem 3.1 follow from the corresponding expressions satisfied by ρ_σ in the limit $\sigma \to 0$.

3.2.6 Time-uniform convergence of (ρ_σ).

The following lemma is needed in the proof of Proposition 3.3. It is essentially a consequence of the $L^2(0,T;H^{-1}(\mathbb{R}^d))$ bound of $\partial_t\rho_\sigma$ and the Ascoli–Arzelà theorem.

Corollary 3.13. *Under the assumptions of Theorem 3.1, it holds for all $\phi \in L^\infty(\mathbb{R}^d)$ that, possibly for a subsequence,*

$$\int_{\mathbb{R}^d} \rho_\sigma\phi\mathrm{d}x \to \int_{\mathbb{R}^d} \rho\phi\mathrm{d}x \quad \textit{uniformly in } [0,T].$$

Proof. Let $\phi \in C_0^1(\mathbb{R}^d)$ and $0 \leq t_1 < t_2 \leq T$. The uniform $L^2(0,T;H^{-1}(\mathbb{R}^d))$ bound of $\partial_t\rho_\sigma$ implies that

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \rho_\sigma(t_2)\phi\mathrm{d}x - \int_{\mathbb{R}^d} \rho_\sigma(t_1)\phi\mathrm{d}x \right| &= \left| \int_{t_1}^{t_2} \langle \partial_t\rho_\sigma, \phi\rangle \mathrm{d}t \right| \\ &\leq |t_2 - t_1|^{1/2} \|\partial_t\rho_\sigma\|_{L^2(0,T;H^{-1}(\mathbb{R}^d))} \|\phi\|_{H^1(\mathbb{R}^d)} \leq C|t_2 - t_1|^{1/2}\|\phi\|_{H^1(\mathbb{R}^d)}. \end{aligned}$$

Hence, the sequence of functions $t \mapsto \int_{\mathbb{R}^d} \rho_\sigma(t)\phi \mathrm{d}s$ is bounded and equicontinuous in $[0,T]$. By the Ascoli–Arzelá theorem, up to a ϕ-depending subsequence, $\int_{\mathbb{R}^d} \rho_\sigma \phi \mathrm{d}x \to \xi_\phi$ strongly in $C^0([0,T])$ as $\sigma \to 0$. Since $\rho_\sigma \rightharpoonup^* \rho$ weakly* in $L^\infty(0,T;L^\infty(\mathbb{R}^d))$, we can identify the limit, $\xi_\phi = \int_{\mathbb{R}^d} \rho\phi \mathrm{d}x$. Since $H^1(\mathbb{R}^d)$ is separable, a Cantor diagonal argument together with a density argument allows us to find a subsequence (which is not relabeled) such that for all $\phi \in H^1(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \rho_\sigma \phi \mathrm{d}x \to \int_{\mathbb{R}^d} \rho\phi \mathrm{d}x \quad \text{strongly in } C^0([0,T]). \tag{3.36}$$

Since (ρ_σ) is bounded in $L^\infty(0,T;L^2(\mathbb{R}^d))$, another density argument shows that this limit also holds for all $\phi \in L^2(\mathbb{R}^d)$.
Now, let $\phi \in L^\infty(\mathbb{R}^d)$. Using $\phi 1_{\{|x|<R\}} \in L^2(\mathbb{R}^d)$, it follows from (3.36) and the moment estimate for ρ_σ that

$$\begin{aligned}
&\limsup_{\sigma\to 0} \sup_{0<t<T} \left| \int_{\mathbb{R}^d} \rho_\sigma(t)\phi \mathrm{d}x - \int_{\mathbb{R}^d} \rho(t)\phi \mathrm{d}x \right| \\
&\quad \le \limsup_{\sigma\to 0} \sup_{0<t<T} \left| \int_{\mathbb{R}^d} \rho_\sigma(t)\phi 1_{\{|x|>R\}} \mathrm{d}x - \int_{\mathbb{R}^d} \rho(t)\phi 1_{\{|x|>R\}} \mathrm{d}x \right| \\
&\quad \le R^{-2d/(d-2s)} \|\phi\|_\infty \limsup_{\sigma\to 0} \sup_{0<t<T} \int_{\mathbb{R}^d} (\rho_\sigma(x,t) + \rho(x,t))|x|^{2d/(d-2s)} \mathrm{d}x \\
&\quad \le C(T) R^{-2d/(d-2s)} \|\phi\|_\infty \to 0 \quad \text{as } R \to \infty.
\end{aligned}$$

This shows that

$$\lim_{\sigma\to 0} \sup_{0<t<T} \left| \int_{\mathbb{R}^d} \rho_\sigma(t)\phi \mathrm{d}x - \int_{\mathbb{R}^d} \rho(t)\phi \mathrm{d}x \right| = 0,$$

concluding the proof. □

3.3 Analysis of equation of the regularised equation (3.5)

This section is devoted to the analysis of equation (3.5),

$$\begin{aligned}
&\partial_t \rho_{\sigma,\beta,\zeta} - \sigma\Delta\rho_{\sigma,\beta,\zeta} = \operatorname{div}\left(\rho_{\sigma,\beta,\zeta} \nabla \mathcal{K}_\zeta * f_\sigma(\mathcal{W}_\beta * \rho_{\sigma,\beta,\zeta})\right), \quad t > 0, \\
&\rho_{\sigma,\beta,\zeta}(0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d,
\end{aligned} \tag{3.37}$$

where $\mathcal{K}_\zeta = \widetilde{\mathcal{K}}_\zeta * \mathcal{W}_\zeta$ and $\mathcal{W}_\beta$ is defined in (3.10), as well as to an estimate for the difference $\rho_{\sigma,\beta,\zeta} - \rho_\sigma$, which is needed in the mean-field analysis. The existence and uniqueness of a strong solution to (3.37) follows from standard parabolic theory, since we regularized the singular kernel and smoothed the nonlinearity.

Proposition 3.14 (Uniform estimates). *Let Hypotheses (H1)–(H3) hold and let $T > 0$, $p > d$. Set $a := \min\{1, d-2s\}$, let ρ_σ be the strong solution to (3.7), and let $\rho_{\sigma,\beta,\zeta}$ be the strong solution to (3.5). Then there exist constants $C_1 > 0$, and $\varepsilon_0 > 0$, both depending on σ, p, and T, such that if $\beta + \zeta^a < \varepsilon_0$ then*

$$\|\rho_{\sigma,\beta,\zeta} - \rho_\sigma\|_{L^\infty(0,T;W^{2,p}(\mathbb{R}^d))} \le C_1(\beta + \zeta^a), \tag{3.38}$$

$$\|\rho_{\sigma,\beta,\zeta}\|_{L^\infty(0,T;W^{2,p}(\mathbb{R}^d))} \le C_1. \tag{3.39}$$

Furthermore, for every $q \ge 2$, there exists $C_2 = C_2(q) > 0$, depending on σ and T, such that

$$\|(\mathcal{K}_\zeta - \mathcal{K}) * \rho_\sigma\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C_2\zeta^a, \tag{3.40}$$

$$\|\rho_{\sigma,\beta,\zeta}\|_{L^\infty(0,T;W^{2,1}(\mathbb{R}^d)\cap W^{3,q}}(\mathbb{R}^d)) \le C_2. \tag{3.41}$$

The proof is presented in the following subsections. The most difficult part is the proof of (3.38) in Section 3.3.1. We first prove an estimate for $\mathrm{D}^2(\rho_{\sigma,\beta,\zeta} - \rho_\sigma)$ that depends on a lower-order estimate of this difference. Second, this lower-order estimate is shown by testing the equation satisfied by the difference $\rho_{\sigma,\beta,\zeta} - \rho_\sigma$ with a suitable nonlinear test function. Based on the arguments of this section, estimates (3.39)–(3.41) are then shown in Sections 3.3.2–3.3.4, respectively.

3.3.1 Proof of (3.38).

We introduce the difference $u := \rho_{\sigma,\beta,\zeta} - \rho_\sigma$, which satisfies

$$\begin{aligned}\partial_t u - \sigma\Delta u &= \operatorname{div}\big[(u+\rho_\sigma)\nabla\mathcal{K}_\zeta * f_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - \rho_\sigma\nabla\mathcal{K} * f_\sigma(\rho_\sigma)\big] \\ &= D[u] + R[\rho_\sigma, u] + S[\rho_\sigma, u] \quad \text{in } \mathbb{R}^d,\ t > 0,\end{aligned} \tag{3.42}$$

and the initial datum $u(0) = 0$ in $\mathbb{R}^d$, where

$$\begin{aligned} D[u] &= \operatorname{div}\big[u\nabla\mathcal{K} * f_\sigma(\mathcal{W}_\beta * u)\big], \\ R[\rho_\sigma, u] &= \operatorname{div}\big[u\nabla\mathcal{K} * \big(f_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f_\sigma(\mathcal{W}_\beta * u)\big) \\ &\quad + \rho_\sigma\nabla\mathcal{K} * \big(f_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big) + \rho_\sigma\nabla\mathcal{K} * \big(f_\sigma(\mathcal{W}_\beta * \rho_\sigma) - f_\sigma(\rho_\sigma)\big)\big], \\ S[\rho_\sigma, u] &= \operatorname{div}\big[(u+\rho_\sigma)\nabla(\mathcal{K}_\zeta - \mathcal{K}) * f_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\big]. \end{aligned}$$

We show first an estimate for $\mathrm{D}^2 u$ that depends on a lower-order estimate for u.

Lemma 3.15 (Conditional estimate for u)**.** *For any $p > d$, there exists a number $\Gamma_p \in (0,1)$ such that, if $\sup_{0<t<T}\|u(t)\|_{W^{1,p}(\mathbb{R}^d)} \le \Gamma_p$ then*

$$\|\mathrm{D}^2 u\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\big(\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + \beta + \zeta^a\big),$$

recalling that $a = \min\{1, d-2s\}$, and where $C > 0$ is independent of u, β, and ζ, but may depend on σ.

Proof. Let $\Gamma_p \in (0,1)$ be such that $\sup_{0<t<T}\|u(t)\|_{W^{1,p}(\mathbb{R}^d)} \le \Gamma_p$. We will find a constraint for Γ_p at the end of the proof. The aim is to derive an estimate for the right-hand side of (3.42) in $L^p(0,T;L^p(\mathbb{R}^d))$. We observe that $\|u(t)\|_1 \le \|\rho_{\sigma,\beta,\zeta}\|_1 + \|\rho_\sigma\|_1 \le 2\|\rho^0\|_1$ for $t \in [0,T]$. In the following, we denote by $C > 0$ a generic constant that may depend on σ, without making this explicit. Furthermore, we denote by μ a generic exponent in $(0,1)$, whose value may vary from line to line.

Step 1: Estimate of $D[u]$. Let $1/2 < s < 1$. Then, by the Hardy–Littlewood–Sobolev-type inequality (3.68),

$$\begin{aligned}\|D[u]\|_p &\le \|\nabla u \cdot \nabla \mathcal{K} * f_\sigma(\mathcal{W}_\beta * u)\|_p + \|u \nabla \mathcal{K} * [f_\sigma'(\mathcal{W}_\beta * u)\mathcal{W}_\beta * \nabla u]\|_p \\ &\le C\|\nabla u\|_p \|f_\sigma(\mathcal{W}_\beta * u)\|_{d/(2s-1)} + C\|u\|_{d/(2s-1)}\|f_\sigma'(\mathcal{W}_\beta * u)\|_\infty \|\nabla u\|_p.\end{aligned}$$

We use the Young convolution inequality, the Gagliardo–Nirenberg inequality, the smoothness of f_σ, the property $f_\sigma(0) = 0$, and the fact $\|\mathcal{W}_\beta\|_{L^1(\mathbb{R}^d)} = 1$ to estimate the terms on the right-hand side:

$$\begin{aligned}\|\mathcal{W}_\beta * u\|_\infty &\le \|u\|_\infty \le \|u\|_1^{1-\lambda}\|\nabla u\|_p^\lambda \le C\Gamma_p^\lambda \le C, \\ \|f_\sigma(\mathcal{W}_\beta * u)\|_\infty &\le \max_U |f_\sigma'|\, \|\mathcal{W}_\beta * u\|_\infty \le C, \\ \|f_\sigma'(\mathcal{W}_\beta * u)\|_\infty &\le |f_\sigma'(0)| + \max_U |f_\sigma''| \|\mathcal{W}_\beta * u\|_\infty \le C, \\ \|u\|_{d/(2s-1)} &\le \|u\|_1^{1-\mu}\|u\|_\infty^\mu \le C\|u\|_{W^{1,p}(\mathbb{R}^d)}^\mu \le C\Gamma_p^\mu \le C,\end{aligned}$$

where $U := [-\|\mathcal{W}_\beta * u\|_\infty, \|\mathcal{W}_\beta * u\|_\infty]$ and $\lambda > 0$, $\mu > 0$. Therefore, $\|D[u]\|_p \le C\|\nabla u\|_p$ and

$$\|D[u]\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))}. \tag{3.43}$$

Next, let $0 < s \le 1/2$. Then we write

$$\begin{aligned}D[u] &= \nabla u \cdot \mathcal{K} * [f_\sigma'(\mathcal{W}_\beta * u)\mathcal{W}_\beta * \nabla u] + u\mathcal{K} * [f_\sigma''(\mathcal{W}_\beta * u)|\mathcal{W}_\beta * \nabla u|^2] \\ &\quad + u\mathcal{K} * [f_\sigma'(\mathcal{W}_\beta * u)\mathcal{W}_\beta * \Delta u] =: D_1 + D_2 + D_3.\end{aligned}$$

By the Hardy–Littlewood–Sobolev-type inequality (Lemma 3.22),

$$\|D_1\|_p \le C\|\nabla u\|_{d/(2s)}\|f_\sigma'(\mathcal{W}_\beta * u)\mathcal{W}_\beta * \nabla u\|_p \le C\|\nabla u\|_{d/(2s)}\|\nabla u\|_p.$$

Next, we apply the Gagliardo–Nirenberg inequality with $\lambda = (1+1/d-2s/d)/(1+2/d-1/p)$:

$$\|\nabla u\|_{d/(2s)} \le C\|u\|_1^{1-\lambda}\|\mathrm{D}^2 u\|_p^\lambda \le C\|\mathrm{D}^2 u\|_p^\lambda,$$

which is possible as long as $\lambda \ge 1/2$ or equivalently $d \ge 2s$, which is true. Consequently, using $\Gamma_p \le 1$,

$$\|D_1\|_p \le C\|\nabla u\|_p\|\mathrm{D}^2 u\|_p^\lambda \le C\Gamma_p^\lambda\|\nabla u\|_p^{1-\lambda}\|\mathrm{D}^2 u\|_p^\lambda \le C(\delta)\|\nabla u\|_p + \delta\|\mathrm{D}^2 u\|_p,$$

where $\delta > 0$ is arbitrary. It follows from the Hardy-Littlewood-Sobolev-type inequality and the Gagliardo–Nirenberg inequality

$$\|\nabla u\|_{2p}^2 \le C\|\mathrm{D}^2 u\|_p^{d/p}\|\nabla u\|_p^{2-d/p} \le C\Gamma_p\|\mathrm{D}^2 u\|_p^{d/p}\|\nabla u\|_p^{1-d/p}$$

that

$$\|D_2\|_p \le C\|u\|_{d/2s}\Gamma_p\|D^2 u\|_p^{d/p}\|\nabla u\|_p^{1-d/p} \le C(\delta)\|\nabla u\|_p + \delta\|\mathrm{D}^2 u\|_p.$$

Finally, using similar ideas, we obtain

$$\|D_3\|_p \le C\|u\|_{d/(2s)}\|\Delta u\|_p \le C\Gamma_p^\mu\|\mathrm{D}^2 u\|_p.$$

Summarizing the estimates for D_1, D_2, and D_3 and integrating in time leads to

$$\|D[u]\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C\Gamma_p^\mu \|\mathrm{D}^2 u\|_{L^p(0,T;L^p(\mathbb{R}^d))}. \tag{3.44}$$

Step 2: Estimate of $R[\rho_\sigma, u]$. We write $R[\rho_\sigma, u] = R_1 + R_2 + R_3$ for the three terms in the definition of $R[\rho_\sigma, u]$ below (3.42).
Step 2a: Estimate of R_1. If $s > 1/2$, we can argue similarly as in the derivation of (3.43), which gives

$$\|R_1\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))}.$$

If $0 < s \le 1/2$, we write $R_1 = R_{11} + \cdots + R_{16}$, where

$$\begin{aligned}
R_{11} &= \nabla u \cdot \mathcal{K} * \big[f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\mathcal{W}_\beta * \nabla\rho_\sigma\big], \\
R_{12} &= u\mathcal{K} * \big[f''_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))(\mathcal{W}_\beta * \nabla\rho_\sigma)\cdot(\mathcal{W}_\beta * \nabla(u+\rho_\sigma))\big], \\
R_{13} &= u\mathcal{K} * \big[f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\mathcal{W}_\beta * \Delta\rho_\sigma\big], \\
R_{14} &= \nabla u \cdot \mathcal{K} * \big[\big(f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f'_\sigma(\mathcal{W}_\beta * u)\big)\mathcal{W}_\beta * \nabla u\big], \\
R_{15} &= u\mathcal{K} * \big[\big(f''_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\mathcal{W}_\beta * \nabla(u+\rho_\sigma) \\
&\quad - f''_\sigma(\mathcal{W}_\beta * u)(\mathcal{W}_\beta * \nabla u)\big)\cdot(\mathcal{W}_\beta * \nabla u)\big], \\
R_{16} &= u\mathcal{K} * \big[\big(f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f'_\sigma(\mathcal{W}_\beta * u)\big)\mathcal{W}_\beta * \Delta u\big].
\end{aligned}$$

All terms except the last one can be treated by the Hardy–Littlewood–Sobolev and Gagliardo–Nirenberg inequalities as before. For the last term, we use these inequalities and the $L^\infty(\mathbb{R}^d)$ bound for ρ_σ:

$$\begin{aligned}
\|R_{16}\|_p &\le C\|u\|_{d/(2s)}\big\|\big(f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f'_\sigma(\mathcal{W}_\beta * u)\big)\mathcal{W}_\beta * \Delta u\big\|_p \\
&\le C\|u\|_{d/(2s)}\|f''_\sigma\|_\infty \|\mathcal{W}_\beta * \rho_\sigma\|_\infty \|\mathcal{W}_\beta * \Delta u\|_p \le C\|u\|_{d/(2s)}\|\Delta u\|_p \le C\Gamma_p^\mu\|\mathrm{D}^2 u\|_p.
\end{aligned}$$

We infer that (possibly with a different $\mu > 0$ than before)

$$\|R_1\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C\Gamma_p^\mu\|\mathrm{D}^2 u\|_{L^p(0,T;L^p(\mathbb{R}^d))}.$$

Step 2b: Estimate of R_2. Since $|f'_\sigma|$ is bounded on the interval $[-\|u\|_\infty - \|\rho_\sigma\|_\infty, \|u\|_\infty + \|\rho_\sigma\|_\infty]$, we obtain for $s > 1/2$,

$$\|R_2\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))}.$$

For $0 < s \le 1/2$, we write $R_2 = R_{21} + \cdots + R_{27}$, where

$$\begin{aligned}
R_{21} &= \nabla\rho_\sigma \cdot \mathcal{K} * \big[f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\mathcal{W}_\beta * \nabla u\big], \\
R_{22} &= \rho_\sigma\mathcal{K} * \big[f''_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\mathcal{W}_\beta * \nabla(u+\rho_\sigma)\cdot(\mathcal{W}_\beta * \nabla u)\big], \\
R_{23} &= \rho_\sigma\mathcal{K} * \big[f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))\mathcal{W}_\beta * \Delta u\big], \\
R_{24} &= \nabla\rho_\sigma \cdot \mathcal{K} * \big[\big(f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f'_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big)\mathcal{W}_\beta * \nabla\rho_\sigma\big], \\
R_{25} &= \rho_\sigma\mathcal{K} * \big[f''_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma))(\mathcal{W}_\beta * \nabla u)\cdot(\mathcal{W}_\beta * \nabla\rho_\sigma)\big], \\
R_{26} &= \rho_\sigma\mathcal{K} * \big[\big(f''_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f''_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big)|\mathcal{W}_\beta * \nabla\rho_\sigma|^2\big], \\
R_{27} &= \rho_\sigma\mathcal{K} * \big[\big(f'_\sigma(\mathcal{W}_\beta * (u+\rho_\sigma)) - f'_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big)\mathcal{W}_\beta * \Delta\rho_\sigma\big].
\end{aligned}$$

Similar estimations as before allow us to treat all terms except the third one:

$$\begin{aligned}\|R_{23}\|_p &\le \big\|\rho_\sigma \mathcal{K} * \big[\big(f_\sigma'(\mathcal{W}_\beta * (u+\rho_\sigma)) - f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\big)\mathcal{W}_\beta * \Delta u\big]\big\|_p \\ &\quad + \|\rho_\sigma \mathcal{K} * [f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * \Delta u]\|_p =: Q_{231} + Q_{232}.\end{aligned}$$

The first term can be estimated similarly as above by $Q_{231} \le C\Gamma_p^\mu \|\mathrm{D}^2 u\|_p$, while

$$\begin{aligned}Q_{232} &\le \|\rho_\sigma \Delta\mathcal{K} * [f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * u]\|_p + \|\rho_\sigma \mathcal{K} * [\Delta f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * u]\|_p \\ &\quad + 2\|\rho_\sigma \mathcal{K} * [\nabla f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\cdot(\mathcal{W}_\beta * \nabla u)]\|_p.\end{aligned}$$

It follows from $-\Delta\mathcal{K} * v = (-\Delta)^{1-s} v$ and the fractional Gagliardo–Nirenberg inequality (Lemma 3.24) that

$$\begin{aligned}Q_{232} &\le C\|u\|_{W^{1,p}(\mathbb{R}^d)} + \|\rho_\sigma(-\Delta)^{1-s}[f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * u]\|_p \\ &\le C\|u\|_{W^{1,p}(\mathbb{R}^d)} + C\|\rho_\sigma\|_\infty \|f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * u\|_p^s \|\mathrm{D}^2[f_\sigma'(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * u]\|_p^{1-s} \\ &\le C\|u\|_{W^{1,p}(\mathbb{R}^d)} + C\|u\|_p^s\big(\|u\|_p^{1-s} + \|\nabla u\|_p^{1-s} + \|\mathrm{D}^2 u\|_p^{1-s}\big) \\ &\le C\|u\|_{W^{1,p}(\mathbb{R}^d)} + C\Gamma_p \|\mathrm{D}^2 u\|_p.\end{aligned}$$

This shows that $\|R_{23}\|_p \le C\|u\|_{W^{1,p}(\mathbb{R}^d)} + C\Gamma_p^\mu \|\mathrm{D}^2 u\|_p$, and we conclude that

$$\|R_2\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C\Gamma_p^\mu \|\mathrm{D}^2 u\|_{L^p(0,T;L^p(\mathbb{R}^d))}.$$

Step 2c: Estimate of R_3. We write $R_3 = R_{31} + \cdots + R_{37}$, where

$$\begin{aligned}R_{31} &= \nabla\rho_\sigma \cdot \mathcal{K} * \big[\big(f_\sigma'(\mathcal{W}_\beta * \rho_\sigma) - f_\sigma'(\rho_\sigma)\big)\mathcal{W}_\beta * \nabla\rho_\sigma\big], \\ R_{32} &= \rho_\sigma \mathcal{K} * \big[\big(f_\sigma''(\mathcal{W}_\beta * \rho_\sigma) - f_\sigma''(\rho_\sigma)\big)|\mathcal{W}_\beta * \nabla\rho_\sigma|^2\big], \\ R_{33} &= \rho_\sigma \mathcal{K} * \big[f_\sigma''(\rho_\sigma)(\mathcal{W}_\beta * \nabla\rho_\sigma - \nabla\rho_\sigma)\cdot(\mathcal{W}_\beta * \nabla\rho_\sigma)\big], \\ R_{34} &= \nabla\rho_\sigma \cdot \mathcal{K} * \big[f_\sigma'(\rho_\sigma)(\mathcal{W}_\beta * \nabla\rho_\sigma - \nabla\rho_\sigma)\big], \\ R_{35} &= \rho_\sigma \mathcal{K} * \big[f_\sigma''(\rho_\sigma)\nabla\rho_\sigma \cdot (\mathcal{W}_\beta * \nabla\rho_\sigma - \nabla\rho_\sigma)\big], \\ R_{36} &= \rho_\sigma \mathcal{K} * \big[f_\sigma'(\rho_\sigma)(\mathcal{W}_\beta * \Delta\rho_\sigma - \Delta\rho_\sigma)\big] \\ R_{37} &= \rho_\sigma \mathcal{K} * \big[(f_\sigma'(\mathcal{W}_\beta * \rho_\sigma) - f_\sigma'(\rho_\sigma))\mathcal{W}_\beta * \Delta\rho_\sigma\big]\end{aligned}$$

We start with the estimate of R_{31}. We use the Hardy–Littlewood–Sobolev inequality (Lemma 3.22) and Lemma 3.21 to estimate $\mathcal{W}_\beta * \rho_\sigma - \rho_\sigma$:

$$\begin{aligned}R_{31} &\le C\|\nabla\rho_\sigma\|_{d/s}\|f_\sigma'(\mathcal{W}_\beta * \rho_\sigma) - f_\sigma'(\rho_\sigma)\|_p \|\mathcal{W}_\beta * \nabla\rho_\sigma\|_{d/s} \\ &\le C\|\nabla\rho_\sigma\|_{d/s}^2 \max_{[0,2\|\rho_\sigma\|_\infty]} |f_\sigma''|\, \|\mathcal{W}_\beta * \rho_\sigma - \rho_\sigma\|_p \le C(\sigma)\beta,\end{aligned}$$

also taking into account the $L^\infty(0,T;L^q(\mathbb{R}^d))$ bound for $\nabla\rho_\sigma$; see Proposition 3.10. With this regularity, we can estimate all other terms except R_{34} and R_{36}. Since they have similar structures, we only treat R_{34}. This term is delicate since the factor $f_\sigma'(\rho_\sigma)$ cannot be bounded in $L^q(\mathbb{R}^d)$ for any $q < \infty$. Therefore, one might obtain via Hardy–Littlewood–Sobolev's inequality factors like $\|\nabla\rho_\sigma\|_{q_1}$ and $\|\mathrm{D}^2\rho_\sigma\|_{q_2}$ with either $q_1 < 2$ or $q_2 < 2$.

However, for such factors, an L^∞ bound in time is currently lacking (Proposition 3.10 provides such a bound only for $q \geq 2$). Our idea is to add and subtract the term $f'_\sigma(0)$ since

$$|f'_\sigma(\rho_\sigma) - f'_\sigma(0)| \leq \rho_\sigma \max_{[0,\|\rho_\sigma\|_\infty]} |f''_\sigma| \leq C\rho_\sigma$$

can be controlled. This leads to

$$\begin{aligned}\|R_{34}\|_p &\leq \|\nabla\rho_\sigma \cdot \mathcal{K} * [(f'_\sigma(\rho_\sigma) - f'_\sigma(0))(\mathcal{W}_\beta * \nabla\rho_\beta - \nabla\rho_\beta)\|_p \\ &\quad + \|f'_\sigma(0)\nabla\rho_\sigma \cdot \mathcal{K} * (\mathcal{W}_\beta * \nabla\rho_\sigma - \nabla\rho_\sigma)\|_p \\ &\leq C\beta + |f'_\rho(0)|\|\nabla\rho_\sigma \cdot \mathcal{K} * (\mathcal{W}_\beta * \nabla\rho_\sigma - \nabla\rho_\sigma)\|_p =: C\beta + Q_{341},\end{aligned}$$

as the first term can be estimated in a standard way. For the estimate of Q_{341}, we need to distinguish two cases.
If $1/2 < s \leq 1$, we infer from the Hardy–Littlewoord–Sobolev-type inequality (3.68) that

$$Q_{341} \leq C\|\nabla\rho_\sigma\|_{d/(2s-1)}\|\mathcal{W}_\beta * \rho_\sigma - \rho_\sigma\|_p \leq C\|\nabla\rho_\sigma\|_{d/(2s-1)}\|\nabla\rho_\sigma\|_p\beta \leq C\beta.$$

Next, let $0 < s \leq 1/2$. Then we apply the Hardy–Littlewoord–Sobolev-type inequality (3.67), the standard Gagliardo–Nirenberg inequality for some $\lambda > 0$, and Lemma 3.21:

$$Q_{341} \leq C\|\nabla\rho_\sigma\|_{d/(2s)}\|\mathcal{W}_\beta * \nabla\rho_\sigma - \nabla\rho_\sigma\|_p \leq C\|\rho_\sigma\|_1^{1-\lambda}\|\mathrm{D}^2\rho_\sigma\|_p^\lambda(\beta\|\mathrm{D}^2\rho_\sigma\|_p) \leq C\beta.$$

We conclude that $\|R_{34}\|_p \leq C\beta$ and eventually

$$\|R_3\|_{L^p(0,T;L^p(\mathbb{R}^d))} \leq C\beta.$$

Summarizing the estimates for R_1, R_2, and R_3 finishes this step:

$$\|R[\rho_\sigma, u]\|_{L^p(0,T;L^p(\mathbb{R}^d))} \leq C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C\beta + C\Gamma_p^\mu\|\mathrm{D}^2u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))}. \tag{3.45}$$

Step 3: Estimate of $S[\rho_\sigma, u]$. We formulate this term as $S[\rho_\sigma, u] = S_1 + \cdots + S_4$, where

$$\begin{aligned}S_1 &= \operatorname{div}\big[u\nabla(\mathcal{K}_\zeta - \mathcal{K}) * \big(f_\sigma(\mathcal{W}_\beta * (u + \rho_\sigma)) - f_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big)\big], \\ S_2 &= \operatorname{div}\big(u\nabla(\mathcal{K}_\zeta - \mathcal{K}) * f_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big), \\ S_3 &= \operatorname{div}\big[\rho_\sigma\nabla(\mathcal{K}_\zeta - \mathcal{K}) * \big(f_\sigma(\mathcal{W}_\beta * (u + \rho_\sigma)) - f_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big)\big], \\ S_4 &= \operatorname{div}\big(\rho_\sigma\nabla(\mathcal{K}_\zeta - \mathcal{K}) * f_\sigma(\mathcal{W}_\beta * \rho_\sigma)\big).\end{aligned}$$

The terms S_1, S_2, and S_3 can be treated as the terms in $R[\rho_\sigma, u]$, since they have the same structure and the techniques used to estimate integrals involving $\mathcal{K}$ can be applied to those involving $\mathcal{K}_\zeta$. This leads to (for some $\mu > 0$)

$$\|S_1 + S_2 + S_3\|_{L^p(0,T;L^p(\mathbb{R}^d))} \leq C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C\Gamma_p^\mu\|\mathrm{D}^2u\|_{L^p(0,T;L^p(\mathbb{R}^d))}. \tag{3.46}$$

It remains to estimate S_4. We write $S_4 = S_{41} + S_{42} + S_{43}$, where

$$\begin{aligned}S_{41} &= \nabla\rho_\sigma \cdot (\mathcal{K}_\zeta - \mathcal{K}) * \big[f'_\sigma(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * \nabla\rho_\sigma\big], \\ S_{42} &= \rho_\sigma(\mathcal{K}_\zeta - \mathcal{K}) * \big[f''_\sigma(\mathcal{W}_\beta * \rho_\sigma)|\mathcal{W}_\beta * \nabla\rho_\sigma|^2\big],\end{aligned}$$

$$S_{43} = \rho_\sigma(\mathcal{K}_\zeta - \mathcal{K}) * \big[f'_\sigma(\mathcal{W}_\beta * \rho_\sigma)\mathcal{W}_\beta * \Delta\rho_\sigma\big].$$

Observe that, because of the definition of $\mathcal{K}_\zeta = \widetilde{\mathcal{K}}_\zeta * \mathcal{W}_\zeta$ with $\widetilde{\mathcal{K}}_\zeta = \mathcal{K}\omega_\zeta$ (defined in (3.11)), we have $(\mathcal{K}_\zeta - \mathcal{K}) * v = \mathcal{K} * (\mathcal{W}_\zeta * v - v) - (\mathcal{K}(1-\omega_\zeta)) * \mathcal{W}_\zeta * v$ for every function v for which the convolution is defined, and therefore, by the Hardy–Littlewood–Sobolev-type inequality (3.67), Young's convolution inequality, and Lemma 3.21,

$$\begin{aligned}\|\rho_\sigma(\mathcal{K}_\zeta - \mathcal{K}) * v\|_p &\le C\|\rho_\sigma\|_{d/(2s)}\|\mathcal{W}_\zeta * v - v\|_p + C\|\rho_\sigma\|_p\|(\mathcal{K}(1-\omega_\zeta)) * v\|_\infty \\ &\le C\|\rho_\sigma\|_{d/(2s)}\|\nabla v\|_p\zeta + C\|\rho_\sigma\|_p\|\mathcal{K}1_{\mathbb{R}^d\setminus B(0,\zeta^{-1})}\|_\infty\|v\|_1 \\ &\le C\|\rho_\sigma\|_{d/(2s)}\|\nabla v\|_p\zeta + C\zeta^{d-2s}\|\rho_\sigma\|_p\|v\|_1,\end{aligned}$$

Given the regularity properties of ρ_σ (see Lemma 3.11) and the assumptions on f_σ, it follows that

$$\|S_4\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\zeta^{\min\{1,d-2s\}}. \tag{3.47}$$

We conclude from (3.46) and (3.47) that

$$\|S[\rho_\sigma, u]\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C\zeta^a + C\Gamma_p^\mu\|\mathrm{D}^2u\|_{L^p(0,T;L^p(\mathbb{R}^d))}, \tag{3.48}$$

where $a := \min\{1, d-2s\}$.

Step 4: End of the proof. Summarizing (3.44), (3.45), and (3.48), we infer that the right-hand side of (3.42) can be bounded (for some $\mu > 0$) by

$$\begin{aligned}&\|D[u] + R[\rho_\sigma, u] + S[\rho_\sigma, u]\|_{L^p(0,T;L^p(\mathbb{R}^d))} \\ &\qquad \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C(\beta + \zeta^a) + C\Gamma_p^\mu\|\mathrm{D}^2u\|_{L^p(0,T;L^p(\mathbb{R}^d))}.\end{aligned}$$

By parabolic regularity (3.71),

$$\|\mathrm{D}^2u\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))} + C(\beta + \zeta^a) + C\Gamma_p^\mu\|\mathrm{D}^2u\|_{L^p(0,T;L^p(\mathbb{R}^d))}.$$

Choosing $\Gamma_p > 0$ sufficiently small finishes the proof. □

It remains to estimate the $L^p(0,T;W^{1,p}(\mathbb{R}^d))$ norm of u. This is done in the following lemma.

Lemma 3.16 (Unconditional estimate for u)**.** *For any $p > d$, there exist constants $C > 0$, and $\varepsilon_0 > 0$, both depending on σ, p, and T, such that for $\beta + \zeta^a < \varepsilon_0$,*

$$\|u\|_{L^\infty(0,T;W^{1,p}(\mathbb{R}^d))} \le C(\beta + \zeta^a).$$

recalling that $a := \min\{1, d-2s\}$.

Proof. The idea is to test (3.42) with $p|u|^{p-2}u - p\,\mathrm{div}(|\nabla u|^{p-2}\nabla u)$. Integration by parts and some elementary computations lead to

$$\int_{\mathbb{R}^d} p\,\mathrm{div}(|\nabla u|^{p-2}\nabla u)\Delta u \mathrm{d}x = -p\sum_{i,j}\int_{\mathbb{R}^d}|\nabla u|^{p-2}\partial_i u\partial_i\partial_{jj}^2 u\mathrm{d}x$$

$$\begin{aligned}
&= p\sum_{i,j}\int_{\mathbb{R}^d}\partial_j(|\nabla u|^{p-2}\partial_i u)\partial_{ij}^2 u\mathrm{d}x \\
&= p\int_{\mathbb{R}^d}|\nabla u|^{p-2}|D^2u|^2\mathrm{d}x + \frac{p}{2}\sum_j\int_{\mathbb{R}^d}\partial_j(|\nabla u|^{p-2})\partial_j(|\nabla u|^2)\mathrm{d}x \\
&= p\int_{\mathbb{R}^d}|\nabla u|^{p-2}|D^2u|^2\mathrm{d}x + \sum_j\int_{\mathbb{R}^d}\frac{4}{p}(p-2)\big(\partial_j(|\nabla u|^{p/2})\big)^2\mathrm{d}x.
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
p\|u(t)\|^p_{W^{1,p}(\mathbb{R}^d)} &+ \sigma p(p-1)\int_0^t\int_{\mathbb{R}^d}|u|^{p-2}|\nabla u|^2\mathrm{d}x\mathrm{d}s \\
&+ \sigma\int_0^t\int_{\mathbb{R}^d}\big(p|\nabla u|^{p-2}|D^2u|^2 + 4(p-2)p^{-1}\big|\nabla(|\nabla u|^{p/2})\big|^2\big)\mathrm{d}x\mathrm{d}s \\
&= p\int_0^t\int_{\mathbb{R}^d}\big(|u|^{p-2}u - \operatorname{div}(|\nabla u|^{p-2}\nabla u)\big)\big(D[u]+R[\rho_\sigma,u]+S[\rho_\sigma,u]\big)\mathrm{d}x\mathrm{d}s \\
&=: Q[u].
\end{aligned} \tag{3.49}$$

We infer from Lemmas 3.20 and 3.26 that $u\in C^0([0,T];W^{1,p}(\mathbb{R}^d))$. Therefore, since $u(0)=0$, it holds that $\|u(t)\|_{W^{1,p}(\mathbb{R}^d)}\le\Gamma_p$ for all $t\in[0,T^*]$ and $T^*:=\sup\{t_0\in(0,T):\|u(t)\|_{W^{1,p}(\mathbb{R}^d)}\le\Gamma_p$ for $0\le t\le t_0\}$. Let $t\in[0,T^*]$. We have shown in the proof of the previous lemma that

$$\|D[u]+R[\rho_\sigma,u]+S[\rho_\sigma,u]\|_{L^p(0,t;L^p(\mathbb{R}^d))}\le C\|u\|_{L^p(0,t;W^{1,p}(\mathbb{R}^d))}+C(\beta+\zeta^a).$$

Hence, we can estimate the right-hand side $Q[u]$ of (3.49) as follows:

$$\begin{aligned}
Q[u] &\le C\int_0^t\int_{\mathbb{R}^d}\big(|u|^{p-1}+|\nabla u|^{p-2}|\mathrm{D}^2u|\big)\big|D[u]+R[\rho_\sigma,u]+S[\rho_\sigma,u]\big|\mathrm{d}x\mathrm{d}s \\
&\le C\big(\|u\|^{p-1}_{L^p(0,t;L^p(\mathbb{R}^d))}+\|\nabla u\|^{p/2-1}_{L^p(0,t;L^p(\mathbb{R}^d))}\||\nabla u|^{p/2-1}|\mathrm{D}^2u|\|_{L^2(0,t;L^2(\mathbb{R}^d))}\big) \\
&\quad\times\big(\|u\|_{L^p(0,t;W^{1,p}(\mathbb{R}^d))}+\beta+\zeta^a\big) \\
&\le C(\delta,p,t)\big(\|u\|^p_{L^p(0,t;W^{1,p}(\mathbb{R}^d))}+(\beta+\zeta^a)^p\big)+\delta\||\nabla u|^{p/2-1}|\mathrm{D}^2u|\|^2_{L^2(0,t;L^2(\mathbb{R}^d))},
\end{aligned}$$

where $\delta>0$. Choosing δ sufficiently small, the last term is absorbed by the corresponding expression on the left-hand side of (3.49), and we infer from (3.49) that for $0\le t\le T^*$,

$$\|u(t)\|^p_{W^{1,p}(\mathbb{R}^d)}\le C(p,t)\int_0^t\|u\|^p_{W^{1,p}(\mathbb{R}^d)}\mathrm{d}s+C(p,t)(\beta+\zeta^a)^p.$$

We assume without loss of generality that $C(p,t)$ is nondecreasing in t. Then Gronwall's lemma implies that for $0\le t\le T^*$,

$$\|u(t)\|^p_{W^{1,p}(\mathbb{R}^d)}\le C(p,T)(\beta+\zeta^a)^p\int_0^t e^{C(p,T)(t-s)}\mathrm{d}s\le(\beta+\zeta^a)e^{C(p,T)t}.$$

Choosing $\varepsilon_0=\frac{1}{2}\Gamma_p\exp(-C(p,T)T/p)<1$, we find that $\|u(t)\|_{W^{1,p}(\mathbb{R}^d)}\le\Gamma_p/2$ for $\beta+\zeta^a<\varepsilon_0$ and $0\le t\le T^*$. By definition of T^*, it follows that $T^*=T$. In particular, $\|u(t)\|_{W^{1,p}(\mathbb{R}^d)}\le C(\beta+\zeta^a)$ for $0<t<T$, which finishes the proof. □

3.3.2 Proof of (3.38) and (3.39).

Combining Lemmas 3.15 and 3.16 leads to

$$\|u\|_{L^p(0,T;W^{2,p}(\mathbb{R}^d))} \le C(\sigma,p,T)(\beta+\zeta^a), \quad \text{where } a=\min\{1,d-2s\}, \tag{3.50}$$

as long as $\beta+\zeta^a<\varepsilon_0$ and $p>d$. Next, we differentiate (3.42) with respect to x_i (writing ∂_i for $\partial/\partial x_i$):

$$\partial_t(\partial_i u)-\sigma\Delta(\partial_i u)=\partial_i\big(D[u]+R[\rho_\sigma,u]+S[\rho_\sigma,u]\big), \quad \partial_i u(0)=0 \quad \text{in } \mathbb{R}^d.$$

Taking into account estimate (3.50) and arguing as in the proof of Lemma 3.15, we can show that for $\delta>0$,

$$\|\partial_i(D[u]+R[\rho_\sigma,u]+S[\rho_\sigma,u])\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C(p,\sigma,\delta)(\beta+\zeta^a)+\delta\|\mathrm{D}^3u\|_{L^p(0,T;L^p(\mathbb{R}^d))}.$$

We infer from parabolic regularity (Lemma 3.26) for sufficiently small $\delta>0$ that

$$\|\partial_t\mathrm{D}u\|_{L^p(0,T;L^p(\mathbb{R}^d))}+\|\mathrm{D}^3u\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C(p,\sigma)(\beta+\zeta^a).$$

Then Lemma 3.20, applied to $\mathrm{D}u$, leads to (3.38), which with Proposition 3.10 implies (3.39).

3.3.3 Proof of (3.40).

Let $x\in\mathbb{R}^d$. We use the definitions of $\mathcal{K}_\zeta$ and $\mathcal{W}_\zeta$ to find that

$$\begin{aligned}
|(\mathcal{K}_\zeta-\mathcal{K})*\rho_\sigma(x)| &= \left|\int_{\mathbb{R}^d}\mathcal{W}_\zeta(x-y)\big((\mathcal{K}*\rho_\sigma)(x)-((\mathcal{K}\omega_\zeta)*\rho_\sigma)(y)\big)\mathrm{d}y\right| \\
&\le \int_{\mathbb{R}^d}\mathcal{W}_\zeta(x-y)|x-y|\frac{|(\mathcal{K}*\rho_\sigma)(x)-(\mathcal{K}*\rho_\sigma)(y)|}{|x-y|}\mathrm{d}y+\|(\mathcal{K}(1-\omega_\zeta))*\rho_\sigma\|_\infty \\
&\le \|\nabla\mathcal{K}*\rho_\sigma\|_\infty\int_{\mathbb{R}^d}\mathcal{W}_\zeta(z)|z|\mathrm{d}z+\|\mathcal{K}1_{\mathbb{R}^d\setminus B(0,\zeta^{-1})}\|_\infty\|\rho_\sigma\|_1 \\
&\le \zeta\|\nabla\mathcal{K}*\rho_\sigma\|_\infty\int_{\mathbb{R}^d}\mathcal{W}_1(y)|y|\mathrm{d}y+\zeta^{d-2s}\|\rho_\sigma\|_1.
\end{aligned}$$

Let $\phi\in C_0^\infty(\mathbb{R}^d)$ be such that $\mathrm{supp}(\phi)\subset B_2$ and $\phi=1$ in B_1. Then (since we can assume without loss of generality that $\zeta<1$), by arguing like in the derivation of (3.47), we obtain

$$|(\mathcal{K}_\zeta-\mathcal{K})*\rho_\sigma(x)| \le C\zeta^{\min\{1,d-2s\}}\big(\|\nabla(\mathcal{K}\phi)*\rho_\sigma\|_\infty+\|\nabla(\mathcal{K}(1-\phi))*\rho_\sigma\|_\infty+\|\rho_\sigma\|_1\big),$$

A computation shows that for $p>\max\{d/(2s),2\}$,

$$\begin{aligned}
\|\nabla(\mathcal{K}\phi)*\rho_\sigma\|_\infty &= \|(\mathcal{K}\phi)*\nabla\rho_\sigma\|_\infty \le \|\mathcal{K}\phi\|_{p/(p-1)}\|\nabla\rho_\sigma\|_p \le C\|\nabla\rho_\sigma\|_p, \\
\|\nabla(\mathcal{K}(1-\phi))*\rho_\sigma\|_\infty &\le \|\nabla(\mathcal{K}(1-\phi))\|_\infty\|\rho_\sigma\|_1 \le C\|\rho_\sigma\|_1,
\end{aligned}$$

where we note that $\mathcal{K}1_{B_2}\in L^{p/(p-1)}$ if $p>d/(2s)$. Then, in view of the regularity of ρ_σ in Lemma 3.11, we find that

$$\|(\mathcal{K}_\zeta-\mathcal{K})*\rho_\sigma\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\zeta^a.$$

3.3.4 Proof of (3.41).

The $L^\infty(0,T;W^{2,1}(\mathbb{R}^d)\cap W^{3,q}(\mathbb{R}^d))$ bound for $\rho_{\sigma,\beta,\zeta}$ is shown in a similar way as the corresponding bound for ρ_σ in Lemma 3.11.

3.4 Mean-field analysis

This section is devoted to the proof of Proposition 3.3 and Theorem 3.2. The existence of solutions to (3.4) and (3.6) as well as the existence of density functions is shown in Section 3.4.1. In Section 3.4.2, we estimate the difference $X_i^N - \bar{X}_i^N$ of the processes of the original system (3.3) and the intermediate system (3.4), while the difference $\bar{X}_i^N - \widehat{X}_i^N$ of the processes of the intermediate system (3.4) and the macroscopic system (3.6) is estimated in Section 3.4.3. These estimates are combined in Section 3.4.4 to conclude with the proof of Proposition 3.3 and Theorem 3.2.

3.4.1 Existence of density functions for (3.4) and (3.6)

First, we show that the coefficients of the stochastic differential equation (3.6), satisfied by $\widehat{X}^N$, are globally Lipschitz continuous and of at most linear growth. The latter condition follows from

$$\begin{aligned}
|\nabla\mathcal{K}*f_\sigma(\rho_\sigma(x,t))| &\le \|\mathcal{K}*\nabla f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\
&\le C\|\mathcal{K}*\nabla f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;W^{1,p}(\mathbb{R}^d))} \le C\|\nabla f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;W^{1,r}(\mathbb{R}^d))} \le C(\sigma),
\end{aligned}$$

where $p>d$ and $r=dp/(d+2s)$ according to the Hardy–Littlewood–Sobolev inequality, and we used the regularity bounds for ρ_σ from Lemma 3.26. The global Lipschitz continuity is a consequence of the mean-value theorem, the Hardy–Littlewood–Sobolev inequality, and the $W^{2,\infty}(\mathbb{R}^d)$ regularity of ρ_σ from Lemma 3.11:

$$\begin{aligned}
\sup_{0<t<T}\big|\nabla\mathcal{K}*f_\sigma(\rho_\sigma(x,t)) - \nabla\mathcal{K}*f_\sigma(\rho_\sigma(y,t))\big| &\le \sup_{0<t<T}\|\mathrm{D}^2\mathcal{K}*f_\sigma(\rho_\sigma(\cdot,t))\|_\infty|x-y| \\
= \sup_{0<t<T}\big\|\mathcal{K}*\big(f_\sigma''(\rho_\sigma)\nabla\rho_\sigma\otimes\nabla\rho_\sigma + f_\sigma'(\rho_\sigma)\mathrm{D}^2\rho_\sigma\big)(\cdot,t)\big\|_\infty &|x-y| \le C(\sigma)|x-y|.
\end{aligned}$$

These two conditions yield the existence and uniqueness of solutions to the associated particle systems [67, Theorems 2.5 and 2.9]. Moreover, by [89, Theorem 2.3.1], the law of the process $\widehat{X}_i^N$ is absolutely continuous with respect to the Lebesgue measure. By Radon-Nikodym's theorem, there exists a density function $\widehat{u}(t)$ for all $t>0$ on $\mathbb{R}^d$, which is measurable and integrable with respect to the Lebesgue measure. (Since all $\widehat{X}_i^N$ are copies of the same process, their density functions are the same almost everywhere.) The processes $\widehat{X}_i^N(t)$ have continuous paths, which implies the continuity of the distribution function of $\widehat{X}_i^N(t)$ with respect to time, and this implies in turn the Bochner measurability of $\widehat{u}(t)$. Clearly, we have $\sup_{0<t<T}\|\widehat{u}(t)\|_{L^1(\mathbb{R}^d)}=1$, which shows that $\widehat{u}\in L^\infty(0,T;L^1(\mathbb{R}^d))$.
Similar arguments show that $\bar{X}_i^N(t)$ has a density function $\bar{u}\in L^\infty(0,T;L^1(\mathbb{R}^d))$.
Next, we show that $\widehat{u}$ and $\bar{u}$ can be identified with the weak solutions ρ_σ and $\rho_{\sigma,\beta,\zeta}$, respectively, using Itô's lemma. Indeed, let $\phi\in C_0^\infty(\mathbb{R}^d\times[0,T])$. We infer from Itô's

formula that

$$\phi(\widehat{X}_i^N(t),t) = \phi(\widehat{X}_i^N(0),0) + \int_0^t \partial_s\phi(\widehat{X}_i^N(s),s)\mathrm{d}s + \sigma\int_0^t \Delta\phi(\widehat{X}_i^N(s),s)\mathrm{d}s - \int_0^t \nabla\mathcal{K} * f_\sigma(\rho_\sigma(\widehat{X}_i^N(s),s))\cdot\nabla\phi(\widehat{X}_i^N(s),s)\mathrm{d}s + \sqrt{2\sigma}\int_0^t \nabla\phi(\widehat{X}_i^N(s),s)\cdot\mathrm{d}W_i(s).$$

Taking the expectation, the Itô integral vanishes, and we end up with

$$\int_{\mathbb{R}^d}\phi(x,t)\widehat{u}(x,t)\mathrm{d}x = \int_{\mathbb{R}^d}\phi(x,0)\rho_\sigma^0(x)\mathrm{d}x + \int_0^t\int_{\mathbb{R}^d}\partial_s\phi(x,s)\widehat{u}(x,s)\mathrm{d}x\mathrm{d}s + \sigma\int_0^t\int_{\mathbb{R}^d}\Delta\phi(x,s)\widehat{u}(x,s)\mathrm{d}x\mathrm{d}s - \int_0^t\int_{\mathbb{R}^d}\nabla\mathcal{K} * f_\sigma(\rho_\sigma(x,s))\cdot\nabla\phi(x,s)\widehat{u}(x,s)\mathrm{d}x\mathrm{d}s. \tag{3.51}$$

Hence, $\widehat{u}$ is a very weak solution in the space $L^\infty(0,T;L^1(\mathbb{R}^d))$ to the linear equation

$$\partial_t\widehat{u} = \sigma\Delta\widehat{u} + \operatorname{div}(\widehat{u}\nabla\mathcal{K} * f_\sigma(\rho_\sigma)), \quad \widehat{u}(0) = \rho_\sigma^0 \quad \text{in } \mathbb{R}^d, \tag{3.52}$$

where ρ_σ is the unique solution to (3.7).
It can be shown that (3.52) is uniquely solvable in the class of functions in $L^\infty(0,T;L^1(\mathbb{R}^d))$. This implies that $\widehat{u} = \rho_\sigma$ in $\mathbb{R}^d\times(0,T)$ (and similarly $\bar{u} = \rho_{\sigma,\beta,\zeta}$). The proof is technical but standard; see, e.g., [25, Theorem 7] for a sketch of a proof.
Another approach is as follows. Because of the linearity of (3.51), it is sufficient to prove that $\widehat{u}\equiv 0$ in $\mathbb{R}^d\times(0,T)$ if $\rho_\sigma^0 = 0$. First, we verify that $v := \nabla\mathcal{K} * f_\sigma(\rho_\sigma) \in L^\infty(0,T;W^{1,\infty}(\mathbb{R}^d))$ and $\widehat{u}\in L^p(0,T;L^p(\mathbb{R}^d))$ for $p < d/(d-1)$. Then, by density, (3.51) holds for all $\phi\in W^{1,q}(0,T;L^q(\mathbb{R}^d))\cap L^q(0,T;W^{2,q}(\mathbb{R}^d))$ with $q > d$ and $\phi(T) = 0$. Choosing ψ to be the unique strong solution to the dual problem

$$\partial_t\psi + \sigma\Delta\psi = v\cdot\nabla\psi + g, \quad \psi(T) = 0 \quad \text{in } \mathbb{R}^d$$

in the very weak formulation of (3.51), we find that $\int_0^T\int_{\mathbb{R}^d}\widehat{u}g\mathrm{d}x\mathrm{d}t = 0$ for all $g\in C_0^\infty(\mathbb{R}^d\times(0,T))$, which implies that $\widehat{u} = 0$.

3.4.2 Estimate of $X_i^N - \bar{X}_i^N$

We derive an estimate for the expectation of the difference $X_i^N - \bar{X}_i^N$. To this end, we need to estimate the difference of the microscopic average $N^{-1}\sum_{j=1,\,j\neq i}^N \mathcal{W}_\beta(X_j^N - X_i^N)$ and the macroscopic average $\mathcal{W}_\beta * \rho_{\beta,\zeta,\sigma}(\bar{X}_i^N)$. By a careful choice of β and ζ, we show that this estimate is of the order $N^{-1/4+\delta}$ for $\delta > 0$.

Lemma 3.17. *Let X_i^N and $\bar{X}_i^N$ be the solutions to (3.3) and (3.4), respectively, and let $\delta\in(0,1/4)$. Under the assumptions of Theorem 3.3 on β and ζ, it holds that*

$$\mathbb{E}\bigg(\sup_{0<s<T}\max_{i=1,\ldots,N}|(X_i^N - \bar{X}_i^N)(s)|\bigg) \le CN^{-1/4+\delta}.$$

Proof. To simplify the presentation, we set

$$\Psi(x,t) := f_\sigma\bigg(\frac{1}{N}\sum_{j=1,\,j\neq i}^{N} \mathcal{W}_\beta(X_j^N(t)-x)\bigg),\ \bar\Psi(x,t) := f_\sigma\bigg(\frac{1}{N}\sum_{j=1,\,j\neq i}^{N} \mathcal{W}_\beta(\bar X_j^N(t)-x)\bigg),$$

and we write $\rho := \rho_{\sigma,\beta,\zeta}$. Taking the difference of equations (3.3) and (3.4) in the integral formulation leads to

$$\sup_{0<s<t}|(X_i^N-\bar X_i^N)(s)| \le \int_0^t \big|\nabla\mathcal{K}_\zeta * \big(\Psi(X_i^N(s),s) - f_\sigma(\mathcal{W}_\beta * \rho(\bar X_i^N(s),s))\big)\big|\mathrm{d}s \tag{3.53}$$

$$\le \int_0^t \big|\nabla\mathcal{K}_\zeta * \big(\Psi(X_i^N(s),s) - \bar\Psi(\bar X_i^N(s),s)\big)\big|\mathrm{d}s$$

$$+ \int_0^t \big|\nabla\mathcal{K}_\zeta * \big(\bar\Psi(\bar X_i^N(s),s) - f_\sigma(\mathcal{W}_\beta * \rho(\bar X_i^N(s),s))\big)\big|\mathrm{d}s =: I_1 + I_2.$$

Step 1: Estimate of I_1. To estimate I_1, we formulate $I_1 = I_{11}+I_{12}+I_{13}$, where

$$I_{11} = \int_0^t \big|\nabla\mathcal{K}_\zeta * \big(\Psi(X_i^N(s),s) - \Psi(\bar X_i^N(s),s)\big)\big|\mathrm{d}s,$$
$$I_{12} = \int_0^t \big|\nabla\mathcal{K}_\zeta * \big(\Psi(\bar X_i^N(s),s) - \bar\Psi(X_i^N(s),s)\big)\big|\mathrm{d}s,$$
$$I_{13} = \int_0^t \big|\nabla\mathcal{K}_\zeta * \big(\bar\Psi(X_i^N(s),s) - \bar\Psi(\bar X_i^N(s),s)\big)\big|\mathrm{d}s.$$

We start with the first integral:

$$I_{11} \le \int_0^t \|\mathrm{D}^2\mathcal{K}_\zeta * \Psi(\cdot,s)\|_\infty \sup_{0<r<s}\max_{i=1,\dots,N}|(X_i^N-\bar X_i^N)(r)|\mathrm{d}s.$$

We claim that

$$\|\mathrm{D}^k\mathcal{K}_\zeta * \Psi(\cdot,s)\|_\infty \le C(\sigma)\beta^{-(k+1)(d+k)-1}, \quad k\in\mathbb{N}. \tag{3.54}$$

For the proof, we introduce

$$\Phi(x,y) := f_\sigma\bigg(\frac{1}{N}\sum_{j=1}^{N-1}\mathcal{W}_\beta(y_j-x)\bigg) \quad \text{for } x\in\mathbb{R}^d,\ y=(y_1,\dots,y_{N-1})\in\mathbb{R}^{(N-1)d}.$$

Then, by definition of $\mathcal{K}_\zeta$,

$$\|\mathrm{D}^k\mathcal{K}_\zeta * \Psi(\cdot,t)\|_\infty \le \sup_{y\in\mathbb{R}^{N-1}}\|\mathcal{W}_\zeta * \mathcal{K}\omega_\zeta * \mathrm{D}^k\Phi(\cdot,y)\|_\infty.$$

We estimate the right-hand side:

$$\begin{aligned}\|\mathcal{W}_\zeta * (\mathcal{K}\omega_\zeta * \mathrm{D}^k\Phi(\cdot,y))\|_\infty &\le \|\mathcal{W}_\zeta\|_1\|\mathcal{K}\omega_\zeta * \mathrm{D}^k\Phi(\cdot,y)\|_\infty \le C\|\mathcal{K}\omega_\zeta * \mathrm{D}^k\Phi(\cdot,y)\|_{W^{1,p}(\mathbb{R}^d)}\\ &\le C\|\mathcal{K} * |\mathrm{D}^k\Phi(\cdot,y)|\|_p + C\|\mathcal{K} * |\mathrm{D}^{k+1}\Phi(\cdot,y)|\|_p\\ &\le C\|\mathrm{D}^k\Phi(\cdot,y)\|_r + C\|\mathrm{D}^{k+1}\Phi(\cdot,y)\|_r,\end{aligned}$$

where we used the Hardy–Littlewood–Sobolev inequality for $r = dp/(d+2ps)$ in the last step. It follows from the Faà di Bruno formula, after an elementary computation, that the last term is estimated according to

$$\begin{aligned}
\|\mathrm{D}^{k+1}\Phi(\cdot,y)\|_r^r &= \int_{\mathbb{R}^d}\bigg|\mathrm{D}^{k+1}\bigg(f_\sigma\bigg(\frac{1}{N}\sum_{j=1}^{N-1}\mathcal{W}_\beta(y_j-x)\bigg)\bigg)\bigg|^r\mathrm{d}x \\
&\le C(k,N)\max_{\ell=1,\dots,k+1}\|f_\sigma^{(\ell)}\|_\infty^r\|\mathrm{D}^k\mathcal{W}_\beta\|_\infty^{kr}\max_{0\le j\le k}\int_{\mathbb{R}^d}|\mathrm{D}^{j+1}\mathcal{W}_\beta(x)|^r\mathrm{d}x \\
&\le C(k,N)\max_{\ell=1,\dots,k+1}\|f_\sigma^{(\ell)}\|_\infty^r\beta^{-(d+k)kr}\beta^{-(d+k+1)r+d} \le C(k,N,\sigma)\beta^{-(d+k)(k+1)r-r},
\end{aligned}$$

since $\|\mathrm{D}^k\mathcal{W}_\beta\|_\infty \le C\beta^{-(d+k)}$ and $\|\mathrm{D}^{j+1}\mathcal{W}_\beta\|_r \le C\beta^{-(d+j+1)+d/r}$. This verifies (3.54). We infer from (3.54) with $k=2$ that

$$I_{11} \le C\beta^{-3d-7}\int_0^t \sup_{0<r<s}\max_{i=1,\dots,N}|(X_i^N-\bar{X}_i^N)(r)|\mathrm{d}s.$$

The term I_{13} is estimated in a similar way, with Ψ replaced by $\bar{\Psi}$:

$$I_{13} \le C\beta^{-3d-7}\int_0^t \sup_{0<r<s}\max_{i=1,\dots,N}|(X_i^N-\bar{X}_i^N)(r)|\mathrm{d}s.$$

The estimate of the remaining term I_{12} is more involved. Since $\mathcal{W}_\beta$ is assumed to be symmetric, we find that

$$\begin{aligned}
I_{12} = \bigg|\int_0^t\int_{\mathbb{R}^d}&\mathcal{K}_\zeta(y)\nabla\bigg\{f_\sigma\bigg(\frac{1}{N}\sum_{j=1,\,j\neq i}^N\mathcal{W}_\beta(X_j^N(s)-\bar{X}_i^N(s)+y)\bigg) \\
&- f_\sigma\bigg(\frac{1}{N}\sum_{j=1,\,j\neq i}^N\mathcal{W}_\beta(\bar{X}_j^N(s)-X_i^N(s)+y)\bigg)\bigg\}\mathrm{d}y\mathrm{d}s\bigg| \\
\le C\int_0^t\int_{\mathbb{R}^d}&\mathcal{K}_\zeta(y)\bigg|f_\sigma'\bigg(\frac{1}{N}\sum_{j\neq i}\mathcal{W}_\beta(X_j^N(s)-\bar{X}_i^N(s)+y)\bigg) \\
&\times\frac{1}{N}\sum_{j\neq i}\nabla\big(\mathcal{W}_\beta(X_j^N(s)-\bar{X}_i^N(s)+y)-\mathcal{W}_\beta(\bar{X}_j^N(s)-X_i^N(s)+y)\big) \\
&+\bigg\{f_\sigma'\bigg(\frac{1}{N}\sum_{j\neq i}\mathcal{W}_\beta(X_j^N(s)-\bar{X}_i^N(s)+y)\bigg)-f_\sigma'\bigg(\frac{1}{N}\sum_{j\neq i}\mathcal{W}_\beta(\bar{X}_j^N(s)-X_i^N(s)+y)\bigg)\bigg\} \\
&\times\frac{1}{N}\sum_{j\neq i}\nabla\mathcal{W}_\beta(\bar{X}_j^N(s)-X_i^N(s)+y)\bigg|\mathrm{d}y\mathrm{d}s \\
\le C\|f_\sigma'\|_\infty&\int_0^t\sup_{0<s<t}\max_{i=1,\dots,N}|(X_i^N-\bar{X}_i^N)(s)|\frac{1}{N}\sum_{j\neq i}\int_{\mathbb{R}^d}\mathcal{K}_\zeta(y)|\mathrm{D}^2\mathcal{W}_\beta(y+\xi_{ij}(s))|\mathrm{d}y\mathrm{d}s \\
&+C\|f_\sigma''\|_\infty\int_0^t\sup_{0<s<t}\max_{i=1,\dots,N}|(X_i^N-\bar{X}_i^N)(s)|\,\|\mathcal{K}_\zeta*\nabla\mathcal{W}_\beta\|_\infty\mathrm{d}s,
\end{aligned}$$

where $\xi_{ij}(s)$ is a random value. We write $\mathcal{K}^1 = \mathcal{K}|_{B_1}$, $\mathcal{K}^2 = \mathcal{K}|_{\mathbb{R}^d\setminus B_1}$ and note that $\widetilde{\mathcal{K}}_\zeta \le \mathcal{K}$ for all $\zeta > 0$. Then

$$\begin{aligned}\int_{\mathbb{R}^d} \mathcal{K}_\zeta(y)|\mathrm{D}^2\mathcal{W}_\beta(y+\xi_{ij}(s))|\mathrm{d}y &\le \int_{B_{1+\zeta}} (\mathcal{K}^1 * \mathcal{W}_\zeta)(y)|\mathrm{D}^2\mathcal{W}_\beta(y+\xi_{ij}(s))|\mathrm{d}y \\ &\quad + \int_{\mathbb{R}^d\setminus B_{1-\zeta}} (\mathcal{K}^2 * \mathcal{W}_\zeta)(y)|\mathrm{D}^2\mathcal{W}_\beta(y+\xi_{ij}(s))|\mathrm{d}y \\ &\le \|\mathcal{K}^1 * \mathcal{W}_\zeta\|_{L^{\theta/(\theta-1)}(B_{1+\zeta})}\|\mathrm{D}^2\mathcal{W}_\beta(\cdot+\xi_{ij}(s))\|_{L^\theta(B_{1+\zeta})} \\ &\quad + \|\mathcal{K}^2 * \mathcal{W}_\zeta\|_\infty \|\mathrm{D}^2\mathcal{W}_\beta(\cdot+\xi_{ij}(s))\|_{L^1(\mathbb{R}^d\setminus B_{1-\zeta})} \\ &\le \|\mathcal{K}^1\|_{L^{\theta/(\theta-1)}(B_1)}\|\mathrm{D}^2\mathcal{W}_\beta(\cdot+\xi_{ij}(s))\|_{L^\theta(B_{1+\zeta})} + \|\mathcal{K}^2\|_\infty\|\mathrm{D}^2\mathcal{W}_\beta(\cdot+\xi_{ij}(s))\|_{L^1(\mathbb{R}^d)} \\ &\le C\big(\|\mathrm{D}^2\mathcal{W}_\beta\|_\infty + \|\mathrm{D}^2\mathcal{W}_\beta\|_1\big) \le C\beta^{-d-2}.\end{aligned}$$

Observe that we did not use the compact support for $\widetilde{\mathcal{K}}_\zeta$ (which depends on ζ), because a negative exponent of ζ at this point would lead to a logarithmic connection between ζ and N in the end, which we wish to avoid.
Furthermore, by the convolution, Sobolev, and Hardy–Littlewood–Sobolev inequalities as well as the fact that $|\widetilde{\mathcal{K}}_\zeta * \nabla\mathcal{W}_\beta| = |(\mathcal{K}w_\zeta) * \mathcal{W}_\zeta * \nabla\mathcal{W}_\beta| \le \mathcal{K} * |\mathcal{W}_\zeta| * |\nabla\mathcal{W}_\beta|$,

$$\begin{aligned}\|\mathcal{K}_\zeta * \nabla\mathcal{W}_\beta\|_\infty &= \|\mathcal{W}_\zeta * \widetilde{\mathcal{K}}_\zeta * \nabla\mathcal{W}_\beta\|_\infty \le \|\widetilde{\mathcal{K}}_\zeta * \nabla\mathcal{W}_\beta\|_\infty \le \|\widetilde{\mathcal{K}}_\zeta * \nabla\mathcal{W}_\beta\|_\infty \\ &\le C\|\widetilde{\mathcal{K}}_\zeta * \nabla\mathcal{W}_\beta\|_{W^{1,p}(\mathbb{R}^d)} \le C(\|\mathcal{K} * |\nabla\mathcal{W}_\beta|\|_p^p + \|\mathcal{K} * |D^2\mathcal{W}_\beta|\|_p^p)^{1/p} \\ &\le C\|\nabla\mathcal{W}_\beta\|_{W^{1,r}(\mathbb{R}^d)} \le C\beta^{-d-2+d/r},\end{aligned}$$

where we recall that $r > d/(2s)$ and we choose $p > d$ satisfying $1/p = 2s/d - 1/r$. The previous two estimates lead to

$$I_{12} \le C(\sigma)\beta^{-d-2}\int_0^t \sup_{0<r<s}\max_{i=1,\dots,N}|(X_i^N - \bar{X}_i^N)(r)|\mathrm{d}s.$$

We summarize:

$$I_1 \le C(\sigma)\beta^{-3d-7}\int_0^t \sup_{0<r<s}\max_{i=1,\dots,N}|(X_i^N - \bar{X}_i^N)(r)|\mathrm{d}s. \tag{3.55}$$

Step 2: Estimate of I_2. We take the expectation of I_2 and use the mean-value theorem:

$$\begin{aligned}\mathbb{E}(I_2) &= \int_0^t \mathbb{E}\bigg|\int_{\mathbb{R}^d} \nabla\mathcal{K}_\zeta(y)\bigg\{f_\sigma\bigg(\frac{1}{N}\sum_{j\ne i}\mathcal{W}_\beta(\bar{X}_j^N(s) - \bar{X}_i^N(s) + y)\bigg) \\ &\quad - f_\sigma\big(\mathcal{W}_\beta * \rho(\bar{X}_i^N(s) - y, s)\big)\bigg\}\mathrm{d}y\bigg|\mathrm{d}s \\ &\le N^{-1}\|f_\sigma'\|_\infty\|\widetilde{\mathcal{K}}_\zeta * \nabla\mathcal{W}_\zeta\|_1 \int_0^t \sup_{y\in\mathbb{R}^d}\mathbb{E}\bigg(\sum_{j\ne i}|b_{ij}(y,s)|\bigg)\mathrm{d}s,\end{aligned} \tag{3.56}$$

where

$$b_{ij}(y,s) = \mathcal{W}_\beta(\bar{X}_j^N(s) - \bar{X}_i^N(s) + y) - \frac{N}{N-1}\mathcal{W}_\beta * \rho(\bar{X}_i^N(s) - y, s).$$

We deduce from $\|\nabla \mathcal{W}_\zeta\|_{L^1(\mathbb{R}^d)} \le C\zeta^{-1}$ that

$$\|\widetilde{\mathcal{K}}_\zeta * \nabla \mathcal{W}_\zeta\|_1 \le C\zeta^{-1}\|\widetilde{\mathcal{K}}_\zeta\|_1 \le C\zeta^{-2s-1},$$

due to the compact support of $\widetilde{\mathcal{K}}_\zeta(x) = |x|^{2s-d}\omega_\zeta(x) \le C|x|^{2s-d}1_{|x|\le 2\zeta^{-1}}$ and

$$\int_{\{|x|<2/\zeta\}} |x|^{2s-d}\mathrm{d}x = \int_{\{|y|<2\}} \zeta^{-d}|y/\zeta|^{2s-d}\mathrm{d}y = C\zeta^{-2s}.$$

We claim that $\mathbb{E}(\sum_{j\ne i}|b_{ij}(y,s)|) \le C(\sigma)\beta^{-d/2}N^{1/2}$ for all $y\in\mathbb{R}^d$. To show the claim, we compute the expectation $\mathbb{E}[(\sum_{j\ne i} b_{ij}(y,s))^2]$. We estimate first the terms with $k\ne j$ (omitting the argument (y,s) to simplify the notation). Then an elementary but tedious computation leads to

$$\begin{aligned}
\mathbb{E}(b_{ji}b_{ki}) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\bigg(\mathcal{W}_\beta(x_j - x_i + y) - \frac{N}{N-1}\mathcal{W}_\beta * \rho(x_i - y)\bigg) \\
&\quad \times \bigg(\mathcal{W}_\beta(x_k - x_i + y) - \frac{N}{N-1}\mathcal{W}_\beta * \rho(x_i - y)\bigg)\rho(x_i)\rho(x_j)\rho(x_k)\mathrm{d}x_i\mathrm{d}x_j\mathrm{d}x_k \\
&= \int_{\mathbb{R}^d}\bigg(\mathcal{W}_\beta * \rho(x_i - y) - \frac{N}{N-1}\mathcal{W}_\beta * \rho(x_i - y)\bigg)^2\rho(x_i)\mathrm{d}x_i \\
&\le N^{-2}\|\rho\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\|\mathcal{W}_\beta * \rho\|^2_{L^\infty(0,T;L^2(\mathbb{R}^d))} \\
&\le C(\sigma)N^{-2}\|\mathcal{W}_\beta\|_1^2 \le C(\sigma)N^{-2}.
\end{aligned}$$

The diagonal terms contribute in the following way:

$$\begin{aligned}
\mathbb{E}(b_{ji}^2) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\bigg(\mathcal{W}_\beta(x_j - x_i + y) - \frac{N}{N-1}\mathcal{W}_\beta * \rho(x_i - y)\bigg)^2\rho(x_i)\rho(x_j)\mathrm{d}x_i\mathrm{d}x_j \\
&= \int_{\mathbb{R}^d}\bigg((\mathcal{W}_\beta^2 * \rho)(x_i - y) - \frac{2N}{N-1}(\mathcal{W}_\beta * \rho)(x_i - y)^2 \\
&\quad + \frac{N^2}{(N-1)^2}(\mathcal{W}_\beta * \rho)(x_i - y)^2\bigg)\rho(x_i)\mathrm{d}x_i \\
&\le C(\sigma)\big(\|\mathcal{W}_\beta^2 * \rho\|_{L^\infty(0,T;L^1(\mathbb{R}^d))} + \|\mathcal{W}_\beta * \rho\|^2_{L^\infty(0,T;L^2(\mathbb{R}^d))}\big) \le C(\sigma)\beta^{-d},
\end{aligned}$$

since $\|\mathcal{W}_\beta^2 * \rho\|_2 \le \|\mathcal{W}_\beta^2\|_1\|\rho\|_2 \le C\|\mathcal{W}_\beta\|_2^2 \le \beta^{-d}C$. This shows that

$$\mathbb{E}\bigg(\sum_{j\ne i}|b_{ji}(y,s)|\bigg) \le \bigg(\mathbb{E}\bigg[\sum_{j\ne i} b_{ji}(y,s)\bigg]^2\bigg)^{1/2} \le C(\sigma)\beta^{-d/2}N^{1/2}.$$

We infer that (3.56) becomes

$$I_2 \le C(\sigma)\zeta^{-2s-1}\beta^{-d/2}N^{-1/2}. \tag{3.57}$$

Step 3: End of the proof. We insert (3.55) and (3.57) into (3.53) to infer that

$$E_1(t) := \mathbb{E}\bigg(\sup_{0<s<t}\max_{i=1,\dots,N}|(X_i^N - \bar{X}_i^N)(s)|\bigg)$$

$$\leq C(\sigma)\beta^{-3d-7}\int_0^t E_1(s)\mathrm{d}s + C(\sigma)\zeta^{-2s-1}\beta^{-d/2}N^{-1/2}.$$

By Gronwall's lemma,

$$E_1(t) \leq C(\sigma)\zeta^{-2s-1}\beta^{-d/2}N^{-1/2}\exp\big(C(\sigma)\beta^{-3d-7}T\big), \quad 0 \leq t \leq T.$$

We choose $\varepsilon = \widetilde{\delta}/(C(\sigma)T)$ for some arbitrary $\widetilde{\delta} \in (0,1/4)$. Then, since by assumption, $\beta^{-d/2} \leq \beta^{-3d-7} \leq \varepsilon \log N$ and $\zeta^{-2s-1} \leq C_1 N^{1/4}$, we find that

$$E_1(t) \leq C(\sigma)C_1\varepsilon\log(N)N^{-1/4}\exp\big(C(\sigma)T\varepsilon\log N\big) = \frac{C_1\widetilde{\delta}}{T}\log(N)N^{-1/4+\widetilde{\delta}},$$

proving the result. □

3.4.3 Estimate of $\bar{X}_i^N - \widehat{X}_i^N$

Next, we compute the expectation of $\bar{X}_i^N - \widehat{X}_i^N$ by estimating the difference between $\nabla\mathcal{K}_\zeta$ and $\nabla\mathcal{K}$ as well as the difference between $\mathcal{W}_\beta * \rho(\widehat{X}_i^N)$ and $\rho_\sigma(\widehat{X}_i^N)$. The estimate depends on β and ζ.

Lemma 3.18. *Let $\bar{X}_i^N$ and $\widehat{X}_i^N$ be the solutions to (3.4) and (3.6), respectively. Then there exists a constant $C > 0$, depending on σ, such that*

$$\mathbb{E}\bigg(\sup_{0<t<T}\max_{i=1,\ldots,N}|(\bar{X}_i^N - \widehat{X}_i^N)(t)|\bigg) \leq C(\beta + \zeta^a),$$

where $a := \min\{1, d-2s\}$.

Proof. We compute the difference

$$\begin{aligned}|(\bar{X}_i^N - \widehat{X}_i^N)(t)| &= \bigg|\int_0^t \big(\nabla\mathcal{K}_\zeta * f_\sigma(\mathcal{W}_\beta * \rho(\bar{X}_i^N(s),s)) - \nabla\mathcal{K} * f_\sigma(\rho_\sigma(\widehat{X}_i^N(s),s))\big)\mathrm{d}s\bigg| \\ &\leq J_1 + J_2 + J_3,\end{aligned}$$

where $\rho := \rho_{\sigma,\beta,\zeta}$, the convolution is taken with respect to x_i, and

$$\begin{aligned}J_1 &= \bigg|\int_0^t \nabla\mathcal{K}_\zeta * \big(f_\sigma(\mathcal{W}_\beta * \rho(\bar{X}_i^N(s),s)) - f_\sigma(\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s),s))\big)\mathrm{d}s\bigg|, \\ J_2 &= \bigg|\int_0^t \nabla\mathcal{K}_\zeta * \big(f_\sigma(\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s),s)) - f_\sigma(\rho_\sigma(\widehat{X}_i^N(s),s))\big)\mathrm{d}s\bigg|, \\ J_3 &= \bigg|\int_0^t \nabla(\mathcal{K}_\zeta - \mathcal{K}) * f_\sigma(\rho_\sigma(\widehat{X}_i^N(s),s))\mathrm{d}s\bigg|.\end{aligned}$$

Step 1: Estimate of J_1. We write $\nabla\mathcal{K}_\zeta * f_\sigma(\cdots) = \mathcal{K}_\zeta * \nabla f_\sigma$ and add and subtract the expression $f'_\sigma(\mathcal{W}_\beta * \rho(\bar{X}_i^N - y))\nabla\mathcal{W}_\beta * \rho(\widehat{X}_i^N - y)$:

$$J_1 = \int_0^t\int_{\mathbb{R}^d}\mathcal{K}_\zeta(y)\Big(f'_\sigma(\mathcal{W}_\beta * \rho(\bar{X}_i^N(s) - y))\nabla\mathcal{W}_\beta * \big[\rho(\bar{X}_i^N(s) - y) - \rho(\widehat{X}_i^N(s) - y)\big]$$

$$
\begin{aligned}
&\quad - \big[f_\sigma'(\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s)-y)) - f_\sigma'(\mathcal{W}_\beta * \rho(\bar{X}_i^N(s)-y))\big]\nabla\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s)-y)\Big)\mathrm{d}y\mathrm{d}s \\
&\le \|f_\sigma'\|_\infty \int_0^t \int_{\mathbb{R}^d} \Big|\mathcal{K}_\zeta(y)\nabla\mathcal{W}_\beta * \big(\rho(\bar{X}_i^N(s)-y) - \rho(\widehat{X}_i^N(s)-y)\big)\Big|\mathrm{d}y\mathrm{d}s \\
&\quad + \|f_\sigma''\|_\infty \|\nabla\mathcal{W}_\beta * \rho\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\
&\quad \times \int_0^t \int_{\mathbb{R}^d} \big|\mathcal{K}_\zeta(y)\mathcal{W}_\beta * \big(\rho(\widehat{X}_i^N(s)-y) - \rho(\bar{X}_i^N(s)-y)\big)\big|\mathrm{d}y\mathrm{d}s.
\end{aligned}
$$

By the mean-value theorem and using $\|\mathcal{W}_\beta\|_1 = 1$, we obtain for some random variable $\xi_{ij}(s)$,

$$
\begin{aligned}
J_1 &\le \|f_\sigma\|_{W^{2,\infty}(\mathbb{R})}\|\nabla\rho\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \int_0^t \sup_{0<r<s}\sup_{i=1,\ldots,N} |(\bar{X}_i^N - \widehat{X}_i^N)(r)| \\
&\qquad \times \int_{\mathbb{R}^d} \sum_{k=1}^2 \big|\mathcal{K}_\zeta(y)\mathrm{D}^k\mathcal{W}_\beta * \rho(y+\xi_{ij}(s),s)\big|\mathrm{d}y\mathrm{d}s.
\end{aligned} \tag{3.58}
$$

We need to estimate the last integral. For this, we write for $k=1,2$

$$
\begin{aligned}
&\int_{\mathbb{R}^d} \big|\mathcal{K}_\zeta(y)\mathrm{D}^k\mathcal{W}_\beta * \rho(y+\xi_{ij}(s),s)\big|\mathrm{d}y \le K_1^k + K_2^k, \quad \text{where} \\
&K_1^k := \int_{B_{1+\zeta}} \big|\mathcal{K}^1 * \mathcal{W}_\zeta(y)\mathrm{D}^k\mathcal{W}_\beta * \rho(y+\xi_{ij}(s),s)\big|\mathrm{d}y, \\
&K_2^k := \int_{\mathbb{R}^d\setminus B_{1-\zeta}} \big|\mathcal{K}^2 * \mathcal{W}_\zeta(y)\mathrm{D}^k\mathcal{W}_\beta * \rho(y+\xi_{ij}(s),s)\big|\mathrm{d}y,
\end{aligned}
$$

where $\mathcal{K}^1 = \mathcal{K}|_{B_1}$ and $\mathcal{K}^2 = \mathcal{K}|_{\mathbb{R}^d\setminus B_1}$. Note that $\widetilde{\mathcal{K}}_\zeta \le \mathcal{K}$. A similar argument as for the estimate of I_{12} in the proof of Lemma 3.17 shows that for $\theta > \max\{d/(2s), d\}$,

$$
\begin{aligned}
K_1^k + K_2^k &\le C\big(\|\mathrm{D}^k\mathcal{W}_\beta * \rho\|_{L^\infty(0,T;L^\theta(\mathbb{R}^d))} + \|\mathrm{D}^k\mathcal{W}_\beta * \rho\|_{L^\infty(0,T;L^1(\mathbb{R}^d))}\big) \\
&\le C\big(\|\mathrm{D}^k\rho\|_{L^\infty(0,T;L^\theta(\mathbb{R}^d))} + \|\mathrm{D}^k\rho\|_{L^\infty(0,T;L^1(\mathbb{R}^d))}\big) \le C(\sigma),
\end{aligned}
$$

where we used Proposition 3.14 ((3.39) and (3.41)) with $p=\theta$ in the last inequality. We conclude from (3.58) that

$$
J_1 \le C(\sigma)\int_0^t \sup_{0<r<s}\max_{i=1,\ldots,N} |(\bar{X}_i^N - \widehat{X}_i^N)(r)|\mathrm{d}s. \tag{3.59}
$$

Step 2: Estimate of J_2. We treat the two cases $s<1/2$ and $s\ge 1/2$ separately. Let first $s \ge 1/2$. Then

$$
\begin{aligned}
J_2 &= \bigg|\int_0^t \nabla\widetilde{\mathcal{K}}_\zeta * \mathcal{W}_\zeta * \big(f_\sigma(\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s),s)) - f_\sigma(\rho_\sigma(\widehat{X}_i^N(s),s))\big)\mathrm{d}s\bigg| \\
&\le T\|\nabla\widetilde{\mathcal{K}}_\zeta * (f_\sigma(\mathcal{W}_\beta * \rho) - f_\sigma(\rho_\sigma))\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}.
\end{aligned}
$$

Recalling the definition of $\widetilde{\mathcal{K}}_\zeta = \mathcal{K}\omega_\zeta$ in (3.11) and writing $\nabla\widetilde{\mathcal{K}}_\zeta * u = \nabla\mathcal{K} * u - [(1-\omega_\zeta)\nabla\mathcal{K}] * u + [\mathcal{K}\nabla\omega_\zeta] * u$ for $u = f_\sigma(\mathcal{W}_\beta * \rho) - f_\sigma(\rho_\sigma)$, we find that

$$J_2 \le C(T)\big(\|\nabla\mathcal{K} * u\|_{L^\infty([0,T];L^\infty(\mathbb{R}^d))} + \|[(1-\omega_\zeta)\nabla\mathcal{K}] * u\|_{L^\infty([0,T];L^\infty(\mathbb{R}^d))} + \|[\mathcal{K}\nabla\omega_\zeta] * u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\big). \tag{3.60}$$

We estimate the right-hand side term by term. Because of

$$\nabla\mathcal{K} * v = \begin{cases} \nabla(-\Delta)^{-1/2}v & \text{for } s = 1/2 \\ (\nabla\mathcal{K}) * v & \text{for } s > 1/2, \end{cases}$$

we use Sobolev's embedding $W^{1,p}(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d)$ for any $p > d$ and then the boundedness of the Riesz operator $\nabla(-\Delta)^{-1/2} : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ [108, Chapter IV, §3.1] in case $s = 1/2$ or the Hardy–Littlewood–Sobolev inequality for $\widetilde{\alpha} = \alpha - 1/2 > 0$ (see Lemma 3.22) in case $s > 1/2$ to control the first norm in (3.60) by

$$\begin{aligned}\|\nabla\mathcal{K} * u\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} &\le C\bigg(\|\nabla\mathcal{K} * u\|_{L^\infty(0,T;L^p(\mathbb{R}^d))} + \sum_{j=1}^d \|\nabla\mathcal{K} * D^j u\|_{L^\infty(0,T;L^p(\mathbb{R}^d))}\bigg) \\ &\le C\|u\|_{L^\infty(0,T;W^{1,r}(\mathbb{R}^d))} = C\|f_\sigma(\mathcal{W}_\beta * \rho) - f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;W^{1,r}(\mathbb{R}^d))},\end{aligned}$$

where $r = p$ in case $s = 1/2$ and $r = pd/(d+2s-1)$ in case $s > 1/2$. Choosing $p > d+(2s-1)$ guarantees that $r > d$ always holds.
For the second norm in (3.60), Hölder's inequality yields for $q > d$ and $1/q + 1/q' = 1$, for every $t > 0$,

$$\begin{aligned}\|[(1-\omega_\zeta)\nabla\mathcal{K}] * u(t)\|_{L^\infty(\mathbb{R}^d)} &\le \|1-\omega_\zeta\|_{L^\infty(\mathbb{R}^d)}\|\nabla\mathcal{K}\|_{L^{q'}(\{|x|>2\zeta^{-1}\})}\|u(t)\|_{L^q(\mathbb{R}^d)} \\ &\le \|\nabla\mathcal{K}\|_{L^{q'}(\{|x|>2\zeta^{-1}\})}\|u(t)\|_{L^q(\mathbb{R}^d)},\end{aligned}$$

which can be bounded by $C\zeta^{1-2s+d/q}\|u(t)\|_{L^q(\mathbb{R}^d)}$, since

$$\begin{aligned}\|\nabla\mathcal{K}\|^{q'}_{L^{q'}(\{|x|>2\zeta^{-1}\})} &\le C\int_{\{|x|>2\zeta^{-1}\}} |x|^{(2s-d-1)q'}\mathrm{d}x = C\zeta^{-d}\int_{\{|y|>2\}} |y/\zeta|^{(2s-d-1)q'}\mathrm{d}y \\ &\le C\zeta^{-d+(1+d-2s)q'}.\end{aligned}$$

By similar arguments and the fact that $\|\nabla\omega_\zeta\|_{L^\infty} \le C\zeta$, we find that

$$\|\mathcal{K}\nabla\omega_\zeta\|_{L^{q'}(\{|x|<2\zeta^{-1}\})} \le C\zeta^{1+d-2s-d/q'},$$

and hence, using $q' = q/(q-1)$, we conclude for the second and third term in (3.60) that

$$\|[(1-\omega_\zeta)\nabla\mathcal{K}] * u(t)\|_{L^\infty(\mathbb{R}^d)} + \|[\mathcal{K}\nabla\omega_\zeta] * u(t)\|_{L^\infty(\mathbb{R}^d)} \le C\zeta^{1-2s+d/q}\|u(t)\|_{L^q(\mathbb{R}^d)}.$$

The choice $d < q \le d/(2s-1)$ guarantees on the one hand that $q > d$ and on the other hand that the exponent $1-2s+d/q$ is strictly positive (which allows us to use the property $\zeta^{1-2s+d/q} < 1$).

Using these estimates in (3.60), we arrive (for $s \geq 1/2$) at

$$J_2 \leq C(T)\big(\|f_\sigma(\mathcal{W}_\beta * \rho) - f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;W^{1,r}(\mathbb{R}^d))} + \|f_\sigma(\mathcal{W}_\beta * \rho) - f_\sigma(\rho_\sigma)\|_{L^\infty(0,T;L^q(\mathbb{R}^d))}\big),$$

where we recall that $r, q > d$. These norms can be estimated by $\|f_\sigma(\mathcal{W}_\beta * \rho(t)) - f_\sigma(\rho_\sigma(t))\|_{L^q(\mathbb{R}^d)} \leq \|f_\sigma'\|_\infty \|\mathcal{W}_\beta * \rho(t) - \rho_\sigma(t)\|_{L^q(\mathbb{R}^d)}$ and

$$\begin{aligned}\|\nabla(f_\sigma(\mathcal{W}_\beta * \rho) - f_\sigma(\rho_\sigma))(t)\|_{L^r(\mathbb{R}^d)} &\leq \|f_\sigma'\|_\infty \|(\mathcal{W}_\beta * \nabla\rho - \nabla\rho_\sigma)(t)\|_{L^r(\mathbb{R}^d)} \\ &\quad + \|f_\sigma''\|_\infty \|(\mathcal{W}_\beta * \rho - \rho_\sigma)(t)\|_{L^r(\mathbb{R}^d)} \|\nabla\rho_\sigma(t)\|_{L^\infty(\mathbb{R}^d)}.\end{aligned}$$

The $L^\infty(\mathbb{R}^d \times (0,T))$ bound for $\nabla\rho_\sigma$ from Lemma 3.11 and the definition of f_σ finally show for $s \geq 1/2$ and $r, q > d$ that

$$J_2 \leq C(\sigma,T)\big(\|\mathcal{W}_\beta * \rho - \rho_\sigma\|_{L^\infty(0,T;W^{1,r}(\mathbb{R}^d))} + \|\mathcal{W}_\beta * \rho - \rho_\sigma\|_{L^\infty(0,T;L^q(\mathbb{R}^d))}\big). \tag{3.61}$$

Now, let $s < 1/2$. In this case, we cannot estimate $\nabla\mathcal{K}$ and put the gradient to the second factor of the convolution. Adding and subtracting an appropriate expression as in Step 1, using the embedding $W^{1,p}(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d)$ for $p > d$, the estimate $\mathcal{K}_\zeta \leq \mathcal{K}$, and the Hardy–Littlewood–Sobolev inequality, we find that

$$\begin{aligned}J_2 &= \bigg|\int_0^t \int_{\mathbb{R}^d} \mathcal{K}_\zeta(y)\Big(\big(f_\sigma'(\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s) - y)) - f_\sigma'(\rho_\sigma(\widehat{X}_i^N(s) - y))\big)\nabla\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s) - y) \\ &\qquad - f_\sigma'(\rho_\sigma(\widehat{X}_i^N(s) - y))\big(\nabla\rho_\sigma(\widehat{X}_i^N(s) - y) - \nabla\mathcal{W}_\beta * \rho(\widehat{X}_i^N(s) - y)\big)\Big)\mathrm{d}y\mathrm{d}s\bigg| \\ &\leq \|f_\sigma''\|_\infty \|\mathcal{W}_\beta * \nabla\rho\|_\infty \int_0^t \int_{\mathbb{R}^d} \mathcal{K}_\zeta(y)\big|\rho_\sigma(\widehat{X}_i^N(s) - y) - \mathcal{W}_\beta * \rho(\widehat{X}_i^N(s) - y)\big|\mathrm{d}y\mathrm{d}s \\ &\quad + \|f_\sigma'\|_\infty \int_0^t \int_{\mathbb{R}^d} \mathcal{K}_\zeta(y)\big|\nabla\rho_\sigma(\widehat{X}_i^N(s) - y) - \mathcal{W}_\beta * \nabla\rho(\widehat{X}_i^N(s) - y)\big|\mathrm{d}y\mathrm{d}s \\ &\leq \max\{\|\nabla\rho\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}, 1\}\|f_\sigma'\|_{W^{1,\infty}} T\big(\|\mathcal{K} * |(\mathcal{W}_\beta * \rho - \rho_\sigma)|\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\ &\quad + \|\mathcal{K} * |(\mathcal{W}_\beta * \nabla\rho - \nabla\rho_\sigma)|\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\big) \\ &\leq C(\sigma,T)\big(\|\nabla\rho\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} + 1\big) \sum_{|\alpha|\leq 2} \|\mathcal{W}_\beta * \mathrm{D}^\alpha\rho - \mathrm{D}^\alpha\rho_\sigma\|_{L^\infty(0,T;L^r(\mathbb{R}^d))},\end{aligned}$$

where $r > d$ is such that $1/r = 2s/d + 1/p$ (this is needed for the Hardy–Littlewood–Sobolev inequality) and $p > d$ (because of Sobolev's embedding). Note that $r > d$ can be only guaranteed if $s < 1/2$. Together with the fact that $\|\nabla\rho\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \leq C(\sigma)$ (choose $q > d$ in (3.41) and use Sobolev's embedding), this shows that for $s < 1/2$,

$$J_2 \leq C(\sigma,T) \sum_{|\alpha|\leq 2} \|\mathcal{W}_\beta * \mathrm{D}^\alpha\rho - \mathrm{D}^\alpha\rho_\sigma\|_{L^\infty(0,T;L^r(\mathbb{R}^d))}. \tag{3.62}$$

It follows from estimate (3.38) and Lemma 3.21 in Appendix 3.A for $p > d$ that

$$\|(\mathcal{W}_\beta * D^\alpha\rho - D^\alpha\rho_\sigma)(t)\|_{L^p(\mathbb{R}^d)} \leq C\big(\|D^\alpha\nabla\rho\|_{L^p(\mathbb{R}^d)}\beta + \beta + \zeta^a\big) \leq C(\sigma,T)(\beta + \zeta^a),$$

where we used the $L^\infty(0,T;W^{3,p}(\mathbb{R}^d))$ estimate for $\rho = \rho_{\sigma,\beta,\zeta}$ in (3.41). Then we deduce from estimates (3.61) and (3.62) that for all $0 < s < 1$,

$$J_2 \le C(\sigma,T)(\beta + \zeta^a),$$

where we recall that $a = \min\{1, d-2s\}$.
Step 3: Estimate of J_3 and end of the proof. Arguing similarly as in Section 3.3.3, we have

$$\|(\mathcal{K}_\zeta - \mathcal{K}) * \nabla\rho_\sigma\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\zeta^a\big(\|\mathrm{D}^2\rho_\sigma\|_{L^\infty(0,T;L^p(\mathbb{R}^d))} + \|\nabla\rho_\sigma\|_{L^\infty(0,T;L^1(\mathbb{R}^d))}\big).$$

This implies that

$$J_3 \le \|f'_\sigma\|_\infty \|(\mathcal{K}_\zeta - \mathcal{K}) * \nabla\rho_\sigma\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C(\sigma)\zeta^a. \tag{3.63}$$

Taking the expectation, we infer from (3.59)–(3.63) that

$$E_2(t) := \mathbb{E}\bigg(\sup_{0<s<t}\max_{i=1,\ldots,N}|(\bar{X}_i^N - \widehat{X}_i^N)(s)|\bigg) \le C(\sigma)(\beta+\zeta^a) + C(\sigma)\int_0^t E_2(s)\mathrm{d}s,$$

An application of Gronwall's lemma gives the result. □

3.4.4 Proof of Theorem 3.2 and Proposition 3.3

Lemmas 3.17 and 3.18 show that

$$\mathbb{E}\bigg(\sup_{0<s<T}\max_{i=1,\ldots,N}|(X_i^N - \widehat{X}_i^N)(s)|\bigg) \le C(N^{-1/4+\delta} + \beta + \zeta^{\min\{1,d-2s\}}),$$

and this expression converges to zero as $N \to \infty$ and $(\beta,\zeta) \to 0$ under the conditions stated in Theorem 3.3. This result implies the convergence in probability of the k-tuple $(X_1^N, \ldots, X_k^N)$ to $(\widehat{X}_1^N, \ldots, \widehat{X}_k^N)$. Since convergence in probability implies convergence in distribution, we obtain

$$\lim_{N\to\infty,\,(\beta,\zeta)\to 0} \mathrm{P}^k_{N,\beta,\sigma}(t) = \mathrm{P}_\sigma^{\otimes k}(t) \quad \text{locally uniform in time},$$

where $\mathrm{P}^k_{N,\beta,\sigma}(t)$ and $\mathrm{P}_\sigma^{\otimes k}(t)$ denote the joint distributions of $(X_1^N, \ldots, X_k^N)(t)$ and $(\widehat{X}_1^N, \ldots, \widehat{X}_k^N)(t)$, respectively. By Section 3.4.1, $\mathrm{P}_\sigma(t)$ is absolutely continuous with the density function $\rho_\sigma(t)$. Using the test function $\phi = 1_{(-\infty,x]^d}$ in Corollary 3.13, we have, up to a subsequence,

$$\mathrm{P}_\sigma(t,(-\infty,x]^d) = \int_{(-\infty,x]^d}\rho_\sigma(y,t)\mathrm{d}y \to \int_{(-\infty,x]^d}\rho(y,t)\mathrm{d}y =: \mathrm{P}(t,(-\infty,x]^d)$$

locally uniformly for $t > 0$. Since the convergence also holds for the initial condition, the result is shown.

3.A Auxiliary results

We recall some known results. The following result is proved in [11, Theorem 4.33].

Lemma 3.19 (Young's convolution inequality). *Let* $1 \le p,\ r \le \infty$, $u \in L^p(\mathbb{R}^d)$, $v \in L^q(\mathbb{R}^d)$, *and* $1/p + 1/q = 1 + 1/r$. *Then* $u * v \in L^r(\mathbb{R}^d)$ *and*

$$\|u * v\|_r \le \|u\|_p \|v\|_q.$$

The following lemma slightly extends [101, Lemma 7.3] from the L^2 to the L^p setting.

Lemma 3.20. *Let* $p \ge 2$ *and* $T > 0$. *Then the following embedding is continuous:*

$$L^p(0,T;W^{1,p}(\mathbb{R}^d)) \cap W^{1,p}(0,T;W^{-1,p}(\mathbb{R}^d)) \hookrightarrow C^0([0,T];L^p(\mathbb{R}^d)).$$

Proof. Let $u \in L^p(0,T;W^{1,p}(\mathbb{R}^d)) \cap W^{1,p}(0,T;W^{-1,p}(\mathbb{R}^d))$ and $0 \le t_1 \le t_2 \le T$. Then

$$\begin{aligned} \left| \int_{\mathbb{R}^d} |u(t_2)|^p \mathrm{d}x - \int_{\mathbb{R}^d} |u(t_1)|^p \mathrm{d}x \right| &= \left| \int_{t_1}^{t_2} \langle \partial_t u, p|u|^{p-2}u \rangle \mathrm{d}t \right| \\ &\le p\|\partial_t u\|_{L^p(t_1,t_2;W^{-1,p}(\mathbb{R}^d))} \||u|^{p-2}u\|_{L^{p'}(t_1,t_2;W^{1,p'}(\mathbb{R}^d))}, \end{aligned} \tag{3.64}$$

where $p' = p/(p-1)$. Direct computations using Young's inequality lead to

$$\begin{aligned} \||u|^{p-2}u\|^{p'}_{L^{p'}(t_1,t_2;W^{1,p'}(\mathbb{R}^d))} &= C\int_{t_1}^{t_2} \int_{\mathbb{R}^d} \left(|u|^p + |u|^{p'(p-2)}|\nabla u|^{p'}\right) \mathrm{d}x\mathrm{d}t \\ &\le C\int_{t_1}^{t_2} \|u(t)\|^p_{W^{1,p}(\mathbb{R}^d)} \mathrm{d}t. \end{aligned}$$

We infer from (3.64) and the continuity of the integrals with respect to the time integration boundaries that $t \mapsto \|u(t)\|_p$ is continuous and

$$\sup_{0<t<T} \|u(t)\|_p \le \|u(0)\|_p + C\|\partial_t u\|_{L^p(t_1,t_2;W^{-1,p}(\mathbb{R}^d))} + C\|u\|_{L^p(0,T;W^{1,p}(\mathbb{R}^d))}. \tag{3.65}$$

Next, let $t \in [0,T]$ be arbitrary and let $\tau_n \to 0$ as $n \to \infty$ such that $t + \tau_n \in [0,T]$. Estimate (3.65) implies that $(u(t+\tau_n))_{n\in\mathbb{N}}$ is bounded in $L^p(\mathbb{R}^d)$. Thus, there exists a subsequence $(\tau_{n'})$ of (τ_n) such that $u(t+\tau_{n'}) \rightharpoonup v(t)$ weakly in $L^p(\mathbb{R}^d)$ as $n' \to \infty$ for some $v(t) \in L^p(\mathbb{R}^d)$. We can show, using estimate (3.65) and dominated convergence for the integral

$$\int_0^T \int_{\mathbb{R}^d} (u(t+\tau_{n'},x) - v(t,x))\phi(t,x)\mathrm{d}x \quad \text{for } \phi \in C_0^\infty(\mathbb{R}^d \times (0,T))$$

that in the limit $n' \to \infty$

$$\int_0^T \int_{\mathbb{R}^d} (u(t,x) - v(t,x))\phi(t,x)\mathrm{d}x = 0,$$

which yields $v(t) = u(t)$.
Moreover, since $t \mapsto \|u(t)\|_p$ is continuous, we have $\|u(t+\tau_{n'})\|_p \to \|u(t)\|_p$. Since $L^p(\mathbb{R}^d)$ is uniformly convex, we deduce from [11, Prop. 3.32] that $u(t+\tau_{n'}) \to u(t)$ strongly in $L^p(\mathbb{R}^d)$. Since the limit is unique, the whole sequence converges. Together with (3.65), this concludes the proof. □

Let $\mathcal{W}_1 \in C_0^\infty(\mathbb{R}^d)$ be nonnegative with $\int_{\mathbb{R}^d} \mathcal{W}_1(x)\mathrm{d}x = 1$ and define $\mathcal{W}_\beta(x) = \beta^{-d}\mathcal{W}_1(\beta^{-1}x)$ for $x \in \mathbb{R}^d$ and $\beta > 0$.

Lemma 3.21. *Let $1 \le p < \infty$ and $u \in W^{1,p}(\mathbb{R}^d)$. Then*

$$\|\mathcal{W}_\beta * u - u\|_p \le C\beta\|\nabla u\|_p.$$

Proof. We use Hölder's inequality and the fact that $\|\mathcal{W}_\beta\|_{L^1(\mathbb{R}^d)} = 1$ to find that

$$\begin{aligned}
\|\mathcal{W}_\beta * u - u\|_p^p &= \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\mathcal{W}_\beta(x-y)(u(x)-u(y))\mathrm{d}y\right|^p\mathrm{d}x \\
&\le \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\mathcal{W}_\beta(x-y)\mathrm{d}y\right)^{p-1}\left(\int_{\mathbb{R}^d}\mathcal{W}_\beta(x-y)|u(x)-u(y)|^p\mathrm{d}y\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathcal{W}_\beta(z)|z|^p\frac{|u(y+z)-u(y)|^p}{|z|^p}\mathrm{d}y\mathrm{d}z \\
&\le \|\nabla u\|_p^p\int_{\mathbb{R}^d}\mathcal{W}_\beta(z)|z|^p\mathrm{d}z \le C\beta^p\|\nabla u\|_p^p,
\end{aligned}$$

which shows the lemma. □

3.B Fractional Laplacian

We recall that the fractional Laplacian $(-\Delta)^s$ for $0 < s < 1$ can be written as the pointwise formula

$$(-\Delta)^s u(x) = c_{d,s}\int_{\mathbb{R}^d}\frac{u(x)-u(y)}{|x-y|^{d+2s}}\mathrm{d}y, \quad \text{where } c_{d,s} = \frac{4^s\Gamma(d/2+s)}{\pi^{d/2}|\Gamma(-s)|}, \tag{3.66}$$

$u \in H^s(\mathbb{R}^d)$, and the integral is understood as principal value if $1/2 \le s < 1$ [110, Theorem 2]. The inverse fractional Laplacian $(-\Delta)^{-s}$ is defined in (3.2). The following lemma can be found in [108, Chapter V, Section 1.2].

Lemma 3.22 (Hardy–Littlewood–Sobolev inequality)**.** *Let $0 < s < 1$ and $1 < p < \infty$. Then there exists a constant $C > 0$ such that for all $u \in L^p(\mathbb{R}^d)$,*

$$\|(-\Delta)^{-s}u\|_q \le C\|u\|_p, \quad \textit{where } \frac{1}{p} = \frac{1}{q} + \frac{2s}{d}.$$

Applying Hölder's and then Hardy–Littlewood–Sobolev's inequality gives the following result.

Lemma 3.23. *Let $0 < s < 1$ and $1 \le p < q < \infty$. Then there exists $C > 0$ such that for all $u \in L^q(\mathbb{R}^d)$, $v \in L^r(\mathbb{R}^d)$,*

$$\|u(-\Delta)^{-s}v\|_p \le C\|u\|_q\|v\|_r, \quad \frac{1}{q} + \frac{1}{r} = \frac{1}{p} + \frac{2s}{d}, \tag{3.67}$$

$$\|u\nabla(-\Delta)^{-s}v\|_p \le C\|u\|_q\|v\|_r, \quad \frac{1}{q} + \frac{1}{r} = \frac{1}{p} + \frac{2s-1}{d}, \quad s > \frac{1}{2}. \tag{3.68}$$

Lemma 3.24 (Fractional Gagliardo–Nirenberg inequality I). *Let $d \geq 2$ and $1 < p < \infty$. Then there exists $C > 0$ such that for all $u \in W^{1,p}(\mathbb{R}^d)$ or $u \in W^{2,p}(\mathbb{R}^d)$, respectively,*

$$\|(-\Delta)^s u\|_p \leq C\|u\|_p^{1-2s}\|\nabla u\|_p^{2s} \quad \text{if } 0 < s \leq 1/2,$$
$$\|(-\Delta)^s u\|_p \leq C\|u\|_p^{1-s}\|\mathrm{D}^2 u\|_p^{s} \quad \text{if } 1/2 < s \leq 1.$$

Proof. It follows from the properties of the Riesz and Bessel potentials [108, Theorem 3, page 96] that the operator $(-\Delta)^s : W^{1,p}(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ is bounded for $0 < s \leq 1/2$, while the operator $(-\Delta)^s : W^{2,p}(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ is bounded for $1/2 < s \leq 1$. Thus, if $0 < s \leq 1/2$,

$$\|(-\Delta)^s u\|_p \leq C(\|u\|_p + \|\nabla u\|_p) \quad \text{for } u \in W^{1,p}(\mathbb{R}^d).$$

Replacing u by $u_\lambda(x) = \lambda^{d/p-2s}u(\lambda x)$ with $\lambda > 0$ yields

$$\|(-\Delta)^s u\|_p = \|(-\Delta)^s u_\lambda\|_p \leq C(\|u_\lambda\|_p + \|\nabla u_\lambda\|_p) = C\lambda^{-2s}(\|u\|_p + \lambda\|\nabla u\|_p).$$

We minimize the right-hand side with respect to λ giving the value $\lambda_0 = 2s(1-2s)^{-1}\|u\|_p \|\nabla u\|_p^{-1}$ and therefore,

$$\|(-\Delta)^s u\|_p \leq C\|u\|^{1-2s}\|\nabla u\|_p^{2s}.$$

The case $1/2 < s \leq 1$ is proved in a similar way. □

Lemma 3.25 (Fractional Gagliardo–Nirenberg inequality II). *Let $d \geq 2$, $0 < s \leq 1/2$, $p \in (1,\infty)$, and $q \in [p,\infty)$. If $p < d/(2s)$, we assume additionally that $q \leq dp/(d-2sp)$. Then there exists $C > 0$ such that for all $u \in W^{1,p}(\mathbb{R}^d)$,*

$$\|(-\Delta)^{-s}\nabla u\|_q \leq C\|u\|_p^{1-\theta}\|\nabla u\|_p^{\theta},$$

where $\theta = 1 + d/p - d/q - 2s \in [0,1]$.

Proof. The statement is true for $s = 1/2$ since the operator $(-\Delta)^{-1/2}\nabla : L^q(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is bounded for any $q \in (1,\infty)$ [108, Theorem 3, page 96]. Then the inequality follows from the standard Gagliardo–Nirenberg inequality.
Thus, let $0 < s < 1/2$. We claim that it is sufficient to prove that $(-\Delta)^{-s}\nabla : W^{1,p}(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is bounded. Indeed, assume that

$$\|(-\Delta)^{-s}\nabla u\|_q \leq C(\|u\|_p + \|\nabla u\|_p) \quad \text{for } u \in W^{1,p}(\mathbb{R}^d). \tag{3.69}$$

Replacing, as in the proof of Lemma 3.24, u by $u_\lambda(x) = \lambda^{d/q-1+2s}u(\lambda x)$ with $\lambda > 0$ yields

$$\|(-\Delta)^{-s}\nabla u\|_q \leq C\lambda^{-\theta}(\|u\|_p + \lambda\|\nabla u\|_p),$$

where θ is defined in the statement of the theorem. Minimizing the right-hand side with respect to λ gives the value $\lambda_0 = \theta(1-\theta)^{-1}\|u\|_p\|\nabla u\|_p^{-1}$ and therefore,

$$\|(-\Delta)^{-s}\nabla u\|_q \leq C\|u\|_p^{1-\theta}\|\nabla u\|_p^{\theta}.$$

It remains to show (3.69). To this end, we distinguish two cases. First, let $p < d/(2s)$. By assumption, $p \leq q \leq r(1) := dp/(d-2sp)$. We apply the Hardy–Littlewood–Sobolev inequality (Lemma 3.22) to find that

$$\|(-\Delta)^{-s}\nabla u\|_{r(1)} \leq C\|\nabla u\|_p \leq C(\|u\|_p + \|\nabla u\|_p).$$

Furthermore, by using (in this order) the boundedness of $(-\Delta)^{-1/2}\nabla : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$, Lemma 2 in [108, page 133], equation (40) in [108, page 135], and Theorem 3 in [108, page 135f],

$$\begin{aligned} \|(-\Delta)^{-s}\nabla u\|_p &= \|\nabla(-\Delta)^{-1/2}(-\Delta)^{1/2-s}u\|_p \le C\|(-\Delta)^{1/2-s}u\|_p \\ &\le C\|(I-\Delta)^{1/2-s}u\|_p \le C\|(I-\Delta)^{1/2}u\|_p \le C(\|u\|_p + \|\nabla u\|_p). \end{aligned} \tag{3.70}$$

These inequalities hold for any $p \in (1,\infty)$. Now, it is sufficient to interpolate with $1/q = \mu/p + (1-\mu)/r(1)$:

$$\|(-\Delta)^{-s}\nabla u\|_q \le \|(-\Delta)^{-s}\nabla u\|_p^{\mu}\|(-\Delta)^{-s}\nabla u\|_{r(1)}^{1-\mu} \le C(\|u\|_p + \|\nabla u\|_p).$$

Second, let $p \ge d/(2s)$. We choose $\lambda \in (0, d/(2sp)) \subset (0,1)$ and apply the Hardy–Littlewoord–Sobolev inequality:

$$\|(-\Delta)^{-s}\nabla u\|_{r(\lambda)} = \|(-\Delta)^{-\lambda s}(-\Delta)^{-(1-\lambda)s}\nabla u\|_{r(\lambda)} \le C\|(-\Delta)^{-(1-\lambda)s}\nabla u\|_p,$$

where $r(\lambda) = dp/(d-2s\lambda p)$. Since $(1-\lambda)s < 1/2$, we deduce from (3.70) that

$$\|(-\Delta)^{-s}\nabla u\|_{r(\lambda)} \le C(\|u\|_p + \|\nabla u\|_p).$$

Since $r(\lambda) \to \infty$ as $\lambda \to d/(2sp)$, the result follows. □

3.C Parabolic regularity

Lemma 3.26 (Parabolic regularity)**.** *Let $1 < p < \infty$, $T > 0$ and let u be the (weak) solution to the heat equation*

$$\partial_t u - \Delta u = v, \quad u(0) = u^0 \quad \text{in } \mathbb{R}^d,$$

where $v \in L^p(0,T;L^p(\mathbb{R}^d))$ and $u^0 \in W^{2,p}(\mathbb{R}^d)$. Then there exists $C > 0$, depending on T and p, such that

$$\|\partial_t u\|_{L^p(0,T;L^p(\mathbb{R}^d))} + \|D^2 u\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\big(\|v\|_{L^p(0,T;L^p(\mathbb{R}^d))} + \|D^2 u^0\|_{L^p(\mathbb{R}^d)}\big). \tag{3.71}$$

Furthermore, if $v = \operatorname{div} w$ for some $w \in L^p(0,T;L^p(\mathbb{R}^d;\mathbb{R}^d))$ then

$$\|\nabla u\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\big(\|w\|_{L^p(0,T;L^p(\mathbb{R}^d))} + T^{1/p}\|\nabla u^0\|_{L^p(\mathbb{R}^d)}\big). \tag{3.72}$$

Proof. We use a known result on the parabolic regularity for the equation

$$\partial_t \widehat{u} - \Delta \widehat{u} = v, \quad \widehat{u}(0) = 0 \quad \text{in } \mathbb{R}^d. \tag{3.73}$$

It holds that [76]

$$\|\partial_t \widehat{u}\|_{L^p(0,T;L^p(\mathbb{R}^d))} + \|D^2 \widehat{u}\|_{L^p(0,T;L^p(\mathbb{R}^d))} \le C\|v\|_{L^p(0,T;L^p(\mathbb{R}^d))}. \tag{3.74}$$

We apply this result to $\widehat{u} = u - e^{t\Delta}u^0$, where $e^{t\Delta}u^0$ is the solution to the homogeneous heat equation in $\mathbb{R}^d$ with initial datum u^0. Then $\widehat{u}$ solves (3.73) and satisfies estimate

(3.74). Inserting the definition of $\widehat{u}$ and observing that $\|\mathrm{D}^2(e^{t\Delta}u^0)\|_p \leq C\|\mathrm{D}^2 u^0\|_p$, we obtain (3.71).
If $v = \operatorname{div} w$ for some $w \in L^p(0,T;L^p(\mathbb{R}^d;\mathbb{R}^d))$, the uniqueness of solutions to the heat equation yields $u = e^{t\Delta}u^0 + \operatorname{div} U$, where U solves

$$\partial_t U - \Delta U = w, \quad U(0) = 0 \quad \text{in } \mathbb{R}^d.$$

Then we deduce from the regularity result of [76] with $\widehat{u} = U$ and $v = w$ that

$$\|\mathrm{D}^2 U\|_{L^p(0,T;L^p(\mathbb{R}^d))} \leq C\|w\|_{L^p(0,T;L^p(\mathbb{R}^d))}.$$

Since $\nabla u = e^{t\Delta}\nabla u^0 + \nabla \operatorname{div} U$, inequality (3.72) follows. □

Dimidium facti qui coepit habet: sapere aude[1]

— Horace

4 Quantitative convergence result for a diffusion model with aggregation

Contents

This chapter shows a joint work with Li Chen[2] *and Ansgar Jüngel*[3].

4.1 Introduction and motivation

The aim of this chapter is to prove a quantitative mean-field result in $L^2(\mathbb{R}^d)$-norm associated to the following interacting stochastic particle system describing the dynamics in time $t \geq 0$ of the spatial position $X_i^{N,\eta}$ of the i-th particle,

$$\mathrm{d}X_i^{N,\eta}(t) = \frac{\kappa}{N}\sum_{j=1}^{N} \nabla V^\eta(X_i^{N,\eta}(t) - X_j^{N,\eta}(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \tag{4.1}$$

$$X_i^{N,\eta}(0) = \zeta_i \quad \text{in } \mathbb{R}^d, \quad i = 1, \ldots, N,$$

where $N \in \mathbb{N}$ denotes the number of particles, the spatial dimension $d \geq 1$, $V^\eta \geq 0$ denotes the interaction potential with interaction radius $\eta > 0$ and $\sigma > 0$ the diffusion coefficient. The parameter $\kappa = \pm 1$ models the type of the dynamics: $\kappa = -1$ corresponds to repulsive interactions and $\kappa = 1$ to aggregating particles. $(W_i)_{i=1}^N$ denotes

[1] Horaz, *Epistulae I, 2. 40f*; English translation: *The one who started has already done half of the work: dare to know!*

[2] University of Mannheim, Department of Mathematics, 68131 Mannheim, Germany

[3] Institute of Analysis and Scientific Computing, Technische Universität Wien, 1040 Wien, Austria

a family of independent d-dimensional Brownian motions on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\le t<T}, \mathbb{P})$ and $(\zeta_i)_{i=1}^N$ are $\mathscr{F}_0$-measurable i.i.d random variables with common density function $u_0 \in W^{2,\infty}(\mathbb{R}^d)$.
For fixed $\eta > 0$ the interaction potential V^η is defined by

$$V^\eta(x) := \eta^{-d} V\left(\frac{|x|}{\eta}\right) \quad \text{for } x \in \mathbb{R}^d, \tag{4.2}$$

where $V : \mathbb{R}^d \to \mathbb{R}$ is a non-negative, smooth, normalized, symmetric function with support on the unit ball $B_1(0) \subset \mathbb{R}^d$. Additionally, we assume that $V = Z * Z$, see Section 4.1.1 for discussions of the assumptions.
In order to perform the mean-field limit $N \to \infty$, the interaction radius $\eta > 0$ is coupled to the number of particles N, such that $\eta = N^{-\beta}$ with $0 < \beta < 1/(d+2)$, which leads to the regime of *moderately interacting particles*, see Section 1.2.2 for an introduction. In order to show the main result of this chapter, we need additional assumptions on the smallness of β (depending on the spatial dimension d), see Theorem 4.1.
There already are some results in the literature concerning particle system (4.1): In the repulsive case $\kappa = -1$, Oelschläger proved in [91] the convergence in law of the so-called empirical measures towards a porous-media type equation when $\eta = N^{-\beta}$ for some $0 < \beta < 1/(d+2)$. Later, he proved in [90] a quantitative mean-square convergence result in expectation of the "smoothed" empirical measure, still for $\kappa = -1$ and $0 < \beta < 1/(2d+4)$. The convergence rate in [90] is of order $O(N^{-1/2-\varepsilon})$ for a small $\varepsilon > 0$. For the more delicate aggregating case $\kappa = 1$, the mean-square convergence in probability of the smoothed empirical measure under the algebraic scaling $\eta = N^{-\beta}$ for a particle system similar to (4.1) modelling chemotaxis was shown in [109], while the (stronger) convergence of the second moments in the path space under the (weaker) logarithmic scaling $\eta \ge C(\log N)^{-1/(2d+4)}$ for some $C > 0$ was derived in [27]. By using the stronger notion of convergence, the authors in [27] derived a convergence rate which scales logarithmically in N, whereas the result in [109] does not provide a convergence rate
In this part of the thesis, we present a conditional L^2 convergence result (Theorem 4.1), which leads to a generalisation of the results of [27, 90, 109] in the sense that we allow for the (more difficult) aggregating case $\kappa = 1$, the (stronger) sense of mean-square convergence in expectation *and* the (stronger) algebraic rate $\eta = N^{-\beta}$ for some $0 < \beta < 1/(d+2)$ in the moderate regime. However, we need to assume that for algebraic scaling of $\eta > 0$ at least convergence in probability holds; see Assumption (C1) and equation (4.16). In Section 4.A, we explain the technical difficulty of proving such a result for interaction potentials approximating the Dirac measure with current techniques and present a proof for an other type of singular potential, the Coulomb potential.

4.1.1 Motivation and main results

First, in order to measure the behavior of the stochastic particle systems (4.1), we need to introduce the so-called empirical measure

$$\mu_{N,\eta}(t,\omega) := \frac{1}{N} \sum_{i=1}^N \delta_{X_i^{N,\eta}(t,\omega)}, \text{ for } t > 0, \ \omega \in \Omega, \tag{4.3}$$

where $X_i^{N,\eta}(t)$ solve (4.1) for $i = 1, \dots, N$.
It is well-known (see e.g. [27, 91]) that the limiting behavior ($N \to \infty, \eta \to 0$) of the particle system (4.1) is captured for both cases $\kappa = \pm 1$ by the local diffusion equation

$$\partial_t u = \sigma \Delta u - \kappa \operatorname{div}(u \nabla u) \quad \text{for } t > 0, \quad u(0) = u_0 \quad \text{in } \mathbb{R}^d. \tag{4.4}$$

In [91], equation (4.4) is derived for $\kappa = -1$ from (4.1) with $\eta = N^{-\beta}$ with $0 < \beta < 1/(d+2)$ by proving convergence of the empirical measure in law towards a (random) Dirac measure $\delta_{\hat{X}(t)}$, where $\hat{X}(t)$ is a process with density function $u(t)$ (with respect to the Lebesgue measure).
In contrast to this, in [27] equation (4.4) is derived from system (4.1) for $\kappa = 1$ with a different scaling in $\eta > 0$: There, the limit

$$N \to \infty \text{ and } \eta \geq C(\log(N))^{-1/(2d+4)} \to 0$$

is considered in two steps, similar to the results in Chapter 2: First, for fixed $\eta > 0$, which corresponds to weakly interacting particles, the mean-field limit leads to the nonlocal diffusion equation

$$\partial_t \bar{u}^\eta = \sigma \Delta \bar{u}^\eta - \kappa \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta), \; t > 0, \quad \bar{u}^\eta(0) = u_0 \text{ in } \mathbb{R}^d. \tag{4.5}$$

This equation is connected to the intermediate particle system

$$\begin{aligned} \mathrm{d}\overline{X}_i^\eta(t) &= \kappa (\nabla V^\eta * \bar{u}^\eta)(\overline{X}_i^\eta(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ \overline{X}_i^\eta(0) &= \zeta_i \quad \text{in } \mathbb{R}^d, \; i = 1, \dots, N, \end{aligned} \tag{4.6}$$

where all particles $\overline{X}_i^\eta(t)$ are independent and possess the common density function $\bar{u}^\eta(t)$. Note that (4.5) still depends on the number N of particles via the interaction radius $\eta = \eta(N)$. Second, since V^η converges to the Dirac delta distribution in the limit $\eta \to 0$ and $\bar{u}^\eta \to u$ ([27, Lemma 2.1]), we have $\nabla V^\eta * \bar{u}^\eta \to \nabla u$ in the sense of distributions, where u solves (4.4). This fact is used in [27] in order to show convergence of

$$\mathbb{E}\big(\sup_{0<t<T} \max_{i=1,\dots,N} |X_i^{N,\eta}(t) - \hat{X}_i(t)| \big) \to 0 \tag{4.7}$$

for $N \to \infty, \eta \to 0$, where all $\hat{X}_i(t)$ possess the common density function $u(t)$ and solve

$$\begin{aligned} \mathrm{d}\hat{X}_i(t) &= \kappa \nabla u(\hat{X}_i(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ \hat{X}_i(0) &= \zeta_i \quad \text{in } \mathbb{R}^d, \; i = 1, \dots, N. \end{aligned} \tag{4.8}$$

At this point, we want to remark that [27] only considers the case $\kappa = 1$, however, with similar arguments, the case $\kappa = -1$ can be shown in the logarithmic scaling by using the concept of the *intermediate system* (4.6), see [26] where a cross-diffusion system is considered with similar arguments and logarithmic scaling. Additionally, the main result in [27] implies that $\mu_{N,\eta}(t) \to u(t)$ in the weak sense for logarithmic scaling of η at a rate $O(\eta^2) = O(\log(N)^{-1/(d+2)})$. As explained in the introduction of this thesis, this can be seen as a *law of large numbers* in the mean-field setting. However, in order to show a *central* *limit theorem* – which will be the goal of future work – we need algebraic convergence rates in terms of N. The reason is that we want to apply the convergence result to determine the limiting behaviour of the so-called *fluctuation process*

$$\xi^N(t) := \sqrt{N}(\mu_{N,\eta}(t) - u(t)). \tag{4.9}$$

Heuristically, if this process converges to a limiting process ξ in an appropriate sense, it can be seen as a correction of the mean-field behaviour since

$$\mu_{N,\eta}(t) = u(t) + N^{-1/2}\xi^N(t) \sim u(t) + N^{-1/2}\xi(t). \tag{4.10}$$

This means that the particle dynamics for sufficiently large N can be captured by the mean-field limit $u(t)$ plus some noise term with scaling $N^{-1/2}$. If ξ is a Gaussian process, this corresponds to a *central limit theorem* in the mean-field setting.
In the setting of moderate interacting particles, we do not expect the convergence of $\mu_{N,\eta}(t) \to u(t)$ to be 'fast enough' such that $\xi^N(t)$ has a Gaussian limit. The reason is that for $\beta \to 0$ the limiting PDE changes from the local model (4.4) to the non-local PDE (4.5) for $\eta = N^0 = 1$. Hence, from an intuitive point of view we have to expect that the convergence to u is very slow for small values of $\beta > 0$ due to the structural change of the limiting PDE, see [90].
In order to still show a fluctuation theorem, we use a similar approach as K. Oelschläger in [90], where we do not compare $\mu_{N,\eta}$ with the solution to the local problem u, but to the non-local *intermediate* solution $\bar{u}^\eta$ of (4.5) plus a deterministic correction $K_\eta \to 0$ for $\eta \to 0$. In order to illustrate the main ideas and motivations here, we ignore the deterministic correction at the moment, since it is not relevant for the study of the L^2 convergence, which is the main part of this chapter.

We define the *intermediate fluctuation process* as follows

$$\xi^N_{inter}(t) := \sqrt{N}(\mu_{N,\eta}(t) - \bar{u}^\eta(t)). \tag{4.11}$$

If $\xi^N_{inter}(t) \to \xi_{inter}(t)$ for $N \to \infty$ (which implies $\eta(N) \to 0$) and $\xi_{inter}(t)$ is a Gaussian process, then by denoting the PDE error with $r^N(t) := \bar{u}^\eta(t) - u(t)$ in the spirit of (4.10), we can approximate the particle dynamics by the mean-field solution u plus the limiting intermediate fluctuations ξ_{inter} and a PDE approximation error:

$$\mu_{N,\eta}(t) = u(t) + N^{-1/2}\xi^N_{inter}(t) + r^N(t) \sim u(t) + N^{-1/2}\xi_{inter} + r^N(t). \tag{4.12}$$

Hence, it is an important question to determine the limiting behaviour of the intermediate fluctuation process ξ^N_{inter} as well as to gain estimates for the PDE error $r^N(t)$.
First, let us remark that the PDE error estimate has already been done in [27] ($\kappa = 1$) and [26] ($\kappa = -1$):

$$\sup_{0<t<T} \|\bar{u}^\eta(t) - u(t)\|^2_{L^2} + \int_0^T \|\nabla(\bar{u}^\eta(t) - u(t))\|^2_{L^2} \mathrm{d}t \le C\eta^2 = CN^{-2\beta}. \tag{4.13}$$

Second, let us take a look at the dynamics of the intermediate fluctuations: Indeed, if $V^\eta = Z^\eta * Z^\eta$ as in [90] be a convolution square and let $\langle \cdot, \cdot \rangle$ be a dual product, then the stochastic differential equation for ξ^N_{inter} reads as

$$\mathrm{d}\langle \xi^N_{inter}(t), \phi \rangle = \frac{\sqrt{2\sigma}}{\sqrt{N}} \sum_{i=1}^N \nabla\phi(X_i^{N,\eta}(t))\mathrm{d}W_i(t) + \sqrt{N}\langle |(\mu_{N,\eta} - \bar{u}^\eta) * Z^\eta|^2, \Delta\phi \rangle \mathrm{d}t + R^N(t), \tag{4.14}$$

where ϕ is a test function and $R^N(t)$ denotes an error term, which determines the test function space for ϕ and converges to zero for $N \to \infty$. For details see Section 4.1.2.
In order to prove Gaussian behaviour in the limit for $\xi^N_{inter}(t)$, we see in (4.14) that the term $(\mu_{N,\eta} - \bar{u}^\eta) * Z^\eta$ should converge in the L^2 norm faster than $N^{-1/2}$ if we assume that $\Delta\phi \in L^\infty(\mathbb{R}^d)$. This motivates us to define (as in [90]) the "smoothed" empirical measure and intermediate PDE solution

$$\begin{aligned} f^{N,\eta}(t,x) &:= (\mu_{N,\eta}(t) * Z^\eta)(x) = \frac{1}{N}\sum_{i=1}^N Z^\eta(x - X_i^{N,\eta}(t)), \\ g^\eta(t,x) &:= (\bar{u}^\eta(t) * Z^\eta)(x) = \int_{\mathbb{R}^d} Z^\eta(x-y)\bar{u}^\eta(y)\mathrm{d}y. \end{aligned} \tag{4.15}$$

Assumptions. Our main result is the $L^2(\mathbb{R}^d)$ convergence of $f^{N,\eta} - g^\eta$ in expectation with rate $N^{-1/2-\varepsilon}$. We impose the following assumptions.

(A1) Parameters: $d \geq 1$, $\sigma > 0$, $\kappa > 0$, $T > 0$.

(A2) Interaction radius: $\eta = N^{-\beta}$ with $0 < \beta < 1/(d+2)$ (moderate regime); for additional assumptions on β, see Theorem 4.1.

(A3) $W_1, \dots, W_N$ are independent d-dimensional Brownian motions on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

(A4) Initial data: $\zeta_1, \dots, \zeta_N$ are $\mathcal{F}_0$-measurable independent and identically distributed (i.i.d.) square-integrable random variables with the common density function $u_0 \in W^{2,\infty}(\mathbb{R}^d)$ whose $(d+1)$st moment is bounded.

(A5) Potential: $V = Z * Z$, where $Z \in C^3(\mathbb{R}^d)$ is symmetric, nonnegative, normalized (i.e. $\|Z\|_{L^1} = 1$), and has compact support in the ball $B_{1/2}(0)$. Define $Z^\eta(x) = \eta^{-d}Z(x/\eta)$ for $x \in \mathbb{R}^d$.

The regularity and the boundedness of the $(d+1)$st moment of u_0 are needed to obtain bounded second derivatives and bounded $(d+1)$st moment for the solution $\bar{u}^\eta(t)$ to (4.5); see Theorem 4.4 . The moment bound is used to estimate $\bar{u}^\eta$ in the "far field", where we need fast decay of the solution; see part 4 (Estimation of $L(T)$) of the proof in Section 4.7. Due to the assumptions on Z and by the definition $V = Z * Z$, V is a symmetric, nonnegative potential with $\|V\|_{L^1} = 1$. The assumption of the compact support of Z

implies that the potential V is compactly supported in the unit ball $B_1(0)$.[4] The definition of Z^η is consistent with (4.2) in the sense that $V^\eta = Z^\eta * Z^\eta$:

$$V^\eta(x) = \eta^{-d}V(x/\eta) = \eta^{-d}\int_{\mathbb{R}^d} Z(x/\eta - y)Z(y)dy$$
$$= \eta^{-2d}\int_{\mathbb{R}^d} Z((x-z)/\eta)Z(z/\eta)\mathrm{d}z = Z^\eta * Z^\eta(x).$$

Regularity of the solution to the non-local PDE (4.5). Before we state our main theorem, we need to impose the following regularity of the solution $\bar{u}^\eta \geq 0$ to the non-local PDE (4.5) for both cases $\kappa = \pm 1$:

(B1) *Regularity and uniform bounds:* $\bar{u}^\eta \in L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d)\cap L^1(\mathbb{R}^d))$, where

$$\|\bar{u}^\eta\|_{L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d))} \leq C$$

with a constant that is independent of η.

(B2) *Smallness in case of $\kappa = 1$ (aggregating case)*: If $\kappa = 1$, then $\|\bar{u}^\eta(t)\|_{L^\infty(\mathbb{R}^d)} < \sigma$.

(B3) *Uniformly bounded $(d+1)$-st moment:*

$$\sup_{0<t<T}\int_{\mathbb{R}^d} |x|^{d+1}\bar{u}^\eta(t,x)\mathrm{d}x \leq C,$$

where $C > 0$ does not depend on η.

In Section 4.3, we state assumptions on u_0 such that there exists a unique solution which fulfils (B1)–(B3); see Theorem 4.4. However, we want to remark that there may be weaker assumptions on the initial condition such that (B1)–(B3) is still satisfied.

Additionally, we need the following (weak) convergence in probability with algebraic rate:

Assumption (C1): Let $0 < \beta < 1/(10d+12)$ and the cut-off rate $\beta(d+3) < \alpha < 1/2 - \beta(d+1)$. Let $(X_i^{N,\eta})_{i=1}^N$ and $(\overline{X}_i^\eta)_{i=1}^N$ be the solutions to systems (4.1) and (4.6), respectively. Then, we assume that for any $\gamma > 0$ and $T > 0$, there exists $C(\gamma, T) > 0$ such that for all $0 < t < T$,

$$\mathbb{P}\Big(\max_{i=1,\dots,N} |X_i^{N,\eta}(t) - \overline{X}_i^\eta(t)| > N^{-\alpha}\Big) \leq C(\gamma,T)N^{-\gamma}. \tag{4.16}$$

In Section 4.A.1, we discuss Assumption (C1) in more detail. In particular, we show that for interaction potentials approximating singular potentials of Coulomb-type, the equivalent formulation of (4.16) indeed holds.
Open Problem: It is still an open problem to prove convergence in probability (4.16) to the intermediate system in the moderate regime with algebraic scaling of the interaction

[4] This condition can be weakened by assuming boundedness of the first moment of Z instead of a compact support

radius. We comment why current methods fail to handle interaction potentials approximating the Dirac measure in Section 4.A.2.

With these considerations at hand, the main theorem of this chapter can be stated in the following way:

Theorem 4.1 (Mean-square convergence with rate $N^{-1/2-\varepsilon}$)**.** *Let Assumptions (A1)–(A4) as well as (B1)–(B3) and (C1) hold and let $\eta = N^{-\beta}$, where $0 < \beta < 1/(10d+12)$. Then, for any $T > 0$, there exists $\varepsilon > 0$ and a constant $C(\beta, d, T) > 0$ such that for sufficiently large number of particles $N > 0$,*

$$\mathbb{E}\left(\sup_{0<t<T} \|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 + \sigma \int_0^T \|\nabla(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 \mathrm{d}t\right) \le C(\beta, d, T) N^{-1/2-\varepsilon}. \tag{4.17}$$

Theorem 4.1 can be summarised in the following way: Given that propagation of chaos with respect to convergence in probability (4.16) holds, even a stronger result holds, which forms an important step for showing a fluctuation result in the regime of aggregating particles, see Section 4.1.2 for an introduction why the L^2-norm is a natural norm to study for rigorously showing a fluctuation theorem.

A similar theorem as Theorem 4.1 was proved by Oelschläger in [90]. In this article, the author showed a fluctuation theorem for the so-called *corrected fluctuations* ($\sqrt{N}(\mu_{N,\eta}(t) - u(t) - c_N(t))$, where c_N is a deterministic correction) in the repulsive case $\kappa = -1$, see [90, Theorem 1] without additional assumption of the convergence in probability. Because of structural reasons, the aggregating regime $\kappa = 1$ is much more involved.
There are three main differences of [90] to Theorem 4.1: First, in our case, we consider the smoothed intermediate solution g^η, whereas Oelschläger is using an approximation of g^η instead. Second, we do not need as strict assumptions on V, especially the assumptions on the Fourier transform of Z (and therefore V) is not needed in the present work. However, we note that maybe in order to prove (4.16) rigorously, more assumptions on V might be needed. Third, [90] only considers the respulsive case $\kappa = -1$, which makes the analysis easier since the negative sign allows to neglect certain terms, which need to be estimated in a different way in case of aggregation.
Similarly as our estimate (4.17), Stevens [109] showed an L^2 convergence result for the smoothed quantities for the chemotaxis equation. Compared to that work, we do not prove the convergence of the smoothed quantities in probability but the stronger convergence in expectation, and we are able to derive a convergence rate, which is absent in [109]. However, our result needs the condition that at least propagation of chaos holds in probability, see Assumption (C1) and (4.16).

Initial condition. Since we have assumed in Assumption (A3) i.i.d. initial data, (4.17) holds at time $t = 0$ in the following way (see the last step Section 4.7 for a proof):

$$\mathbb{E}\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 \le C N^{-1/2-\varepsilon_0}, \quad \text{where } \varepsilon_0 = 1/2 - \beta d > 0. \tag{4.18}$$

4.1.2 Connection to fluctuations

As mentioned in Section 4.1.1, the main motivation to study the quantitative mean-field limit in L^2 norm (see Theorem 4.1) lies in the fact that this result is needed when studying fluctuations around the mean-field limit.

For particle systems like (4.1), an interesting question is whether the *central limit theorem* holds in the limit $N \to \infty$ when the particles become approximately independent, i.e. the question if $\sqrt{N}(\mu_{N,\eta}(0) - u_0) \to$ Gaussian (which is fulfilled if all ζ_i are i.i.d) implies

$$\xi^N(t) = \sqrt{N}(\mu_{N,\eta}(t) - u(t)) \to \text{Gaussian(t)}^5? \tag{4.19}$$

As mentioned in the introduction of this chapter, if (4.19) holds, then we can interpret the limit of the *fluctuation process* $\xi^N(t)$ as next order correction of the mean-field behaviour

$$\mu_{N,\eta}(t) = u + \frac{1}{\sqrt{N}}\xi^N(t) \sim u + \frac{1}{\sqrt{N}}\text{Gaussian(t)} + \mathcal{O}(\frac{1}{\sqrt{N}}). \tag{4.20}$$

This means that for each particle number N fixed (large enough), the particle dynamics can be approximately captured by the mean-field limit u plus some Gaussian noise with scaling $N^{-1/2}$. This is especially interesting since the particle system leading to a certain PDE is not unique, see [47] and [25] for two different particle systems leading to the porous medium equation. We refer the reader also to the end of Section 1.1.3 in the introduction of this thesis for more details. In case of the viscous porous media equation, we expect different fluctuation behaviour for those two particle systems, which would help us to understand the difference between those two systems from a modelling point of view.

Clearly, we can not expect limit (4.19) to hold if the convergence of $\mu_{N,\eta}$ to u (in the weak sense) is slower than $N^{-1/2}$. This motivates the introduction of *intermediate fluctuations* $\xi^N_{inter}(t)$ defined in (4.11). We expect that the intermediate structure captures the limiting behaviour of $\mu_{N,\eta}(t)$ in a better way than the local solution $u(t)$. If

$$\xi^N_{inter}(t) \to \text{Gaussian(t)} \tag{4.21}$$

holds (we do not specify the type of convergence here), then we can approximate the particle dynamics (4.1) by the mean-field limit u plus a deterministic correction

$$\mu_{N,\eta}(t) = u + \frac{1}{\sqrt{N}}\xi^N_{inter}(t) - (u(t) - \bar{u}^\eta(t)) \sim u + \frac{1}{\sqrt{N}}\text{Gaussian(t)} + \mathcal{O}(\frac{1}{\sqrt{N}}), \tag{4.22}$$

where the last term also captures the PDE error (4.13).

Since we want to check whether (4.21) holds, we have to study the SDE, which is fulfilled by $\xi^N_{inter}(t)$ for fixed N and η: *At this point the author wants to remark that the following lines*

[5] We do not specify what 'Gaussian' means in this context; In fact we are talking about generalised Ornstein-Uhlenbeck processes, however, since this is not within the scope of this thesis, we do not go further into details here.

give insights into the study of the intermediate fluctuations which relies on the main theorem of this chapter (Theorem 4.1). However, the arguments will be presented as motivation but without proof. The rigorous study of the limiting behaviour of the intermediate fluctuations is ongoing work and not part of the present thesis.

Using a test function $\phi(t,\cdot)$ (since we only present ideas here, we do not specify the test space), applying Itô's formula leads to

$$\begin{aligned} \mathrm{d}\langle \xi^N_{inter}(t), \phi(t)\rangle &= -\langle \xi^N_{inter}(t), (\mathcal{L}^\eta_t)^*\phi(t)\rangle \mathrm{d}t + \frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla\phi(t, X_i^{N,\eta}(t))\sqrt{2\sigma}\mathrm{d}W_i(t) \\ &\quad + \frac{1}{\sqrt{N}}\langle \xi^N_{inter}(t), \nabla V^\eta * \xi^N_{inter}(t)\nabla\phi(t)\rangle \mathrm{d}t, \end{aligned} \tag{4.23}$$

where $(\mathcal{L}^\eta_t)^*$ denotes the dual formal operator of the linearised version of the non-local PDE (4.5)

$$\begin{aligned} \mathcal{L}^\eta_t\phi &:= \partial_t\phi - \sigma\Delta\phi + \operatorname{div}\left(\phi\nabla V^\eta * \bar{u}^\eta + \bar{u}^\eta\nabla V^\eta * \phi\right) \\ (\mathcal{L}^\eta_t)^*\psi &:= -\partial_t\psi - \sigma\Delta\psi - \nabla V^\eta * \bar{u}^\eta\nabla\psi + \nabla V^\eta * (\bar{u}^\eta\nabla\psi). \end{aligned}$$

Motivated by K. Oelschläger [90], we take a test function which lies in the kernel of $(\mathcal{L}^\eta_t)^*$, which simplifies (4.23) by cancelling the first term on the right-hand side. Additionally, exploiting the "quadratic" structure $V^\eta = Z^\eta * Z^\eta$, the last term in (4.23) can we written as

$$\begin{aligned} \frac{1}{\sqrt{N}}\langle \xi^N_{inter}(t), \nabla V^\eta * \xi^N_{inter}(t)\nabla\phi(t)\rangle &= -\frac{1}{2\sqrt{N}}\langle (Z^\eta * \xi^N_{inter}(t))^2, \Delta\phi(t)\rangle \\ &+ \frac{1}{\sqrt{N}}\langle \nabla Z^\eta * \xi^N_{inter}(t), (Z^\eta * (\xi^N_{inter}\nabla\phi) - Z^\eta * \xi^N_{inter}\nabla\phi)(t)\rangle. \end{aligned}$$

The second term on the right-hand side can be viewed as 'error term' and hence ignored for this motivational section. I want to put the spotlight on the first term on the right-hand side, since it shows the necessity of studying the convergence rate of the smoothed intermediate fluctuations in L^2 norm. Assuming that the test function space is such that $\Delta\phi$ is uniformly bounded, we get

$$\begin{aligned} \left|\frac{1}{2\sqrt{N}}\langle (Z^\eta * \xi^N_{inter}(t))^2, \Delta\phi\rangle\right| &\le \sup_{0<t<T}\sup_{x\in\mathbb{R}^d}|\Delta\phi(x,t)|\frac{1}{2\sqrt{N}}\sup_{0<t<T}\|Z^\eta * \xi^N_{inter}(t)\|^2_{L^2} \\ &\le C\sqrt{N}\sup_{0<t<T}\|Z^\eta * (\mu_{N,\eta}(t) - \bar{u}^\eta(t))\|^2_{L^2}. \end{aligned}$$

Thus, in order to get a Gaussian limiting behaviour in equation (4.23), we need to study the convergence of $\sqrt{N}\sup_{0<t<T}\|Z^\eta * (\mu_{N,\eta}(t) - \bar{u}^\eta(t))\|^2_{L^2}$. Theorem 4.1 shows that $\sup_{0<t<T}\|Z^\eta * (\mu_{N,\eta}(t) - \bar{u}^\eta(t))\|^2_{L^2} \le CN^{-1/2-\varepsilon}$ for $\varepsilon > 0$, which allows us to conclude that

$$\sqrt{N}\sup_{0<t<T}\|Z^\eta * (\mu_{N,\eta}(t) - \bar{u}^\eta(t))\|^2_{L^2} \to 0 \quad \text{for } N \to \infty.$$

This shows that the main theorem of this chapter (Theorem 4.1) is a key step for understanding the limiting fluctuation behaviour of the stochastic interacting particle system (4.1).

Outlook. At the moment of writing this thesis, it is still an open question whether ξ_{inter}^N converges to a generalized Ornstein-Uhlenbeck process - which would mean that the central limit theorem holds for the intermediate fluctuations - or if we need a deterministic correction, like in [90]. Similar to [90], the correction $K_\eta(t)$ would be purely determined from the PDE structure of the intermediate non-local and the local diffusion aggregation equation.
Concerning the limiting structure - based on heuristic consideration - we expect that ξ_{inter}^N (or $\xi_{inter}^N + K_\eta$) converges to a generalised Ornstein-Uhlenbeck process which can be at least formally seen as a solution to the following (linear) SPDE of Dean-Kawasaki-type:

$$\mathrm{d}\xi(t) = \mathcal{L}_t(\xi(t))\mathrm{d}t + \sqrt{2\sigma}\nabla \cdot (\sqrt{u(t)}\xi),$$

where $\xi : L^2([0,T] \times \mathbb{R}^d; \mathbb{R}^d) \to L^2(\Omega; \mathbb{R}^d)$ is a vector-valued space-time white noise, $\mathcal{L}_t$ denotes the formal operator of the linearised version of the diffusion aggregation equation (4.4) and u is the solution to (4.4).

4.1.3 State of the art

Quantitative estimates for mean-field limits are of particular importance since they provide information on the fluctuation process. There are many notions of convergence for showing a *propagation of chaos* result, but not all of them can be directly used for fluctuation theory. For a quantitative estimate on the trajectories for particle system (4.1) in case $\kappa = 1$ and $\eta \geq C(\log(N)^{-1/(2d+4)})$ we refer to [27], where a convergence rate of $O(\log(N)^{-1/(d+2)})$ is derived which is too slow for a central limit theorem. Applying classical techniques (see for instance the lecture notes by Sznitman [113]) mean-field estimates by coupling methods naturally lead to quantitative results, however, we do not expect to get a convergence rate of order $1/\sqrt{N}$ for the moderately interacting particle system (4.1) by coupling. Thus, we will focus in this section on estimates on the empirical measure or the joint law (statistical expressions) in the setting of (4.1) (or related) and refer the reader to the review [21] for more quantitative results using coupling techniques.
Due to the fact that the study of mean-field limits is a timely topic, there are many contributions in this direction. Hence, the following summary should be understood in the sense that we only give a short outline of articles closely related to the present work and it should not be understood as a complete list of articles concerning (quantitative) mean-field results, for more detailed reviews on mean-field limits in general we refer to [21], [62] and [53].
In the regime of weakly interacting particles ($\eta = 1$, $\beta = 0$) and smooth interaction kernels, a propagation of chaos result for the finite marginals in $L^\infty((0,T); L^1(\mathbb{R}^d))$ norm at rate $1/\sqrt{N}$ is shown by relative entropy techniques in [81] in the early 2000s. Almost 20 years later, also using bounds on the relative entropy norm, Jabin and Wang [63] were able to derive a quantitative propagation of chaos result for the joint law of interacting particles for a large class of interaction kernels (with weak assumptions on the regularity)

in the weakly interacting regime at rate $O(1/\sqrt{N})$. Consequently, the authors together with Bresch were able to even derive a quantitative result (in the weakly interacting case) for singular attractive interaction kernels, which includes the Patlak-Keller-Segel system, see [10]. Almost at the same time as [63], a quantitative result concerning Coulomb-type interaction potentials with respect to a modulated energy norm was shown in [43] which generalises the result of [42]. To the best of the author's knowledge, there is no mean-field result concerning the smoothed empirical measure in the weakly interacting regime with regard to the L^2-norm used in Theorem 4.1.

First results on moderately interacting particle systems with repulsive forces have been presented by Oelschläger, showing the (non-quantitative) mean-field limit for $\beta < 1/(d+2)$ in [91], and characterizing some corrected fluctuations as Gaussians for $\beta < 1/(2d+4)$, [90]. In the latter work, he showed a similar result as in Theorem 4.1 for $\kappa = -1$ but with stronger assumptions in the interaction kernel V and not comparing to the intermediate non-local solution directly. However, in case of repulsive particles assumption (C1) (see (4.16)) is not necessary. A (non-quantitative) propagation of chaos result for reaction-diffusion equations in the regime of moderately interacting particles was shown in [94]. Extending Oelschläger's methods developed in [94], Stevens [109] was able to derive a propagation of chaos result for chemotaxis equations, by showing convergence of the smoothed empirical measure in the norm $\sup_{0<t<T} \|f(t)\|_{L^2}^2 + \int_0^T \|\nabla f(s)\|_{L^2}^2 \mathrm{d}s$ (non-quantitative), which is the same norm with respect to time and space as considered in Theorem 4.1. In comparison to the notion of convergence used in Theorem 4.1, this result only holds in probability. In [84], Méléard and Roelly generalized the result [91] for moderately interacting particles by showing a (non-quantitative) propagation of chaos result in the moderate regime by extending the space of convergence and using probabilistic methods. Sequentially, a non-quantitative propagation of chaos for a moderate model leading to a diffusion-convection equation, which does not fulfil the assumptions on the drift coefficient in [84], was shown in [64]. Later, by using probabilistic methods, a fluctuation theorem for moderately interacting particles (even with non-linear diffusion part) was derived by Jourdain and Méléard with logarithmic connection between η and N. In this article, the authors compared the empirical measure directly to the local PDE solution (no deterministic correction) and used a scaling factor different from $\sqrt{N}$, [65]. Concerning a particle approximation of a moderate model with aggregating and repulsive interaction kernels, we refer to [86], where aggregation is modelled by a non-local interaction kernel, which is different to our model.

More recently, using semigroup techniques, Flandoli and Leocata [48] were able to prove (non-quantitative) convergence of the smoothed empirical measure for a biological PDE-ODE system modelling aggregation in the moderate regime. This semigroup approach was also used in other settings of moderately interacting particles, see [49] and [50] for example. Convergence of the smoothed empirical measures in the moderate regime with logarithmic scaling for the regularisation of the (singular) kernel and algebraic scaling in η was recently shown with respect to $L^m(\Omega; L^\infty((0,T); L^p \cap L^1(\mathbb{R}^d))$-norm for some $m \in \mathbb{N}$ and $p > 2$ in [55] (non-quantitative), where singular drift terms, including repulsive Poisson kernels, and environmental noise was considered. Summarising, there are many contributions using moderately interacting particles, however, most of them are not quantitative or consider only repulsive cases.

Since Theorem 4.1 can be used in fluctuation theory, we already mentioned the contributions by Oelschläger [90] as well as Jourdain and Méléard [65] in the moderate setting. In the classical mean-field setting (weakly interacting particles), we refer to the classical work by Tanaka [114] and Dawson [39], based on ideas of Braun and Hepp [9]. For additional treatment of non-linear diffusion terms in the fluctuation setting for weakly interacting particles, we refer to [74] and [46]. Recently, motivated by the quantitative mean-field estimates of [63], a central limit for singular kernels in the weakly interacting setting has been studied in [117]. In fact, most results on fluctuations are presented for weakly interacting particle systems; see also [70, 79, 105]. Up to the author's knowledge, only the articles [90] and [65] are concerned with fluctuations for moderately interacting particle systems.

4.1.4 Main idea of the proof of Theorem 4.1

For the proof of Theorem 4.1, we wish to estimate $f^{N,\eta}-g^\eta = Z^\eta*(\mu_{N,\eta}-\bar{u}^\eta)$ in the $L^2(\mathbb{R}^d)$ norm. In order to illustrate the idea of the methods used in this work, we exemplary pick the following two terms which appear in this or similar ways many times in the proof. Let $\overline{\mu}_{N,\eta}(t)$ denote the empirical measure associated with the intermediate system (4.6) at time $t>0$, then let us define

$$Z_1 := \mathbb{E}\int_0^t |\langle Z^\eta * (\mu_{N,\eta}(s) - \overline{\mu}_{N,\eta}(s)), |\nabla f^{N,\eta} - \nabla g^\eta|^2(s)\rangle|\mathrm{d}s,$$
$$Z_2 := \mathbb{E}\int_0^t |\langle Z^\eta * (\overline{\mu}_{N,\eta}(s) - \bar{u}^\eta(s)), |\nabla f^{N,\eta} - \nabla g^\eta|^2(s)\rangle|\mathrm{d}s.$$

The proof of Theorem 4.1 is mainly based on two considerations. First, the *law-of-large-numbers estimate* (see Lemma 4.2)

$$\sup_{0<s<T} \mathbb{P}\big(|\langle\overline{\mu}_{N,\eta}(s) - \bar{u}^\eta(s), \psi_\eta\rangle| > N^{-\theta}\big) \le C(m,T)\|\psi_\eta\|_{L^\infty}^{2m} N^{2m(\theta-1/2)}, \tag{4.24}$$

valid for any $\theta \ge 0$, $m\in\mathbb{N}$, $\psi_\eta \in L^\infty(\mathbb{R}^d)$, and the *mean-field estimate* (see Assumption (C1))

$$\sup_{0<s<T} \mathbb{P}\Big(\max_{i=1,\dots,N} |X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)| > N^{-\alpha}\Big) \le C(\gamma,T)N^{-\gamma}, \tag{4.25}$$

for $0<t<T$, valid for α lying in a certain interval and for any $\gamma>0$. Note that in both estimates, the algebraic decay can be arbitrarily fast for large values of γ and m, under the conditions that $\theta<1/2$ and $\|\psi_\eta\|_{L^\infty}$ is growing not too fast in terms of $\eta(N)$. Those estimates in probability are motivated by articles by P. Pickl and co-workers, see [72] for instance.

Second – in the sprit of [72, Theorem 4.2] – we split Ω into a set $\mathcal{D}_1$, for which we can apply either (4.24) or (4.25), and its complement, where the integrands of Z_1 and Z_2 (or a related expression) are small.

To fix some ideas, let us choose $\mathcal{D}(s) = \{\omega\in\Omega : N^{-1}\sum_{i=1}^N |(X_i^{N,\eta} - \overline{X}_i^\eta)(\omega,s)| > N^{-\alpha}\}$. Then, by the mean-value theorem applied to $\phi_\eta(\cdot,y) := Z^\eta(\cdot - y)$, $G_\eta(\cdot,s) := |\nabla f^{N,\eta} - \nabla g^\eta|(\cdot,s)$ and $1 = \mathbb{1}_{\mathcal{D}_1^c} + \mathbb{1}_{\mathcal{D}_1}$ on Ω:

$$Z_1 \le \mathbb{E}\bigg(\int_0^t \frac{1}{N}\sum_{i=1}^N \big\langle\big|\phi_\eta(\cdot, X_i^{N,\eta}(s)) - \phi_\eta(\cdot, \overline{X}_i^\eta(s))\big|, |\nabla f^{N,\eta} - \nabla g^\eta|^2(\cdot,s)\big\rangle\mathrm{d}s\bigg)$$

$$
\begin{aligned}
&\leq \|\mathrm{D}\phi^\eta\|_{L^\infty} \frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big(\int_0^t \|G^\eta(s)\|_{L^2}^2 |X_i^{N,\eta} - \overline{X}_i^\eta| \mathbb{1}_{\mathcal{D}_1^c}\,\mathrm{d}s\Big) \\
&+ \sup_{0<s<T}\sup_{\omega\in\Omega} \|G^\eta(s)\|_{L^2}^2 \mathbb{E}\Big(\int_0^t \frac{1}{N}\sum_{i=1}^N \big\|\phi_\eta(\cdot, X_i^{N,\eta}) - \phi_\eta(\cdot, \overline{X}_i^\eta)\big\|_{L^\infty} \mathbb{1}_{\mathcal{D}_1}\mathrm{d}s\Big). \qquad (4.26)
\end{aligned}
$$

Since $|X_i^{N,\eta} - \overline{X}_i^\eta| \leq N^{-\alpha}$ on $\mathcal{D}_1^c$, by assuming that $\mathrm{D}\phi_\eta$ is bounded by N^r for some $r > 0$, the first integral in (4.26) is – by recalling the definition of G_η – bounded by

$$
N^{r-\alpha}\mathbb{E}\int_0^t \|\nabla f^{N,\eta} - \nabla g^\eta(s)\|_{L^2}^2 \mathrm{d}s.
$$

If ϕ_η and $\sup_{0<s<T}\sup_{\omega\in\Omega}\|G^\eta(s)\|_{L^2}^2$ are bounded by N^k for some $k > 0$, the second integral can be estimated by N^k multiplied by the probability of $\mathcal{D}_1$, which is bounded by $CN^{-\gamma}$ due to the mean-field estimate (4.25). Thus, by choosing α and γ sufficiently large, Z_1 can be estimated by

$$
Z_1 \leq \frac{\sigma}{4}\,\mathbb{E}\int_0^t \|\nabla f^{N,\eta} - \nabla g^\eta(s)\|_{L^2}^2 \mathrm{d}s + CN^{-1/2-\varepsilon}.
$$

For Z_2, we define $\mathcal{D}_2(s) = \{\omega : |Z^\eta * \overline{\mu}_{N,\eta}(s,x) - Z^\eta * \bar{u}^\eta(s,x)| > \frac{\sigma}{4}\}$ for fixed $x \in \mathbb{R}^d$. Since $|Z^\eta * \overline{\mu}_{N,\eta}(s,x) - Z^\eta * \bar{u}^\eta(s,x)|$ is small on $\mathcal{D}_2^c$, we find that

$$
\begin{aligned}
Z_2 &\leq \mathbb{E}\Big(\int_0^t \langle |Z^\eta * \overline{\mu}_{N,\eta}(s,\cdot) - Z^\eta * \bar{u}^\eta(s,\cdot)|(\mathbb{1}_{\mathcal{D}_2^c} + \mathbb{1}_{\mathcal{D}_2}), |\nabla f^{N,\eta} - \nabla g^\eta|^2(\cdot,s)\rangle|\mathrm{d}s\Big) \\
&\leq \frac{\sigma}{4}\,\mathbb{E}\int_0^t \|\nabla f^{N,\eta} - \nabla g^\eta(s)\|_{L^2}^2 \mathrm{d}s + C\|Z^\eta\|_{L^\infty} \sup_{\omega\in\Omega} \|G_\eta\|_{L^\infty(0,T;L^2(\mathbb{R}^d))}^2 \mathbb{P}(\mathcal{D}_2(s))
\end{aligned}
$$

It is important to remark that for illustrative reasons we ignored the dependence on x of the set $\mathcal{D}_2(s)$, for a more careful treatment of this (and similar terms) we refer to the proof of Theorem 4.1. The last term is estimated by using the law-of-large-numbers estimate (4.24), which gives, if all other expressions can be bounded,

$$
Z_2 \leq \frac{\sigma}{4}\,\mathbb{E}\int_0^t \|\nabla f^{N,\eta}(s) - \nabla g^\eta(s)\|_{L^2}^2 \mathrm{d}s + CN^{-1/2-\varepsilon},
$$

by taking m large enough in (4.24). Thus,

$$
Z_1(t) + Z_2(t) \leq \frac{\sigma}{2}\,\mathbb{E}\int_0^t \|\nabla f^{N,\eta}(s) - \nabla g^\eta(s)\|_{L^2}^2 \mathrm{d}s + CN^{-1/2-\varepsilon},
$$

for some $\varepsilon > 0$. The first term can be absorbed by a term which is induced by the diffusion of (4.1), whereas the second term gives us the desired rate. Many terms appearing in the analysis of $\|f^{N,\eta}(s) - g^\eta(s)\|_{L^2}$ have a similar structure as Z_1 and Z_2. However, due to an error we make by manipulating the convolution inside the dual bracket, we need more careful estimates, which are sketched in some detail in Section 4.2.

4.1.5 Outline of the chapter and notation

The chapter is organized as follows. We present more details on the proof in Section 4.2, where an outline of the proof is given. This can serve the reader as a guideline through the technical parts of the proof. The existence of solutions to (4.5) (Theorem 4.4) and the particle systems (4.1) and (4.6) as well as some properties of the solution which are needed for Theorem 4.1 are investigated in Section 4.3-4.4. Section 4.5 is concerned with some auxiliary results and estimates needed in the proof of Theorem 4.1. The law-of-large-number estimate (4.24), precisely stated in Lemma 4.2, is shown in Section 4.6. With these preparations, the detailed and rigorous proof of Theorem 4.1 is given in Section 4.7. In the appendix (Section 4.A) we discuss the assumption of propagation of chaos in probability (4.16) in more details and give a proof for convergence in probability for Coulomb-type interactions.

The norm of $L^p(\mathbb{R}^d)$ with $1 \leq p \leq \infty$ is denoted by $\|\cdot\|_{L^p}$. We write $\langle \mu, f\rangle = \int_{\mathbb{R}^d} f(x)\mathrm{d}\mu(x)$ for the dual product between a measure μ and an integrable function f. We denote the inner product on $L^2(\mathbb{R}^d)$ by the same symbol, $\langle u, v\rangle = \int_{\mathbb{R}^d} u(x)v(x)\mathrm{d}x$ for $u, v \in L^2(\mathbb{R}^d)$. The m-th derivative of a smooth function ϕ equals $\mathrm{D}^m\phi$. As usual, we omit the dependence of $\omega \in \Omega$ in most of the expressions. We denote by $C > 0$ a generic constant independent of N and η, whose value may change from line to line.

4.2 Key steps of the proof of Theorem 4.1

I. Law of large numbers in probability. As already mentioned in Section 4.1.4, the first ingredient of the proof is the law-of-large-numbers estimate (4.24). Roughly speaking, we derive an estimate for the probability that $\langle \overline{\mu}_{N,\eta} - \bar{u}^\eta, \phi_\eta\rangle$ or $(\overline{\mu}_{N,\eta} - \bar{u}^\eta) * \psi_\eta$ are outside the ball of radius $N^{-\theta}$ for an arbitrary $\theta \geq 0$.

Lemma 4.2 (Law of large numbers)**.** *Let $(\overline{X}_i^\eta)_{i=1}^N$ be the solution to system (4.6) and let $\bar{u}^\eta$ be the density function associated to $\overline{X}_i^\eta$. Given $\theta \geq 0$ and $\phi_\eta \in L^\infty(\mathbb{R}^d)$, $\psi_\eta \in L^\infty(\mathbb{R}^d; \mathbb{R}^n)$ with $n \in \{1, d, d\times d\}$, we define the sets*

$$\mathcal{A}^N_{\theta,\phi_\eta}(t) := \left\{\omega \in \Omega : \left|\frac{1}{N}\sum_{i=1}^N \phi_\eta(\overline{X}_i^\eta(t)) - \int_{\mathbb{R}^d} \phi_\eta(x)\bar{u}^\eta(t,x)\mathrm{d}x\right| > N^{-\theta}\right\}, \tag{4.27}$$

$$\mathcal{B}^N_{\theta,\psi_\eta}(t) := \bigcup_{i=1}^N \left\{\omega \in \Omega : \left|\frac{1}{N}\sum_{j=1}^N \psi_\eta\big(\overline{X}_i^\eta(t) - \overline{X}_j^\eta(t)\big) - (\psi_\eta * \bar{u}^\eta)(\overline{X}_i^\eta(t))\right| > N^{-\theta}\right\}. \tag{4.28}$$

Then, for every $m \in \mathbb{N}$ and $T > 0$, there exists $C(m) > 0$ such that for all $0 < t < T$,

$$\mathbb{P}(\mathcal{A}_{\theta,\phi_\eta}(t)) \leq C(m)\|\phi_\eta\|^{2m}_{L^\infty} N^{2m(\theta-1/2)},$$
$$\mathbb{P}(\mathcal{B}_{\theta,\psi_\eta}(t)) \leq C(m)\|\psi_\eta\|^{2m}_{L^\infty} N^{2m(\theta-1/2)+1}.$$

Remark 4.3. *Choosing $\theta < 1/2$ and assuming that the dependence of ϕ_η or ψ_η is in such a that the growth of $\|\phi_\eta\|^{2m}_{L^\infty}$ or $\|\psi_\eta\|^{2m}_{L^\infty}$ is sufficiently 'slow' in terms of η (and hence*

N) leads to an arbitrary algebraic decay of the probabilities. Since all $\overline{X}_i^\eta(t)$ are already independent at any time $t \geq 0$ this might not be surprising, however, the author could not find a quantitative result like this in literature. Hence, for the readers convenience we also present a proof in Section 4.6.

Note that the exponent of N for the estimate of the probability of $\mathcal{A}^N_{\theta,\phi_\eta}(t)$ is smaller by one than the exponent in the estimate for $\mathcal{B}^N_{\theta,\psi_\eta}(t)$, since we do not take the union over $i = 1, \dots, N$. By definition of the empirical measures, we can write

$$\mathcal{A}^N_{\theta,\phi_\eta}(t) = \{|\langle(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(t), \phi_\eta\rangle| > N^{-\theta}\},$$
$$\mathcal{B}^N_{\theta,\psi_\eta}(t) = \bigcup_{i=1}^N \{|((\overline{\mu}_{N,\eta} - \bar{u}^\eta) * \psi_\eta)(t, \overline{X}_i^\eta(t))| > N^{-\theta}\}.$$

The proof, detailed in Section 4.6, is rather standard (see, e.g., [44, Sec 2.G] for a similar proof in a slightly easier setting). To shortly summarize the proof, we exemplary pick $\mathcal{B}^N_{\theta,\psi_\eta}(t)$ and apply Markov's inequality to obtain

$$\begin{aligned}\mathbb{P}(\mathcal{B}^N_{\theta,\psi_\eta}(t)) &\leq N^{2m\theta+1} \max_{i=1,\dots,N} \mathbb{E}\left(\frac{1}{N^{2m}}\left|\sum_{j=1}^N h_{ij}(t)\right|^{2m}\right) \\ &= N^{2m(\theta-1)+1} \max_{i=1,\dots,N} \mathbb{E}\left(\left(\sum_{j,k=1}^N h_{ij}(t)h_{ik}(t)\right)^m\right),\end{aligned}$$

where $h_{ij}(t) = \psi_\eta(\overline{X}_i^\eta(t) - \overline{X}_j^\eta(t)) - (\psi_\eta * \bar{u}^\eta)(\overline{X}_i^\eta(t))$. We show that the expectation vanishes except for a number of cases which can be bounded by N^m (up to some constant). As each of the products $h_{ij}(t)h_{ik}(t)$ is bounded by $\|\psi_\eta\|^{2m}_{L^\infty}$, we conclude that

$$\mathbb{P}(\mathcal{B}^N_{\theta,\psi_\eta}(t)) \leq C(m)N^{2m(\theta-1)+1}N^m\|\psi_\eta\|^{2m}_{L^\infty},$$

proving the claim. The probability of $\mathcal{A}^N_{\theta,\phi_\eta}(t)$ is estimated in a similar way, see Section 4.6 for the complete proof.

II. Estimate of the L^2 norm: We turn to the sketch of the proof of Theorem 4.1. To compute the expectation of $\|(f^{N,\eta} - g^\eta)(t)\|^2_{L^2}$, we use Itô's formula to find after some reformulations detailed in Section 4.7:

$$\begin{aligned}&\|(f^{N,\eta} - g^\eta)(t)\|^2_{L^2} - \|(f^{N,\eta} - g^\eta)(0)\|^2_{L^2} + 2\sigma\int_0^t \|\nabla(f^{N,\eta} - g^\eta)\|^2_{L^2}\mathrm{d}s \\ &= -\frac{2\sigma t}{N}\Delta V^\eta(0) + K(t) + L(t) + M(t),\end{aligned} \tag{4.29}$$

where

$$K(t) = C(\sigma)\frac{1}{N}\sum_{i=1}^N \int_0^t ((\mu_{N,\eta} - \bar{u}^\eta) * \nabla V^\eta)(s, X_i^{N,\eta}(s))\mathrm{d}W_i(s),$$

$$L(t) = 2\kappa \int_0^t \big\langle \mu_{N,\eta} - \bar{u}^\eta, (\nabla V^\eta * \bar{u}^\eta) \cdot (\nabla Z^\eta * (f^{N,\eta} - g^\eta))\big\rangle \mathrm{d}s,$$
$$M(t) = 2\kappa \int_0^t \big\langle \mu_{N,\eta}, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2\big\rangle \mathrm{d}s.$$

The term $L(t)$ can be treated in a similar way as the corresponding term in [90, (2.20)] for $\kappa = -1$ except that by exploiting convergence in probability we make the analysis also rigorous for the multidimensional case, while the calculations regarding this term in [90] are restricted to one space dimension. The expression $M(t)$ for $\kappa = -1$ in [90, (2.20)] is negative and can be neglected in that work, but we need to estimate this term.
The idea is to estimate each of the terms on the right-hand side of (4.29) such that they are either of order $N^{-1/2-\varepsilon}$ or can be absorbed by the gradient term on the left-hand side since $\sigma > 0$. In view of the scaling of V^η, the first term on the right-hand side of (4.29) is bounded from above by $CN^{\beta(d+2)-1}$ (see (4.37) below). This expression is of order $N^{-1/2-\varepsilon}$ for some $\varepsilon > 0$ if we assume that $\beta < 1/(2d+4)$. After taking supremum in time and expectation of the expressions in (4.29), by the Burkholder–Davis–Gundy inequality, the stochastic integral $K(t)$ can be estimated by $C/N + \mathbb{E}(\sup_{0<t<T} M(t))$, such that it remains to estimate $L(t)$ and $M(t)$.

II.a. Estimate of the 'quadratic term' $M(t)$ for $\kappa = 1$: The term $M(t)$ (we refer to it as 'quadratic term'), is (for $\kappa = 1$) the most involved one and shows the strength of our new method. The reason lies in the fact that in contrast to the repulsive case ($\kappa = -1$) in the aggregation case ($\kappa = 1$), this term has a positive sign and can therefore not be neglected or used in order to absorb other terms on the left-hand side at a later stage of the estimates, like in [90]. Hence, we have to establish a different strategy in order to estimate it in a proper way such that we can indeed show estimate (4.17) with rate $N^{-1/2-\varepsilon}$.
Using Lemma 4.10, we first observe that

$$M(t) \approx 2\kappa \int_0^t \big\langle Z^\eta * \mu_{N,\eta}, |\nabla(f^{N,\eta} - g^\eta)|^2\big\rangle \mathrm{d}s, \tag{4.30}$$

which only holds up to an error term since the convolution with Z^η is inside the absolute value. However, since we want to illustrate the idea of the proof here, we ignore this error in this section; for details of the proof see Section 4.7.
Unfortunately, we cannot absorb $M(t)$ by the last term on the left-hand side of (4.29), since a naive estimate gives $\|Z^\eta * \mu_{N,\eta}\|_{L^\infty} \le CN^{\beta d}$, which diverges as $N \to \infty$. Hence, we have to estimate it directly. The idea is to add and subtract $\bar{u}^\eta$, leading to $M(t) = M_1(t) + M_2(t)$, where

$$M_1(t) = 2\int_0^t \big\langle \mu_{N,\eta} - \bar{u}^\eta, |Z^\eta * \nabla(f^{N,\eta} - g^\eta)|^2\big\rangle \mathrm{d}s,$$
$$M_2(t) = 2\int_0^t \big\langle \bar{u}^\eta, |Z^\eta * \nabla(f^{N,\eta} - g^\eta)|^2\big\rangle \mathrm{d}s.$$

For a sufficiently small initial datum, the norm $\|\bar{u}^\eta\|_{L^\infty}$ is small too (see Theorem 4.4). Moreover, by assumption, $\|Z^\eta\|_{L^1} = \|Z\|_{L^1} = 1$. Thus, after an application of Young's convolution inequality, M_2 can be absorbed by the last term on the left-hand side of (4.29). The smallness of the initial data is only needed in the aggregating case $\kappa = 1$, for the repulsive case, the term $M(t)$ can be treated as in [90] by absorbing it on the left-hand side of (4.29).

The estimation of M_1 is more delicate. Motivated by PDE techniques, a naive approach would be to estimate $M_1(t)$ similar to $M_2(t)$. Unfortunately, $\|Z^\eta * (\mu_{N,\eta} - \bar{u}^\eta)\|_{L^\infty}$ cannot be bounded uniformly in ω and hence, this naive approach is not applicable. In order to gain estimates for $\mathbb{E}(\sup_{0<t<T} |M_1(t)|)$, we use the convergence in probability (Lemma 4.2 and Assumption (C1); (4.16)), since it allows us to divide Ω in a subset where the distance between the particle dynamics and the mean-field equation is large and its complement. Lemma 4.2 and Assumption (4.16) respectively show that this set has a small probability. Hence, we add and subtract the intermediate empirical measure $\overline{\mu}_{N,\eta} = N^{-1} \sum_{i=1}^N \delta_{\overline{X}_i^\eta(t,\omega)}$. Then $\mathbb{E}(\sup_{0<t<T} |M_1(t)|)$ can be estimated by the sum of M_{11}, M_{12}, and an error term (due to the error we make in (4.30)), where

$$M_{11} = \mathbb{E}\bigg(\sup_{0<t<T} \int_0^t \big\langle Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta}, |\nabla (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \bigg),$$
$$M_{12} = \mathbb{E}\bigg(\sup_{0<t<T} \int_0^t \big\langle Z^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta), |\nabla (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \bigg).$$

The idea for M_{11} and M_{12} has already been explained in Section 4.1.4, since it forms the core element of our idea. Summarizing, we use a combination of the law of large numbers and mean-field estimates in probability ((4.24) and (4.25)) and exploit the diffusion structure of the model.

Finally, we have to estimate the error term $M_1 - M_{11} - M_{12}$. The strategy is inspired by the one in [90], where a Taylor expansion is used. However, by exploiting the idea developed for M_{11} and M_{12}, we see that a first-order expansion is sufficient. Still, the rigorous estimate of the error term is very technical and more complicated than the estimate for the one-dimensional situation of [90]; see estimates of M_{13} starting in (4.68) and (4.69).

II.b. Estimate of $L(t)$: For the term $L(t)$ in (4.29), we also add and subtract $\overline{\mu}_{N,\eta}$ to split the estimate in a mean-field part involving $\mu_{N,\eta} - \overline{\mu}_{N,\eta}$ and a law-of-large-numbers part for $\overline{\mu}_{N,\eta} - \bar{u}^\eta$. Again, the idea is to estimate both terms such that we obtain one contribution of the type $\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}$ and another contribution, which can be bounded by $N^{-1/2-\varepsilon}$, to split further the differences $\mu_{N,\eta} - \overline{\mu}_{N,\eta}$ and $\overline{\mu}_{N,\eta} - \bar{u}^\eta$, and to apply Lemma 4.2 and (4.16) several times. Additionally, this term is the reason why we need the assumption of bounded $(d+1)$-st moment of the initial data; see calculations starting in (4.91).
Combining these estimates, we infer from (4.29) that

$$\mathbb{E}\bigg(\sup_{0<t<T} \|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 - \|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 \bigg)$$
$$+ C(\sigma)\mathbb{E} \int_0^T \|\nabla (f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s \le C(\sigma, T) N^{-1/2-\varepsilon}.$$

Observing that $\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 \le CN^{-1/2-\varepsilon_0}$, see (4.18), we conclude the proof for $0 < \varepsilon \le \varepsilon_0$.

4.3 Results from PDE analysis

The main purpose of this chapter is to show an improved mean-field convergence result in $L^2(\mathbb{R}^d)$ norm of the smoothed quantities $f^{N,\eta}$ and g^η with convergence rate $N^{-1/2-\varepsilon}$, see Theorem 4.1. This result will be essential in the study of the corresponding fluctuations of the particle system. However, since by definition $g^\eta = Z^\eta * \bar{u}^\eta$, where $\bar{u}^\eta$ is the unique weak solution to (4.5), also existence, uniqueness and some bounds from classical non-linear PDE analysis for non-local equations are needed.
The analysis of equation (4.5) in the repulsive case ($\kappa = -1$) is already included in [26], where a mean-field limit to a cross-diffusion system is shown. However, due the fact that [26] holds for cross-diffusion systems, smallness assumptions on u_0 are needed which are not necessary for the non-local viscous porous media equation (4.5) in case $\kappa = -1$. The aggregating case of (4.5) ($\kappa = 1$) was already studied in [27]. For the reader's convenience and due to the fact that the assumptions in [26] and [27] on the initial condition are slightly different, we present the result here with combined assumptions:

4.3.1 Assumptions on the initial data u_0

Let in the following $s > d/2+2$. We denote with C_s the embedding constant of $H^s(\mathbb{R}^d) \hookrightarrow W^{2,\infty}(\mathbb{R}^d)$. Note that for this choice of $s > 0$, it holds that $H^s(\mathbb{R}^d) \hookrightarrow W^{2,\infty}(\mathbb{R}^d)$ continuously, see [75, Theorem 8.8].

Then, we impose the following condition on the initial datum:

$$u_0 \in W^{2,\infty}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d), \quad u_0 \ge 0, \quad \|u_0\|_{L^1} = 1, \quad \int_{\mathbb{R}^d} |x|^{d+1} u_0(x)\mathrm{d}x < \infty \tag{4.31}$$

Additionally, we assume in case $\kappa = 1$ (aggregating case) that $\|u_0\|_{H^s} < \frac{\delta}{C_s}$ for some $\delta > 0$, where we recall that C_s is the embedding constant $H^s(\mathbb{R}^d) \hookrightarrow W^{2,\infty}(\mathbb{R}^d)$.

4.3.2 Existence and uniqueness of the non-local equation (4.5)

We use the following well-posedness theorem for equation (4.5) which holds for repulsive and aggregating potentials.

Theorem 4.4 (Well-posedness of the non-local PDE (4.5) for $\kappa = \pm 1$)**.** *Let $\delta > 0$, $\eta > 0$, $s > d/2+2$, and let u_0 satisfy (4.31) and*

$$\text{if } \kappa = 1\text{: } \|u_0\|_{H^s} < \tfrac{\delta}{C_s} \text{ or if } \kappa = -1\text{: } \|u_0\|_{L^\infty} \le \delta.$$

Then there exists a unique strong solution $\bar{u}^\eta \in L^\infty(0,\infty; W^{2,\infty}(\mathbb{R}^d))$ to (4.5) such that $\bar{u}^\eta(t) \ge 0$ in $\mathbb{R}^d$, $\|\bar{u}^\eta(t)\|_{L^1} = 1$, $\|\bar{u}^\eta(t)\|_{L^\infty} \le \delta$ for $t > 0$ and the moment bound

$$\sup_{0<t<T} \int_{\mathbb{R}^d} |x|^{d+1} \bar{u}^\eta(t,x)\mathrm{d}x < \infty \text{ and } \|\bar{u}^\eta\|_{L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d))} \le C \tag{4.32}$$

hold uniformly in η.
For the stronger assumptions on u_0 in case $\kappa = 1$ the regularity $\bar{u}^\eta \in L^\infty(0,\infty; H^s(\mathbb{R}^d)) \cap L^2_{loc}(0,\infty; H^{s+1}(\mathbb{R}^d))$ holds and there exists $C > 0$ such that for all $\eta > 0$ and $T > 0$,

$$\|\bar{u}^\eta\|_{L^\infty(0,T;H^s(\mathbb{R}^d))} + \|\nabla \bar{u}^\eta\|_{L^2(0,T;H^s(\mathbb{R}^d))} \leq C. \tag{4.33}$$

Proof. As mentioned in the introduction of this section, the proof builds on results of [27] ($\kappa = 1$) and [26] ($\kappa = -1$). However, it is important to remark that [26] deals with cross-diffusion systems and hence there stronger assumptions on u_0 are needed there. We separate the proof in those two cases for existence, uniqueness and uniform bounds. Since the results in [27] and [26] do not include moment bounds of the solution $\bar{u}^\eta$, those will be shown as last step of the proof for both cases.

$\kappa = 1$: Applying [27, Theorem 1] provides a unique weak solution to (4.5), which is nonnegative, normalized, and satisfies the stated regularity. The proof of Theorem 1 in [27] shows that if $\|u_0\|_{L^\infty} \leq M$ then also $\|\bar{u}^\eta(t)\|_{L^\infty} \leq M$ for any $M > 0$ and $t > 0$. Furthermore, $\bar{u}^\eta$ is uniformly bounded in the $L^\infty(0,T;L^2(\mathbb{R}^d))$ norm and $\mathrm{D}\bar{u}^\eta$ is uniformly bounded in the $L^2(0,T;L^2(\mathbb{R}^d))$ norm. The higher-order estimates (4.33) are proved in [27, Theorem 2.2] for $s > d/2+1$, but estimates for $s > d/2+2$ can be achieved in the same way.

Next, let $\kappa = -1$. The result in [26, Propsition 1] for $n = 1$ implies (under stronger assumptions on u_0) existence of a unique, nonnegative, normalized strong solution $\bar{u}^\eta$ to (4.5) with the regularity $\bar{u}^\eta \in L^\infty(0,\infty; H^s(\mathbb{R}^d)) \cap L^2_{\mathrm{loc}}(0,\infty; H^{s+1}(\mathbb{R}^d))$ for $s > d/2+1$. However, since for $\kappa = -1$, we do not need smallness of the $L^\infty(\mathbb{R}^d)$ norm (and we are in an easier setting since no cross-diffusion is present), the assumptions stated in (4.31) are sufficient to derive the desired regularity in Theorem 4.4. We refer the reader to the proof of [26, Propsition 1] for details.

Moment estimates: It remains to derive the moment bound (4.33) for $\kappa = \pm 1$. We present here only the idea since the calculations follow standard arguments. Multiplying (4.5) by $|x|^{d+1}$ and integrating over $\mathbb{R}^d$ yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}|x|^{d+1}\bar{u}^\eta(t,x)\mathrm{d}x &= -\sigma\int_{\mathbb{R}^d}\nabla\bar{u}^\eta(t,x)\cdot\nabla|x|^{d+1}\mathrm{d}x \\ &\quad + 2\kappa\int_{\mathbb{R}^d}\bar{u}^\eta(t,x)\nabla\bar{u}^\eta(t,x)\cdot\nabla|x|^{d+1}\mathrm{d}x =: H_1(t)+H_2(t).\end{aligned} \tag{4.34}$$

Clearly, this formulation is only formal as a rigorous argument needs a cut-off function; we leave the details to the reader. Since $\nabla|x|^{d+1} = (d+1)|x|^{d-1}x$, we obtain from Hölder's inequality and the Sobolev embedding $H^s(\mathbb{R}^d) \hookrightarrow W^{1,d+1}(\mathbb{R}^d)$ with $s \geq d/2 + 1/(d+1)$:

$$\begin{aligned}H_2(t) &\leq C\||\cdot|^d\bar{u}^\eta(t,\cdot)\|_{L^{(d+1)/d}}\|\nabla\bar{u}^\eta(t)\|_{L^{d+1}} \leq C\big(\||\cdot|^d\bar{u}^\eta(t,\cdot)\|_{L^{(d+1)/d}}^{(d+1)/d} + \|\nabla\bar{u}^\eta(t)\|_{L^{d+1}}^{d+1}\big) \\ &\leq C\big(\||\cdot|^d\bar{u}^\eta(t,\cdot)\|_{L^{(d+1)/d}}^{(d+1)/d} + 1\big),\end{aligned}$$

where we have used (4.33) in the last step. As by Sobolev's embedding $\bar{u}^\eta$ is also uniformly bounded in $L^\infty(0,\infty;L^\infty(\mathbb{R}^d))$, we find that

$$\||\cdot|^d\bar{u}^\eta(t,\cdot)\|_{L^{(d+1)/d}}^{(d+1)/d} = \int_{\mathbb{R}^d}|x|^{d+1}\bar{u}^\eta(t,x)\bar{u}^\eta(t,x)^{1/d}\mathrm{d}x \leq C\int_{\mathbb{R}^d}|x|^{d+1}\bar{u}^\eta(t,x)\mathrm{d}x,$$

and consequently,

$$H_2(t) \le C + C\int_{\mathbb{R}^d} |x|^{d+1}\bar{u}^\eta(x,t)\mathrm{d}x.$$

We integrate by parts and apply Young's inequality $|x|^{d-1} \le C(1+|x|^{d+1})$ to the remaining term:

$$\begin{aligned} H_1(t) &= \sigma\int_{\mathbb{R}^d} \bar{u}^\eta(t,x)\Delta|x|^{d+1}\mathrm{d}x = \sigma(d+1)(2d-1)\int_{\mathbb{R}^d} |x|^{d-1}\bar{u}^\eta(t,x)\mathrm{d}x \\ &\le C\int_{\mathbb{R}^d}(1+|x|^{d+1})\bar{u}^\eta(t,x)\mathrm{d}x \le C + C\int_{\mathbb{R}^d}|x|^{d+1}\bar{u}^\eta(t,x)\mathrm{d}x. \end{aligned}$$

Inserting the estimates for $H_1(t)$ and $H_2(t)$ into (4.34) shows

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}|x|^{d+1}\bar{u}^\eta(t,x)\mathrm{d}x \le C + C\int_{\mathbb{R}^d}|x|^{d+1}\bar{u}^\eta(t,x)\mathrm{d}x.$$

Gronwall's lemma and the fact that C does not depend on $\eta > 0$ then concludes the proof. □

4.4 Solvability of the particle systems

The solvability of the particle systems (4.1) and (4.6) was proved in [26]. For the convenience of the reader, we recall the results.

Lemma 4.5 (Solvability of the particle systems)**.** *There exists a unique strong solution $X_i^{N,\eta}$ to system (4.1) on $(0,T)$. Moreover, if the solution $\bar{u}^\eta$ to (4.5) satisfies $\bar{u}^\eta \in L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d))$ then system (4.6) has a unique strong solution $\overline{X}_i^\eta$ with probability density function $\bar{u}^\eta$.*

A strong solution means that $(X_i^{N,\eta}(t))_{t\ge0}$ and $(\overline{X}_i^\eta(t))_{t\ge0}$ are $\mathbb{P}$-a.s. continuous, $\mathbb{R}^d$-valued, $\mathcal{F}_t$-adapted processes satisfying (4.1) and (4.6), respectively, in the sense of Itô. Note that the condition $s > 2+d/2$ in Theorem 4.4 implies that $H^s(\mathbb{R}^d) \hookrightarrow W^{2,\infty}(\mathbb{R}^d)$ which yields a unique solution $\bar{u}^\eta \in L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d))$ if the assumptions on u_0 stated in Theorem 4.4 are fulfilled.

4.5 Auxiliary results

We collect some inequalities which are used several times in the following sections. First, we remark that for continuous functions $F : \mathbb{R}^d \to \mathbb{R}^n$ for $n \in \mathbb{N}$,

$$\langle \mu_{N,\eta}(t), F\rangle = \frac{1}{N}\sum_{i=1}^N F(X_i^{N,\eta}(t)), \tag{4.35}$$

$$(\mu_{N,\eta} * F)(t,x) = \frac{1}{N}\sum_{j=1}^N F(x - X_j^{N,\eta}(t)) \quad \text{for } t>0,\ x\in\mathbb{R}^d. \tag{4.36}$$

Clearly, this also holds for the empirical measure of the intermediate measure $\overline{\mu}_{N,\eta}$. The scaling of V^η and Z^η, see (4.2) with $\eta = N^{-\beta}$, implies the following bounds.

Lemma 4.6. *It holds for $m \in \mathbb{N}_0$ that*

$$\|\mathrm{D}^m V^\eta\|_{L^\infty} + \|\mathrm{D}^m Z^\eta\|_{L^\infty} \le CN^{\beta(d+m)}, \quad \|\mathrm{D}^m V^\eta\|_{L^2} + \|\mathrm{D}^m Z^\eta\|_{L^2} \le CN^{\beta(d+2m)/2}, \tag{4.37}$$

$$\int_{\mathbb{R}^d} Z^\eta(y)|y|\mathrm{d}y \le CN^{-\beta}, \quad \int_{\mathbb{R}^d} Z^\eta(y)^2|y|^2\mathrm{d}y \le CN^{\beta(d-2)}. \tag{4.38}$$

These bounds imply the following result.

Lemma 4.7. *It holds uniformly in Ω that for any $m \in \mathbb{N}_0$*

$$\sup_{0<s<T} \|\mathrm{D}^m V^\eta * (\overline{\mu}_{N,\eta}(s) - \bar{u}^\eta(s))\|_{L^\infty} \le CN^{\beta(d+m)}.$$

Proof. It follows from the definition of $\overline{\mu}_{N,\eta}$, Young's convolution inequality, and estimate (4.37) that

$$\begin{aligned}\|\mathrm{D}^m V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(s)\|_{L^\infty} &\le \left\| \frac{1}{N}\sum_{i=1}^N \mathrm{D}^m V^\eta(x - \overline{X}_i^\eta(s)) \right\|_{L^\infty} + \|\mathrm{D}^m V^\eta\|_{L^\infty}\|\bar{u}^\eta\|_{L^1} \\ &\le 2\|\mathrm{D}^m V^\eta\|_{L^\infty} \le CN^{\beta(d+m)}.\end{aligned}$$

Since all estimates are uniform in ω, this finishes the proof. □

We also need some bounds for $f^{N,\eta} - g^\eta = (\mu_{N,\eta} - \bar{u}^\eta) * Z^\eta$.

Lemma 4.8. *It holds uniformly in Ω that*

$$\sup_{0<s<T} \|\nabla(f^{N,\eta}(s) - g^\eta(s))\|_{L^\infty} \le CN^{\beta(d+1)}, \quad \sup_{0<s<T} \|\nabla(f^{N,\eta}(s) - g^\eta(s))\|_{L^2} \le CN^{\beta(d+2)/2}. \tag{4.39}$$

Proof. The first inequality is shown as in the proof of Lemma 4.7. For the second one, we compute by substitution

$$\begin{aligned}\|\nabla f^{N,\eta}(s)\|_{L^2}^2 &= \left| \frac{1}{N^2}\sum_{i,j=1}^N \int_{\mathbb{R}^d} \nabla Z^\eta\big(y - (X_i^{N,\eta}(s) - X_j^{N,\eta}(s))\big) \cdot \nabla Z^\eta(y)\mathrm{d}y \right| \\ &\le \|\nabla Z^\eta\|_{L^\infty}\|\nabla Z^\eta\|_{L^1} \le CN^{\beta(d+2)},\end{aligned}$$

where we have used Lemma 4.7 and the fact that $\|\nabla Z^\eta\|_{L^1} = N^{\beta(d+1)}\int_{\mathbb{R}^d} |\nabla Z(xN^\beta)|\mathrm{d}x = N^\beta\|\nabla Z\|_{L^1}$.
For $g^\eta = \bar{u}^\eta * Z^\eta$, we see by Young's convolution inequality

$$\|\nabla g^\eta(s)\|_{L^2}^2 = \|\nabla Z^\eta * \bar{u}^\eta(s)\|_{L^2}^2 \le \|\nabla Z^\eta\|_{L^1}^2\|\bar{u}^\eta(s)\|_{L^2}^2 = N^{2\beta}\|\nabla Z\|_{L^1}^2\|\bar{u}^\eta(s)\|_{L^2}^2 \le CN^{2\beta},$$

due to the uniform bounds of $\bar{u}^\eta$; see Theorem 4.4. By triangle inequality, this concludes the proof. □

Lemma 4.9. *It holds that, uniformly in* Ω,

$$\sup_{0<s<T} \left(\|Z^\eta * \mu_{N,\eta}(s)\|_{L^2} + \|Z^\eta * \overline{\mu}_{N,\eta}(s)\|_{L^2}\right) \le CN^{\beta d/2}. \tag{4.40}$$

Proof. The proof is similar to that one of (4.39) (since $Z^\eta * \mu_{N,\eta} = f^{N,\eta}$):

$$\begin{aligned}\|Z^\eta * \mu_{N,\eta}(s)|_{L^2}^2 &= \left|\frac{1}{N^2}\sum_{i,j=1}^N \int_{\mathbb{R}^d} Z^\eta(x - X_i^{N,\eta}(s))Z^\eta(x - X_j^{N,\eta}(s))\mathrm{d}x\right| \\ &\le \|Z^\eta\|_{L^\infty}\|Z^\eta\|_{L^1} \le CN^{\beta d},\end{aligned}$$

using (4.37) and $\|Z^\eta\|_{L^1} = \|Z\|_{L^1}$ in the last step. The estimate for $Z^\eta * \overline{\mu}_{N,\eta}$ is very similar. □

The final result is concerned with the “shift” of the convolution in the inner product of $L^2(\mathbb{R}^d)$.

Lemma 4.10. *Let* $W \in L^1(\mathbb{R}^d)$ *be symmetric and let* u, $v \in L^2(\mathbb{R}^d)$. *Then*

$$\langle W * u, v\rangle = \langle u, W * v\rangle.$$

4.6 Proof of Lemma 4.2 (Law-of-large numbers)

To estimate the probability of $\mathcal{B}_{\theta,\psi_\eta}(t)$, defined in (4.28), we set

$$h_{ij}(t,\omega) := \psi_\eta\big(\overline{X}_i^\eta(t,\omega) - \overline{X}_j^\eta(t,\omega)\big) - (\psi_\eta * \bar{u}^\eta)(\overline{X}_i^\eta(t,\omega))$$

for $t \ge 0$, $\omega \in \Omega$, and $i, j = 1, \dots, N$. Note that depending on the choice of ψ_η, h_{ij} can be a matrix, a vector or a scalar. Then $\mathcal{B}^N_{\theta,\psi_\eta}(t) = \bigcup_{i=1}^N \mathcal{B}_i^N(t)$, where

$$\mathcal{B}_i^N(t) := \left\{\omega \in \Omega : \left|\frac{1}{N}\sum_{j=1}^N h_{ij}(t,\omega)\right| > N^{-\theta}\right\}$$

and $\mathbb{P}(\mathcal{B}^N_{\theta,\psi_\eta}(t)) \le N \max_{i=1,\dots,N} \mathbb{P}(\mathcal{B}_i^N(t))$. By the Markov inequality for $m \in \mathbb{N}$, we have for any $i = 1, \dots, N$

$$\begin{aligned}\mathbb{P}(\mathcal{B}_i^N(t)) &\le N^{2m(\theta-1)}\mathbb{E}\left(\left|\frac{1}{N}\sum_{j=1}^N h_{ij}(t)\right|^{2m}\right) \\ &= N^{2m(\theta-1)}\mathbb{E}\left(\left(\sum_{j,k=1}^N h_{ij}(t)\cdot h_{ik}(t)\right)^m\right).\end{aligned} \tag{4.41}$$

Looking at the summands of $\mathbb{E}((\sum_{j,k=1}^N h_{ij}(t)\cdot h_{ik}(t))^m)$ separately, we distinguish two cases: First, we look at summands such that there exists an index $j \in \{1, \dots, N\}$ so that

h_{ij} appears only once in the product, i.e. terms of the form $\mathbb{E}\big(h_{ij}(t)\prod_{n=1,\,k_n\neq j}^{2m-1} h_{ik_n}(t)\big)$. We claim that the fact that $\overline{X}_i^\eta$ and $\overline{X}_j^\eta$ are i.i.d. for $i \neq j$ implies that

$$\mathbb{E}\Bigg(h_{ij}(t)\prod_{n=1,\,k_n\neq j}^{2m-1} h_{ik_n}(t)\Bigg) = 0.$$

To prove the claim, we assume that ψ_η is scalar. In fact, this case can be extended to vector-valued and matrix-valued functions by taking the sum over its components. Let K denote the set of different indices k_n appearing in the product $\prod_{k_n\neq j} h_{ik_n}$, and for each $\ell \in \{1,\dots,N\}$, let α_ℓ denote its multiplicity in this product. Then, by Fubini's theorem,

$$\begin{aligned}\mathbb{E}\Bigg(h_{ij}(t)\prod_{n=1,\,k_n\neq j}^{2m-1} h_{ik_n}(t)\Bigg) = \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d}\Bigg(\int_{\mathbb{R}^d}\big(\psi_\eta(x_i-x_j)-(\psi_\eta*\bar{u}^\eta)(x_i)\big)\bar{u}^\eta(x_j)\mathrm{d}x_j\Bigg)\\ \times\prod_{\ell\in K}\big(\psi_\eta(x_i-x_\ell)-(\psi_\eta*\bar{u}^\eta)(x_i)\big)^{\alpha_\ell}\bar{u}^\eta(x_\ell)\bar{u}^\eta(x_i)\mathrm{d}x_i\bigotimes_{\ell\in K}\mathrm{d}x_\ell.\end{aligned}$$

Since $\|\bar{u}^\eta\|_{L^1}=1$, the inner integral with respect to x_j vanishes for all $x_i\in\mathbb{R}^d$,

$$\int_{\mathbb{R}^d}\big(\psi_\eta(x_i-x_j)-(\psi_\eta*\bar{u}^\eta)(x_i)\big)\bar{u}^\eta(x_j)\mathrm{d}x_j = (\psi_\eta*\bar{u}^\eta)(x_i)-(\psi_\eta*\bar{u}^\eta)(x_i)\int_{\mathbb{R}^d}\bar{u}^\eta(x_j)\mathrm{d}x_j = 0,$$

which proves the claim.
Next, we consider products of terms h_{ij}, where each factor h_{ij} appears at least twice. Those are the terms which might not have vanishing expectation. We collect them in the set

$$\mathcal{N}_i := \Bigg\{\prod_{n=1}^{2m} h_{ij_n} : \text{all indices } j_n \text{ appear at least twice}\Bigg\}.$$

We claim that the cardinality $|\mathcal{N}_i|$ of this set is, up to some factor, bounded by N^m. Indeed, it holds that if $\prod_{\alpha\in A} h_{i\alpha}\in\mathcal{N}_i$, then the cardinality of A fulfils $|A|\leq m$ since all appearing indices have to appear at least twice.
To estimate the cardinality of $\mathcal{N}_i$, we write $\mathcal{N}^i=\cup_{n=0}^{m-1}\mathcal{N}_n^i$, where

$$\mathcal{N}_i^n := \Bigg\{\prod_{\alpha\in A} h_{i\alpha}\in\mathcal{N}_i : |A| = m-n\Bigg\}.$$

- ▷ We first look at $\mathcal{N}_i^0$: It contains all products, where we have m different indices, i.e. each index appears exactly twice. We can choose $\binom{N}{m}$ such sets of indices. Since N is large in comparison to m, we roughly estimate $\binom{N}{m}\leq C(m)N^m$. By taking into account all permutations of one such selection of indices, we get $|\mathcal{N}_i^0|\leq (2m)!C(m)N^m = C(m)N^m$, where $C(m)$ is a generic constant depending on m.

- ▷ For $\mathcal{N}_i^n$ and $n>0$ with the same argumentation, we get $|\mathcal{N}_n^i|\leq C(m)N^{m-n}$

Hence (since all $\mathcal{N}_i^n$ are disjoint),

$$|\mathcal{N}_i| \le C(m)\underbrace{(N^m + N^{m-1} + \cdots + N)}_{m \text{ summands}} \le C(m)N^m. \tag{4.42}$$

The expectation of $|\sum_{j=1}^N h_{ij}|^{2m}$ can be written as the sum of expectations of products of the form $\prod_{k=1}^{2m} h_{ij_k}$, which is bounded from above by $C\|\psi_\eta\|_{L^\infty}^{2m}$. This leads to

$$\mathbb{E}\left(\left|\sum_{j=1}^N h_{ij}\right|^{2m}\right) \le C|\mathcal{N}_i|\|\psi_\eta\|_{L^\infty}^{2m} \le C(m)N^m\|\psi_\eta\|_{L^\infty}^{2m}.$$

We infer from (4.41) that

$$\mathbb{P}(\mathcal{B}_{\theta,\psi_\eta}^N(t)) \le C(m)N^{2m(\theta-1)+1}N^m\|\psi_\eta\|_{L^\infty}^{2m} = C(m)N^{2m(\theta-1/2)+1}\|\psi_\eta\|_{L^\infty}^{2m}.$$

It remains to show the estimate for $\mathcal{A}_{\theta,\phi_\eta}^N(t)$, which is done in an analogous way as the one for $\mathcal{B}_{\theta,\psi_\eta\eta}^N$. For the reader's convenience we recall the main steps: By Markov's inequality,

$$\begin{aligned}\mathbb{P}(\mathcal{A}_{\theta,\phi_\eta}^N(t)) &\le N^{2m\theta}\mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^N \phi_\eta(\overline{X}_i^\eta(t)) - \int_{\mathbb{R}^d}\phi_\eta(x)\bar{u}^\eta(t,x)\mathrm{d}x\right|^{2m}\right)\\ &\le N^{2m(\theta-1)}\mathbb{E}\left(\left|\sum_{i=1}^N h_i\right|^{2m}\right) = N^{2m(\theta-1)}\mathbb{E}\left(\left|\sum_{i,j=1}^N h_ih_j\right|^m\right),\end{aligned}$$

where $h_i(t) := \phi_\eta(\overline{X}_i^\eta(t)) - \int_{\mathbb{R}^d}\phi_\eta(x)\bar{u}^\eta(t,x)\mathrm{d}x$. Similarly as before, by a short computation the expectation of all terms in the sum such that one index $i \in \{1,\dots,N\}$ appears only once vanish, i.e.

$$\mathbb{E}\left(h_i\sum_{n=1,\,k_n\ne i}^{2m-1} h_{k_n}\right) = 0.$$

To estimate the remaining terms, we introduce

$$\mathcal{N} := \left\{\prod_{n=1}^{2m} h_{i_n} : \text{all indices } i_n \text{ appear at least twice}\right\}.$$

Its size can be estimated as before, leading to $|\mathcal{N}| \le C(m)(N^m + N^{m-1} + \cdots + N) \le C(m)N^m$. Then we deduce from $\mathbb{E}(\prod_{n=1}^{2m} h_{i_n}) \le C\|\phi_\eta\|_{L^\infty}^{2m}$ that

$$\mathbb{P}(\mathcal{A}_{\theta,\phi_\eta}^N(t)) \le C(m)N^{2m(\theta-1/2)}\|\phi_\eta\|_{L^\infty}^{2m}.$$

This finishes the proof.

4.7 Proof of Theorem 4.1 (Quantitative mean-field estimate in L^2 norm)

The proof is split into several steps. The main idea is discussed in Section 4.1.4 and the key steps are presented in Section 4.2.

Proof. **1. First reformulation.** We first reformulate the $L^2(\mathbb{R}^d)$ norm of $(f^{N,\eta} - g^\eta)(t)$ in terms of V^η, $\mu_{N,\eta}$, and $\bar{u}^\eta$. For this, we expand

$$\begin{aligned} &\|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 - \|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 = J_1 + J_2 + J_3, \quad \text{where} \\ &J_1 = \|f^{N,\eta}(t)\|_{L^2(\mathbb{R}^d)}^2 - \|f^{N,\eta}(0)\|_{L^2}^2, \\ &J_2 = \|g^\eta(t)\|_{L^2(\mathbb{R}^d)}^2 - \|g^\eta(0)\|_{L^2}^2, \\ &J_3 = -2\big(\langle f^{N,\eta}(t), g^\eta(t)\rangle - \langle f^{N,\eta}(0), g^\eta(0)\rangle\big). \end{aligned} \tag{4.43}$$

Step 1: Reformulation of J_1. By definition (4.15) of $f^{N,\eta}$, $V^\eta = Z^\eta * Z^\eta$, the symmetry of Z^η and the change of variable $y = x - X_i^{N,\eta}(t)$, we have

$$\begin{aligned} \|f^{N,\eta}(t)\|_{L^2}^2 &= \bigg\|\frac{1}{N}\sum_{i=1}^N Z^\eta(\cdot - X_i^{N,\eta}(t))\bigg\|_{L^2}^2 = \frac{1}{N^2}\int_{\mathbb{R}^d}\bigg(\sum_{i=1}^N Z^\eta(x - X_i^{N,\eta}(t))\bigg)^2 \mathrm{d}x \\ &= \frac{1}{N^2}\sum_{i,j=1}^N \int_{\mathbb{R}^d} Z^\eta\big(y + X_j^{N,\eta}(t) - X_i^{N,\eta}(t)\big)Z^\eta(y)\mathrm{d}y \\ &= \frac{1}{N^2}\sum_{i,j=1}^N (Z^\eta * Z^\eta)((X_i^{N,\eta} - X_j^{N,\eta})(t)) = \frac{1}{N^2}\sum_{i,j=1}^N V^\eta((X_i^{N,\eta} - X_j^{N,\eta})(t)). \end{aligned} \tag{4.44}$$

To reformulate the last expression, we apply Itô's formula. For this, we rewrite the particle system (4.1). In the following, we omit the argument t whenever this simplifies the notation. Using (4.36),

$$(\mu_{N,\eta} * \nabla V^\eta)(X_i^{N,\eta}) = \frac{1}{N}\sum_{j=1}^N \nabla V^\eta(X_i^{N,\eta} - X_j^{N,\eta}),$$

system (4.1) can be written as

$$\mathrm{d}X_i^{N,\eta} = \kappa(\mu_{N,\eta} * \nabla V^\eta)(X_i^{N,\eta}) + \sqrt{2\sigma}\mathrm{d}W_i,$$

and consequently, for the vector $\mathbf{X}_{ij} = (X_i^{N,\eta}, X_j^{N,\eta})^T \in \mathbb{R}^{2d}$ for some $i, j \in \{1, \dots, N\}$,

$$\mathrm{d}\mathbf{X}_{ij}(t) = \kappa\begin{pmatrix}(\mu_{N,\eta} * \nabla V^\eta)(X_i^{N,\eta}(t)) \\ (\mu_{N,\eta} * \nabla V^\eta)(X_j^{N,\eta}(t))\end{pmatrix}\mathrm{d}t + \sqrt{2\sigma}\begin{pmatrix}\mathrm{d}W_i(t) \\ \mathrm{d}W_j(t)\end{pmatrix}.$$

We introduce $g(\mathbf{X}) = V^\eta(X_1 - X_2)$ for $\mathbf{X} = (X_1, X_2)^T \in \mathbb{R}^{2d}$. The derivatives are

$$\mathrm{D}g(\mathbf{X}) = \begin{pmatrix}\nabla V^\eta(X_1 - X_2) \\ -\nabla V^\eta(X_1 - X_2)\end{pmatrix}, \ \mathrm{D}^2 g(\mathbf{X}) = \begin{pmatrix}\mathrm{D}^2 V^\eta(X_1 - X_2) & -\mathrm{D}^2 V^\eta(X_1 - X_2) \\ -\mathrm{D}^2 V^\eta(X_1 - X_2) & \mathrm{D}^2 V^\eta(X_1 - X_2)\end{pmatrix}.$$

Abbreviating $Y_{ij} = X_i^{N,\eta} - X_j^{N,\eta}$, Itô's formula gives

$$\begin{aligned}\mathrm{d}g(\mathbf{X}_{ij}) = \kappa\nabla V^\eta(Y_{ij})\cdot\big((\mu_{N,\eta}*\nabla V^\eta)(X_i^{N,\eta}) - (\mu_{N,\eta}*\nabla V^\eta)(X_j^{N,\eta})\big)\mathrm{d}t \\ + \sqrt{2\sigma}\nabla V^\eta(Y_{ij})\big(\mathrm{d}W_i - \mathrm{d}W_j\big) + 2\sigma\Delta V^\eta(Y_{ij})\mathrm{d}t.\end{aligned} \tag{4.45}$$

After summation over $i,j = 1,\dots,N$ with $i \neq j$ and using the property $\nabla V^\eta(Y_{ij}) = -\nabla V^\eta(Y_{ji})$ in the first term on the right-hand side, the integral formulation of (4.45) becomes

$$\begin{aligned}\sum_{i\neq j}\big(g(\mathbf{X}_{ij}(t)) - g(\mathbf{X}_{ij}(0))\big) = 2\kappa\sum_{i\neq j}\int_0^t \nabla V^\eta(Y_{ij}(s))\cdot(\mu_{N,\eta}*\nabla V^\eta)(X_i^{N,\eta}(s))\mathrm{d}s \\ + 2\sqrt{2\sigma}\sum_{i\neq j}\int_0^t \nabla V^\eta(Y_{ij}(s))\mathrm{d}W_i(s) + 2\sigma\sum_{i\neq j}\int_0^t \Delta V^\eta(Y_{ij}(s))\mathrm{d}s,\end{aligned}$$

where we have used for the Itô integral the definition of Y_{ij} and

$$\begin{aligned}&\sum_{\substack{i,j=1\\ i\neq j}}\int_0^t \nabla V^\eta(X_i^{N,\eta} - X_j^{N,\eta})\mathrm{d}W_i - \sum_{\substack{i,j=1\\ i\neq j}}\int_0^t \nabla V^\eta(X_i^{N,\eta} - X_j^{N,\eta})\mathrm{d}W_j \\ &= 2\sum_{\substack{i,j=1\\ i\neq j}}\int_0^t \nabla V^\eta(X_i^{N,\eta} - X_j^{N,\eta})\mathrm{d}W_i,\end{aligned}$$

due to anti-symmetry of ∇V^η.
The definition of g and J_1, the fact that the difference $V^\eta(Y_{ij}(t)) - V^\eta(Y_{ij}(0))$ vanishes for $i = j$ (since $Y_{ii}(t) = 0$ for all $t \geq 0$), and formulation (4.44) imply that

$$\begin{aligned}J_1 &= \frac{1}{N^2}\sum_{i\neq j}\big(V^\eta(Y_{ij}(t)) - V^\eta(Y_{ij}(0))\big) = \frac{1}{N^2}\sum_{i\neq j}\big(g(\mathbf{X}_{ij}(t)) - g(\mathbf{X}_{ij}(0))\big) \\ &= \frac{2\kappa}{N^2}\sum_{i\neq j}\int_0^t \nabla V^\eta(Y_{ij}(s))\cdot(\mu_{N,\eta}*\nabla V^\eta)(X_i^{N,\eta}(s))\mathrm{d}s \\ &\quad + \frac{2\sqrt{2\sigma}}{N^2}\sum_{i\neq j}\int_0^t \nabla V^\eta(Y_{ij}(s))\mathrm{d}W_i(s) + \frac{2\sigma}{N^2}\sum_{i\neq j}\int_0^t \Delta V^\eta(Y_{ij}(s))\mathrm{d}s.\end{aligned}$$

It follows from (4.36) that $N^{-1}\sum_{j=1}^N \nabla V^\eta(Y_{ij}) = (\mu_{N,\eta}*\nabla V^\eta)(X_i^{N,\eta})$ and hence,

$$\begin{aligned}J_1 &= \frac{2\kappa}{N}\sum_{i=1}^N\int_0^t |(\mu_{N,\eta}*\nabla V^\eta)(X_i^{N,\eta}(s))|^2\mathrm{d}s \\ &\quad + \frac{2\sqrt{2\sigma}}{N}\sum_{i=1}^N\int_0^t (\mu_{N,\eta}*\nabla V^\eta)(X_i^{N,\eta}(s))\mathrm{d}W_i(s) \\ &\quad + \frac{2\sigma}{N}\sum_{i=1}^N\int_0^t (\mu^{N,\eta}*\Delta V^\eta)(X_i^{N,\eta}(s))\mathrm{d}s - \frac{2\sigma}{N}\int_0^t \Delta V^\eta(0)\mathrm{d}s.\end{aligned} \tag{4.46}$$

Note that we have written the sum over $i \neq j$ as the sums over i and j minus the sum of the diagonal $i = j$. In the sum over $i = j$, we need to evaluate $\nabla V^\eta(Y_{ii}(s)) = \nabla V^\eta(0)$, which vanishes due to anti-symmetry of ∇V^η. However, the expression $\Delta V^\eta(Y_{ii}(s)) = \Delta V^\eta(0)$ does generally not vanish, explaining the last term.

Step 2: Reformulation of J_2. Since Z^η is symmetric, we infer from Lemma 4.10 that

$$\|g^\eta(t)\|_{L^2}^2 = \langle Z^\eta * \bar{u}^\eta(t), Z^\eta * \bar{u}^\eta(t)\rangle = \langle \bar{u}^\eta(t), Z^\eta * Z^\eta * \bar{u}^\eta(t)\rangle = \langle \bar{u}^\eta(t), V^\eta * \bar{u}^\eta(t)\rangle.$$

Thus, considering $V^\eta * \bar{u}^\eta$ as a test function in the weak formulation of equation (4.5) for $\bar{u}^\eta$,

$$\begin{aligned}\|g^\eta(t)\|_{L^2}^2 &= \langle \bar{u}^\eta(0), V^\eta * \bar{u}^\eta(0)\rangle + \int_0^t \langle \bar{u}^\eta, V^\eta * \partial_t \bar{u}^\eta\rangle \mathrm{d}s \\ &\quad + \sigma \int_0^t \langle \Delta \bar{u}^\eta, V^\eta * \bar{u}^\eta\rangle \mathrm{d}s - \kappa \int_0^t \langle V^\eta * \bar{u}^\eta, \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)\rangle \mathrm{d}s,\end{aligned}$$

and, after integrating by parts in the third term on the right-hand side,

$$\begin{aligned}J_2 = \|g^\eta(t)\|_{L^2}^2 - \|g^\eta(0)\|_{L^2}^2 &= \int_0^t \langle \bar{u}^\eta, V^\eta * \partial_t \bar{u}^\eta\rangle \mathrm{d}s \\ &\quad + \sigma \int_0^t \langle \bar{u}^\eta, V^\eta * \Delta \bar{u}^\eta\rangle \mathrm{d}s - \kappa \int_0^t \langle V^\eta * \bar{u}^\eta, \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)\rangle \mathrm{d}s.\end{aligned} \tag{4.47}$$

Step 3: Reformulation of J_3. We determine J_3 by first calculating the mixed term

$$\begin{aligned}\langle f^{N,\eta}(t), g^\eta(t)\rangle &= \langle \mu_{N,\eta}(t) * Z^\eta, \bar{u}^\eta * Z^\eta\rangle = \langle \mu_{N,\eta}(t), Z^\eta * Z^\eta * \bar{u}^\eta(t)\rangle \\ &= \langle \mu_{N,\eta}(t), V^\eta * \bar{u}^\eta(t)\rangle = \frac{1}{N}\sum_{i=1}^N V^\eta * \bar{u}^\eta(t, X_i^{N,\eta}(t)),\end{aligned}$$

where we have again used Lemma 4.10 and the symmetry of Z^η. By Itô's lemma applied to every summand $V^\eta * \bar{u}^\eta(t, X_i^{N,\eta}(t))$, as in (4.45),

$$\begin{aligned}J_3 &= -\frac{2}{N}\sum_{i=1}^N \big(V^\eta * \bar{u}^\eta(t, X_i^{N,\eta}(t)) - V^\eta * \bar{u}^\eta(0, X_i^{N,\eta}(0))\big) \\ &= -\frac{2}{N}\sum_{i=1}^N \int_0^t \big[\partial_t(V^\eta * \bar{u}^\eta) + \kappa(\nabla V^\eta * \bar{u}^\eta)(\nabla V^\eta * \mu_{N,\eta}) \\ &\quad + \sigma \Delta V^\eta * \bar{u}^\eta\big](s, X_i^{N,\eta}(s))\mathrm{d}s - \frac{2\sqrt{2\sigma}}{N}\sum_{i=1}^N \int_0^t \nabla V^\eta * \bar{u}^\eta(s, X_i^{N,\eta}(s))\mathrm{d}W_i(s).\end{aligned}$$

Since we have a factor 2 in front of the time derivative, $\partial_t(V^\eta * \bar{u}^\eta) = V^\eta * \partial_t \bar{u}^\eta$, inserting equation (4.5) for $\bar{u}^\eta$ yields

$$2\int_0^t \partial_t(V^\eta * \bar{u}^\eta)(s, X_i^{N,\eta}(s))\mathrm{d}s = \int_0^t V^\eta * \partial_t \bar{u}^\eta(s, X_i^{N,\eta}(s))\mathrm{d}s$$

$$+\int_0^t \big[\sigma V^\eta * \Delta\bar u^\eta(s, X_i^{N,\eta}(s)) - \kappa\nabla V^\eta * \big(\bar u^\eta \nabla V^\eta * \bar u^\eta(s, X_i^{N,\eta}(s))\big)\big]\mathrm{d}s,$$

which allows us to write J_3 as

$$\begin{aligned} J_3 = &-\frac{1}{N}\sum_{i=1}^N \int_0^t V^\eta * \partial_t \bar u^\eta(s, X_i^{N,\eta}(s))\mathrm{d}s \\ &- \frac{1}{N}\sum_{i=1}^N \int_0^t \big[\sigma V^\eta * \Delta\bar u^\eta(s, X_i^{N,\eta}(s)) - \kappa\nabla V^\eta * \big(\bar u^\eta \nabla V^\eta * \bar u^\eta(s, X_i^{N,\eta}(s))\big)\big]\mathrm{d}s \\ &- \frac{2}{N}\sum_{i=1}^N \int_0^t \big[\kappa(\nabla V^\eta * \bar u^\eta)\cdot(\nabla V^\eta * \mu_{N,\eta}) + \sigma\Delta V^\eta * \bar u^\eta\big](s, X_i^{N,\eta}(s))\mathrm{d}s \\ &- \frac{2\sqrt{2\sigma}}{N}\sum_{i=1}^N \int_0^t \nabla V^\eta * \bar u^\eta(s, X_i^{N,\eta}(s))\mathrm{d}W_i(s). \end{aligned} \tag{4.48}$$

We combine estimates (4.46)–(4.48) to find from (4.44) for J_1, J_2, and J_3 that

$$\|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 - \|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 = (K_1 + \cdots + K_6)(t), \quad \text{where} \tag{4.49}$$

$$\begin{aligned} K_1(t) &= -\frac{2\sigma}{N}\int_0^t \Delta V^\eta(0)\mathrm{d}s = -\frac{2\sigma t}{N}\Delta V^\eta(0), \\ K_2(t) &= \frac{\sigma}{N}\sum_{i=1}^N \int_0^t \big(2(\mu_{N,\eta} * \Delta V^\eta)(X_i^{N,\eta}(s)) + \langle \bar u^\eta(s), (\Delta V^\eta * \bar u^\eta)(s)\rangle \\ &\qquad - 3\Delta V^\eta * \bar u^\eta(s, X_i^{N,\eta}(s))\big)\mathrm{d}s, \\ K_3(t) &= \int_0^t \langle \bar u^\eta(s), V^\eta * \partial_t \bar u^\eta(s)\rangle \mathrm{d}s - \frac{1}{N}\sum_{i=1}^N \int_0^t V^\eta * \partial_t \bar u^\eta(s, X_i^{N,\eta}(s))\mathrm{d}s, \\ K_4(t) &= -\kappa \int_0^t \langle V^\eta * \bar u^\eta, \operatorname{div}(\bar u^\eta \nabla V^\eta * \bar u^\eta)\rangle \mathrm{d}s \\ &\qquad + \frac{\kappa}{N}\sum_{i=1}^N \int_0^t \nabla V^\eta * (\bar u^\eta \nabla V^\eta * \bar u^\eta)(s, X_i^{N,\eta}(s))\mathrm{d}s, \\ K_5(t) &= \frac{2\kappa}{N}\sum_{i=1}^N \int_0^t \big(|\nabla V^\eta * \mu_{N,\eta}|^2 - (\nabla V^\eta * \bar u^\eta)\cdot(\nabla V^\eta * \mu_{N,\eta})\big)(X_i^{N,\eta}(s))\mathrm{d}s, \\ K_6(t) &= \frac{\sqrt{8\sigma}}{N}\sum_{i=1}^N \int_0^t \nabla V^\eta * (\mu_{N,\eta} - \bar u^\eta)(s, X_i^{N,\eta}(s))\mathrm{d}W_i(s). \end{aligned}$$

In the next subsection, we rewrite $K_2, \ldots, K_5$ and directly estimate K_1 and K_6 at the end.

2. Second Reformulation. We reformulate the terms $K_2, \ldots, K_5$ in (4.49) in such a way that some terms can be combined or cancel. We start with $K_2(t)$. Using (4.36), we write

$$K_2(t) = 2\sigma \int_0^t \big(\big\langle \mu_{N,\eta}(s), \Delta V^\eta * (\mu_{N,\eta} - \bar u^\eta)(s)\big\rangle \mathrm{d}s$$

$$+ \sigma \int_0^t \big\langle (\bar{u}^\eta - \mu_{N,\eta})(s), \Delta V^\eta * \bar{u}^\eta(s)\big\rangle \mathrm{d}s.$$

Because of $V^\eta = Z^\eta * Z^\eta$ and Lemma 4.10, the first term in K_2 becomes

$$\begin{aligned} 2\sigma \int_0^t \big\langle \mu_{N,\eta}, \Delta V^\eta * (\mu_{N,\eta} - \bar{u}^\eta)\big\rangle \mathrm{d}s &= 2\sigma \int_0^t \big\langle \mu_{N,\eta} * Z^\eta, \Delta Z^\eta * (\mu_{N,\eta} - \bar{u}^\eta)\big\rangle \mathrm{d}s \\ &= 2\sigma \int_0^t \langle f^{N,\eta}, \Delta(f^{N,\eta} - g^\eta)\rangle \mathrm{d}s = -2\sigma \int_0^t \langle \nabla f^{N,\eta}, \nabla(f^{N,\eta} - g^\eta)\rangle \mathrm{d}s. \end{aligned}$$

In a similar way, the second term of $K_2(t)$ can be written as

$$\sigma \int_0^t \langle \bar{u}^\eta - \mu_{N,\eta}, \Delta V^\eta * \bar{u}^\eta\rangle \mathrm{d}s = -\sigma \int_0^t \langle \nabla(g^\eta - f^{N,\eta}), \nabla g^\eta\rangle \mathrm{d}s.$$

This shows that

$$K_2(t) = -2\sigma \int_0^t \big\langle \nabla(f^{N,\eta} - g^\eta), \nabla f^{N,\eta}\big\rangle \mathrm{d}s + \sigma \int_0^t \big\langle \nabla(f^{N,\eta} - g^\eta), \nabla g^\eta\big\rangle \mathrm{d}s.$$

Next, we consider $K_3(t)$. Using the definition $V^\eta = Z^\eta * Z^\eta$, property (4.35), Lemma 4.10, and equation (4.5), we infer that

$$\begin{aligned} K_3(t) &= \int_0^t \langle \bar{u}^\eta, Z^\eta * Z^\eta * \partial_t \bar{u}^\eta\rangle \mathrm{d}s - \int_0^t \langle \mu_{N,\eta}, Z^\eta * Z^\eta * \partial_t \bar{u}^\eta\rangle \mathrm{d}s \\ &= \int_0^t \langle (\bar{u}^\eta - \mu_{N,\eta}) * Z^\eta, Z^\eta * \partial_t \bar{u}^\eta\rangle \mathrm{d}s = \int_0^t \langle g^\eta - f^{N,\eta}, Z^\eta * \partial_t \bar{u}^\eta\rangle \mathrm{d}s \\ &= \sigma \int_0^t \langle g^\eta - f^{N,\eta}, Z^\eta * \Delta \bar{u}^\eta\rangle \mathrm{d}s - \kappa \int_0^t \big\langle g^\eta - f^{N,\eta}, Z^\eta * \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)\big\rangle \mathrm{d}s \\ &= \sigma \int_0^t \langle \nabla(f^{N,\eta} - g^\eta), \nabla g^\eta\rangle \mathrm{d}s - \kappa \int_0^t \big\langle \nabla Z^\eta * (f^{N,\eta} - g^\eta), \bar{u}^\eta \nabla V^\eta * \bar{u}^\eta\big\rangle \mathrm{d}s, \end{aligned}$$

where we integrated by parts in the last step. The first term on the right-hand side is the same as the last term in $K_2(t)$, which shows

$$\begin{aligned} K_2(t) + K_3(t) = &-2\sigma \int_0^t \big\langle \nabla(f^{N,\eta} - g^\eta), \nabla(f^{N,\eta} - g^\eta)\big\rangle \mathrm{d}s \\ &- \kappa \int_0^t \big\langle \nabla Z^\eta * (f^{N,\eta} - g^\eta), \bar{u}^\eta \nabla V^\eta * \bar{u}^\eta\big\rangle \mathrm{d}s. \end{aligned} \tag{4.50}$$

We turn to $K_4(t)$. Using the symmetry of Z^η and Lemma 4.10 again, the first term becomes

$$\begin{aligned} -\kappa\langle V^\eta * \bar{u}^\eta \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)\rangle &= -\kappa\langle Z^\eta * Z^\eta * \bar{u}^\eta, \operatorname{div}(\bar{u}^\eta \nabla Z^\eta * \bar{u}^\eta)\rangle \\ &= -\kappa\langle Z^\eta * \bar{u}^\eta, Z^\eta * \operatorname{div}(\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)\rangle = -\kappa\langle g^\eta, \nabla Z^\eta * (\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)\rangle. \end{aligned}$$

For the second term in $K_4(t)$, we take into account (4.35):

$$\frac{1}{N}\sum_{i=1}^N \nabla V^\eta * (\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)(X_i^{N,\eta}) = \frac{1}{N}\sum_{i=1}^N Z^\eta * \nabla Z^\eta * (\bar{u}^\eta \nabla V^\eta * \bar{u}^\eta)(X_i^{N,\eta})$$

$$= \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} Z^\eta(x - X_i^{N,\eta})\nabla Z^\eta * (\bar u^\eta \nabla Z^\eta * \bar u^\eta)(x)\mathrm{d}x$$
$$= \big\langle \mu_{N,\eta} * Z^\eta, \nabla Z^\eta * (\bar u^\eta \nabla V^\eta * \bar u^\eta)\big\rangle = \langle f^{N,\eta}, \nabla Z^\eta * (\bar u^\eta \nabla V^\eta * \bar u^\eta)\rangle.$$

It follows from the antisymmetry of ∇Z^η that

$$K_4(t) = \kappa \int_0^t \langle f^{N,\eta} - g^\eta, \nabla Z^\eta * (\bar u^\eta \nabla V^\eta * \bar u^\eta)\rangle \mathrm{d}s$$
$$= -\kappa \int_0^t \langle \nabla Z^\eta * (f^{N,\eta} - g^\eta), \bar u^\eta \nabla V^\eta * \bar u^\eta\rangle \mathrm{d}s,$$

which equals the second term of $K_3(t)$. Hence, by (4.50)

$$K_2(t) + K_3(t) + K_4(t) = -2\sigma \int_0^t \big\langle \nabla(f^{N,\eta} - g^\eta), \nabla(f^{N,\eta} - g^\eta)\big\rangle \mathrm{d}s$$
$$- 2\kappa \int_0^t \big\langle \nabla Z^\eta * (f^{N,\eta} - g^\eta), \bar u^\eta \nabla V^\eta * \bar u^\eta\big\rangle \mathrm{d}s. \tag{4.51}$$

Finally, we consider $K_5(t)$. We use $\nabla V^\eta = Z^\eta * \nabla Z^\eta$, the definitions of $f^{N,\eta}$ and g^η, as well as (4.35):

$$K_5(t) = \frac{2\kappa}{N}\sum_{i=1}^N \int_0^t (\nabla V^\eta * \mu_{N,\eta})(X_i^{N,\eta}) \cdot \big(Z^\eta * \nabla Z^\eta * (\mu_{N,\eta}(X_i^{N,\eta}) - \bar u^\eta(X_i^{N,\eta}))\big)\mathrm{d}s$$
$$= \frac{2\kappa}{N}\sum_{i=1}^N \int_0^t (\nabla V^\eta * \mu_{N,\eta})(X_i^{N,\eta}) \cdot \big(Z^\eta * \nabla(f^{N,\eta}(X_i^{N,\eta}) - g^\eta(X_i^{N,\eta}))\big)\mathrm{d}s$$
$$= 2\kappa \int_0^t \big\langle \mu_{N,\eta}, (\nabla V^\eta * \mu_{N,\eta}) \cdot (Z^\eta * \nabla(f^{N,\eta} - g^\eta))\big\rangle \mathrm{d}s.$$

We add the expressions for $K_2, \dots, K_5$:

$$(K_2 + \cdots + K_5)(t) = -2\sigma \int_0^t \big\langle \nabla(f^{N,\eta} - g^\eta), \nabla(f^{N,\eta} - g^\eta)\big\rangle \mathrm{d}s \tag{4.52}$$
$$- 2\kappa \int_0^t \big\langle \nabla Z^\eta * (f^{N,\eta} - g^\eta), \bar u^\eta \nabla V^\eta * \bar u^\eta\big\rangle \mathrm{d}s$$
$$+ 2\kappa \int_0^t \big\langle \mu_{N,\eta}, (\nabla V^\eta * \mu_{N,\eta}) \cdot (\nabla Z^\eta * (f^{N,\eta} - g^\eta))\big\rangle \mathrm{d}s.$$

We rewrite the second term on the right-hand side by adding and subtracting some terms in the second argument of the dual bracket,

$$\bar u^\eta \nabla V^\eta * \bar u^\eta = (\bar u^\eta - \mu_{N,\eta})\nabla V^\eta * \bar u^\eta + \mu_{N,\eta}\nabla V^\eta * (\bar u^\eta - \mu_{N,\eta}) + \mu_{N,\eta}\nabla V^\eta * \mu_{N,\eta}. \tag{4.53}$$

Then the last integral in (4.52) cancels due to the last expression of (4.53) and, because of $\nabla V^\eta * (\bar u^\eta - \mu_{N,\eta}) = \nabla Z^\eta * Z^\eta * (\bar u^\eta - \mu_{N,\eta}) = \nabla Z^\eta * (g^\eta - f^{N,\eta})$, we see that

$$- 2\kappa \int_0^t \big\langle \nabla Z^\eta * (f^{N,\eta} - g^\eta), (\bar u^\eta - \mu_{N,\eta})\nabla V^\eta * \bar u^\eta + \mu_{N,\eta}\nabla V^\eta * (\bar u^\eta - \mu_{N,\eta})\big\rangle \mathrm{d}s$$

$$
\begin{aligned}
&= -2\kappa \int_0^t \big\langle \bar{u}^\eta - \mu_{N,\eta}, (\nabla V^\eta * \bar{u}^\eta) \cdot (\nabla Z^\eta * (f^{N,\eta} - g^\eta))\big\rangle \mathrm{d}s \\
&\quad + 2\kappa \int_0^t \big\langle \mu_{N,\eta}, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s,
\end{aligned}
$$

and end up with

$$
\begin{aligned}
(K_2 + \cdots + K_5)(t) = &-2\sigma \int_0^t \|\nabla (f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s \\
&- 2\kappa \int_0^t \big\langle \bar{u}^\eta - \mu_{N,\eta}, (\nabla V^\eta * \bar{u}^\eta) \cdot (\nabla Z^\eta * (f^{N,\eta} - g^\eta))\big\rangle \mathrm{d}s \\
&+ 2\kappa \int_0^t \big\langle \mu_{N,\eta}, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s.
\end{aligned}
$$

In the repulsive case $\kappa = -1$, the last term is nonpositive and can be not only discarded but also used in order to absorb other terms, see [90]. However, in the attractive case $\kappa = 1$, we need to estimate this expression, which complicates the proof considerably.
We insert the previous formulation for $K_2 + \cdots + K_5$ into (4.49) and take the supremum over $0 < t < T$ and then the expectation:

$$
\begin{aligned}
&\mathbb{E}\Big(\sup_{0<t<T} \|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 \Big) + 2\sigma \mathbb{E} \int_0^T \|\nabla (f^{N,\eta} - g^\eta)(s)\|_{L^2}^2 \mathrm{d}s \\
&\quad \le \mathbb{E}\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 + \mathbb{E}\Big(\sup_{0<t<T} |K_1(t) + K_6(t)| \Big) + L(T) + M(T),
\end{aligned} \tag{4.54}
$$

where

$$
L(T) = 2\kappa \mathbb{E}\bigg(\sup_{0<t<T} \int_0^t \big\langle \mu_{N,\eta} - \bar{u}^\eta, (\nabla V^\eta * \bar{u}^\eta) \cdot (\nabla Z^\eta * (f^{N,\eta} - g^\eta))\big\rangle \mathrm{d}s \bigg), \tag{4.55}
$$

$$
M(T) = 2\kappa \mathbb{E}\bigg(\sup_{0<t<T} \int_0^t \big\langle \mu_{N,\eta}, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \bigg). \tag{4.56}
$$

The term $K_1(t)$ can be estimated directly by using (4.37):

$$
K_1(t) = -\frac{2\sigma t}{N} \Delta V^\eta(0) \le C(T) N^{\beta(d+2)-1}. \tag{4.57}
$$

To estimate $K_6(t)$, we use the Burkholder–Davis–Gundy inequality and Jensen's inequality:

$$
\mathbb{E}\Big(\sup_{0<t<T} |K_6(t)| \Big) \le C \mathbb{E}(\langle K_6 \rangle_T^{1/2}) \le C(\mathbb{E}\langle K_6 \rangle_T)^{1/2}.
$$

where $\langle K_6 \rangle_T$ is the quadratic variation process of K_6 at time $T > 0$. Since for different particles, the Brownian motions W_i are independent, the quadratic variation becomes

$$
(\mathbb{E}\langle K_6 \rangle_T)^{1/2} = \bigg(\frac{8\sigma}{N^2} \sum_{i=1}^N \mathbb{E} \int_0^T \big|\nabla V^\eta * \big(\mu_{N,\eta}(s, X_i^{N,\eta}(s)) - \bar{u}^\eta(s, X_i^{N,\eta}(s))\big)\big|^2 \mathrm{d}s \bigg)^{1/2}
$$

$$= \Big(\frac{8\sigma}{N} \mathbb{E} \int_0^T \big\langle \mu_{N,\eta}(s), |\nabla V^\eta * (\mu_{N,\eta} - \bar{u}^\eta)(s)|^2 \big\rangle \mathrm{d}s \Big)^{1/2}$$
$$= \Big(\frac{8\sigma}{N} \mathbb{E} \int_0^T \big\langle \mu_{N,\eta}(s), |\nabla Z^\eta * (f^{N,\eta} - g^\eta)(s)|^2 \big\rangle \mathrm{d}s \Big)^{1/2}.$$

We infer from Young's inequality, definition (4.56) of $M(t)$, and $|\kappa| = 1$ that for any $\delta > 0$ small

$$\mathbb{E}\Big(\sup_{0<t<T} |K_6(t)| \Big) \le \frac{C(\sigma,\delta)}{N} + \delta \mathbb{E} \int_0^T \big\langle \mu_{N,\eta}, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \le \frac{C(\sigma,\delta)}{N} + \delta |M(T)|. \tag{4.58}$$

It remains to estimate $L(T)$ and $M(T)$. We start with the estimate of $M(T)$ before turning to the slightly easier (and similar) calculation of $L(T)$.

3. Estimation of $M(T)$. Before we start with estimates for $M(T)$, which contain the main novelty of our method, we want to remark that the following calculations hold in both cases $\kappa = \pm 1$. However – as mentioned before – by using the negative sign of $M(T)$ in case $\kappa = -1$, this tedious estimate is not necessary (see [90]). Indeed the following calculations rely on *smallness* of $\|\bar{u}^\eta(t)\|_{L^\infty}$ in comparison to the diffusion parameter σ, which is not needed for $\kappa = -1$.

By adding and subtracting $\bar{u}^\eta$, we use $|\kappa| = 1$ and write $M \le M_1 + M_2$, where

$$M_1(T) = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \big\langle \mu_{N,\eta} - \bar{u}^\eta, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \Big),$$
$$M_2(T) = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \big\langle \bar{u}^\eta, |\nabla Z^\eta * (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \Big).$$

We infer from Young's convolution inequality and $\|Z^\eta\|_{L^1} = 1$ that

$$\begin{aligned} M_2(T) &\le 2\|\bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \|Z^\eta\|_{L^1}^2 \int_0^t \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s \\ &\le 2\delta \int_0^t \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s, \end{aligned} \tag{4.59}$$

if $\|\bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le \delta$, where $\delta > 0$ is some arbitrary number. This smallness condition is possible, due to Theorem 4.4, if the initial datum is small enough. The idea is to absorb the right-hand side of (4.59) by the left-hand side of (4.54), which requires that $\delta < \sigma$. The estimate of M_1 is more involved, and we split this expression as $M_1 \le M_{11} + M_{12} + M_{13}$ by adding and subtracting suitable expressions, where

$$M_{11} = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \big\langle Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta}), |\nabla(f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \Big), \tag{4.60}$$
$$M_{12} = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \big\langle Z^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta), |\nabla(f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \Big),$$

$$M_{13} = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \Big(\big\langle \mu_{N,\eta} - \bar{u}^\eta, |\nabla Z^\eta * (f^{N,\eta} - g^\eta|^2 \big\rangle - \big\langle Z^\eta * (\mu_{N,\eta} - \bar{u}^\eta), |\nabla (f^{N,\eta} - g^\eta)|^2 \big\rangle \Big) \mathrm{d}s \Big),$$

where $\overline{\mu}_{N,\eta}(t) = N^{-1} \sum_{i=1}^N \delta_{\overline{X}_i^\eta(t,\omega)}$ denotes the empirical measure of the intermediate system (4.6).

The term M_{11} is estimated by the mean-field assumption (4.16), while M_{13} is treated by the law-of-large-numbers estimate of Lemma 4.2. The last term M_{13} can be seen as an error term, whose estimation is delicate and which needs a very careful analysis.

Step 1. Estimation of M_{11} (Mean-field estimate). To apply the mean-field result in probability (4.16), which we assumed in Theorem 4.1, we introduce the set

$$\mathcal{C}_\alpha(t) := \Big\{ \omega \in \Omega : \max_{i=1,\dots,N} |X_i^{N,\eta}(t) - \overline{X}_i^\eta(t)| > N^{-\alpha} \Big\}, \tag{4.61}$$

where $\alpha \in (\beta(d+3), 1/2 - \beta(d+1))$. By (4.16), for any $\gamma > 0$ and $T > 0$, there exists $C(\gamma, T) > 0$ such that

$$\sup_{0<t<T} \mathbb{P}(\mathcal{C}_\alpha(t)) \le C(\gamma, T) N^{-\gamma}. \tag{4.62}$$

The idea is to split Ω into the set $\mathcal{C}_\alpha(s)$ and its complement $\mathcal{C}_\alpha^c(s)$ with $s \in (0,t)$ and to estimate M_{11} on these two sets separately. For this, we write $M_{11} \le M_{111} + M_{112}$, where

$$M_{111} = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \big\langle Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta}), \mathbb{1}_{\mathcal{C}_\alpha(s)} |\nabla (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \Big),$$
$$M_{112} = 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \big\langle Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta}), \mathbb{1}_{\mathcal{C}^c_\alpha(s)} |\nabla (f^{N,\eta} - g^\eta)|^2 \big\rangle \mathrm{d}s \Big).$$

Going back to the particle formulation and using $|X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)| \le N^{-\alpha}$ on $\mathcal{C}_\alpha^c(s)$ as well as $\|\nabla Z^\eta\|_{L^\infty} \le CN^{\beta(d+1)}$ from (4.37), we find that

$$\begin{aligned}
&M_{112}(T) \\
&\le 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \Big\langle \frac{1}{N} \sum_{i=1}^N \big| Z^\eta(x - X_i^{N,\eta}) - Z^\eta(x - \overline{X}_i^\eta) \big| \mathbb{1}_{\mathcal{C}^c_\alpha(s)}, |\nabla (f^{N,\eta} - g^\eta)|^2 \Big\rangle \mathrm{d}s \Big) \\
&\le 2\|\nabla Z^\eta\|_{L^\infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \int_0^T \big\langle |X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)| \mathbb{1}_{\mathcal{C}^c_\alpha(s)}, |\nabla (f^{N,\eta} - g^\eta)(s)|^2 \big\rangle \mathrm{d}s \\
&\le CN^{\beta(d+1)-\alpha} \mathbb{E} \int_0^T \|\nabla (f^{N,\eta} - g^\eta)(s)\|_{L^2}^2 \mathrm{d}s \le \delta \mathbb{E} \int_0^T \|\nabla (f^{N,\eta} - g^\eta)(s)\|_{L^2}^2 \mathrm{d}s,
\end{aligned}$$

choosing N sufficiently large such that $CN^{\beta(d+1)-\alpha} \le \delta$. This is possible since

$$\alpha > \beta(d+1).$$

On the remaining set M_{111}, we use $\|Z^\eta\|_{L^\infty} \le CN^{\beta d}$,

$$\sup_{\omega\in\Omega}\sup_{0<s<T}\|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2 \le CN^{\beta(d+2)}$$

from (4.39), and estimate (4.62). Fubini's theorem then gives

$$\begin{aligned} M_{111}(T) &\le 4\|Z^\eta\|_{L^\infty}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\bigg(\sup_{0<s<T}\|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2\int_0^T \mathbb{1}_{\mathcal{C}_\alpha(s)}\mathrm{d}s\bigg) \\ &\le 4TN^{\beta d}\sup_{0<s<T}\Big[\sup_{\omega\in\Omega}\|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2\mathbb{P}(\mathcal{C}_\alpha(s))\Big] \\ &\le C(T)N^{\beta d}N^{\beta(d+2)}N^{-\gamma} = C(T)N^{2\beta(d+1)-\gamma} \le C(T)N^{-1/2-\varepsilon}, \end{aligned}$$

where the last step follows after choosing $\gamma > 1/2 + 2\beta(d+1)$ and where $\varepsilon > 0$ denotes here and in the following a small number with values changing in the proof (here, we can choose $\varepsilon := \gamma - 1/2 - 2\beta(d+1) > 0$). We conclude that

$$M_{11}(T) \le C(T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2\mathrm{d}s, \tag{4.63}$$

for N sufficiently large and $\delta > 0$ arbitrary small.

Step 2. Estimation of M_{12} (Law-of-large numbers). The term M_{12} is treated by the law-of-large-numbers estimate of Lemma 4.2. For this, we introduce for fixed $\delta > 0$ (will be chosen later)

$$\mathcal{D}_\delta(s,x) := \bigg\{\omega\in\Omega : \bigg|\frac{1}{N}\sum_{i=1}^N Z^\eta(x-\overline{X}_i^\eta(s)) - (Z^\eta * \bar{u}^\eta(s,x))\bigg| > \delta\bigg\}, \tag{4.64}$$

and we split Ω into the sets $\mathcal{D}_\delta(s,\cdot)$ and $\mathcal{D}_\delta^c(s,\cdot)$. Then $M_{12} \le M_{121} + M_{122}$, where

$$M_{121}(T) = 2\mathbb{E}\int_0^T \bigg\langle\bigg|\frac{1}{N}\sum_{i=1}^N Z^\eta(\cdot - \overline{X}_i^\eta(s)) - (Z^\eta * \bar{u}^\eta)(s)\bigg|, \mathbb{1}_{\mathcal{D}_\delta^c(s,\cdot)}|\nabla(f^{N,\eta}-g^\eta)(s)|^2\bigg\rangle\mathrm{d}s,$$

$$M_{122}(T) = 2\mathbb{E}\int_0^T \bigg\langle\bigg|\frac{1}{N}\sum_{i=1}^N Z^\eta(\cdot - \overline{X}_i^\eta(s)) - (Z^\eta * \bar{u}^\eta)(s)\bigg|, \mathbb{1}_{\mathcal{D}_\delta(s,\cdot)}|\nabla(f^{N,\eta}-g^\eta)(s)|^2\bigg\rangle\mathrm{d}s.$$

In $\mathcal{D}_\delta(s,\cdot)^c$, we have $|N^{-1}\sum_{i=1}^N Z^\eta(\cdot - \overline{X}_i^\eta(s)) - (Z^\eta * \bar{u}^\eta)(s,\cdot)| \le \delta$ and therefore,

$$M_{121}(T) \le 2\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2\mathrm{d}s.$$

For the second term M_{122}, we have to be careful with the x-dependence of the set $\mathcal{D}_\delta(s,x)$. First, we use $\|Z^\eta * \bar{u}^\eta\|_{L^\infty} \le \|Z^\eta\|_{L^\infty} \le CN^{\beta d}$ from (4.37) and Fubini's theorem

$$M_{122}(T) \le 2\|Z^\eta\|_{L^\infty}\mathbb{E}\int_0^T \langle\mathbb{1}_{\mathcal{D}_\delta(s,\cdot)}, |\nabla(f^{N,\eta}-g^\eta)(s)|^2\rangle\mathrm{d}s \tag{4.65}$$

$$
\begin{aligned}
&\le CN^{\beta d}\int_0^T \big\langle \mathbb{E}(\mathbb{1}_{\mathcal{D}_\delta(s,\cdot)}), \sup_{\omega\in\Omega}|\nabla(f^{N,\eta}-g^\eta)(s)|^2\big\rangle ds \\
&\le CN^{\beta d}\int_0^T \sup_{x\in\mathbb{R}^d}\mathbb{E}(\mathbb{1}_{\mathcal{D}_\delta(s,x)})\|\sup_{\omega\in\Omega}\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2 \mathrm{d}s \\
&\le C(T)N^{\beta d}N^{\beta(d+2)}\sup_{0<s<T}\sup_{x\in\mathbb{R}^d}\mathbb{P}(\mathcal{D}_\delta(s,x)),
\end{aligned}
$$

where we used (analogous to (4.39)) that

$$
\begin{aligned}
\|\sup_{\omega\in\Omega}\nabla f^{N,\eta}\|_{L^2}^2 &\le \int_{\mathbb{R}^d}\frac{1}{N^2}\sum_{i,j=1}^N \sup_{\omega\in\Omega}\big|\nabla Z^\eta(X_i^{N,\eta}(s)-X_j^{N,\eta}(s)+y)\nabla Z^\eta(y)\big|dy \\
&\le \|\nabla Z^\eta\|_{L^\infty}\|\nabla Z^\eta\|_{L^1}\le CN^{\beta(d+2)}
\end{aligned} \tag{4.66}
$$

in the last step.
Now, we apply for fixed $x\in\mathbb{R}^d$ Lemma 4.2 with $\phi_\eta(y)=Z^\eta(x-y)$, $\theta=0$, and $m>0$:

$$
\mathbb{P}(\mathcal{D}_\delta(s,x))\le C(\delta)\|\phi_\eta\|_{L^\infty}^{2m}N^{-m}\le C(\delta)N^{2m\beta d}N^{-m}=C(\delta)N^{-(1-2\beta d)m},
$$

where we note that the right-hand side now depends on $\delta>0$ if we choose $\mathcal{D}_\delta(s,x)$ depending on δ. This is slightly different than in Lemma 4.2, but does not change the computation of the proof.
By assumption, $\beta<1/2d$ such that $1-2\beta d>0$. Thus, if we choose m large enough,

$$
M_{122}(T)\le C(T)N^{2\beta(d+1)-(1-2\beta)m}\le C(T)N^{-1/2-\varepsilon}.
$$

Summarizing the estimates for M_{121} and M_{122}, we infer that

$$
M_{12}(T)=C(T)N^{-1/2-\varepsilon}+2\delta\mathbb{E}\int_0^T\|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2\mathrm{d}s, \tag{4.67}
$$

which finishes Step 2.

Step 3. Estimation of M_{13} (error estimate). We turn to the error term $M_{13}(T)$ defined in (4.60), which is the last and most technical one to estimate for $M(T)$. We add and subtract an expression involving $Z^\eta*(\mu_{N,\eta}-\bar u^\eta)$ to the error term M_{13}, giving $M_{13}\le M_{131}+M_{132}$, where

$$
\begin{aligned}
M_{131}=2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t\Big(&\big\langle\mu_{N,\eta}-\bar u^\eta,|\nabla Z^\eta*(f^{N,\eta}-g^\eta)|^2\big\rangle \\
&-\big\langle Z^\eta*(\mu_{N,\eta}-\bar u^\eta),\nabla(f^{N,\eta}-g^\eta)\cdot\nabla Z^\eta*(f^{N,\eta}-g^\eta)\big\rangle\Big)\mathrm{d}s\bigg)
\end{aligned} \tag{4.68}
$$

$$
\begin{aligned}
M_{132}=2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t\Big(&\big\langle Z^\eta*(\mu_{N,\eta}-\bar u^\eta),\nabla(f^{N,\eta}-g^\eta)\cdot\nabla Z^\eta*(f^{N,\eta}-g^\eta)\big\rangle \\
&-\big\langle Z^\eta*(\mu_{N,\eta}-\bar u^\eta),|\nabla(f^{N,\eta}-g^\eta)|^2\big\rangle\Big)\mathrm{d}s\bigg).
\end{aligned} \tag{4.69}
$$

We start with M_{131}, which is estimated in a similar way as in [90], but we use the estimates (4.16) (Assumption (C1)) and Lemma 4.2 instead of a Taylor expansion as in [90], where the calculations for one space-dimension is done in case $\kappa = -1$.
We split M_{131} into several terms and the technical proof into several sub-steps. We abbreviate $w^{N,\eta} := Z^\eta * (f^{N,\eta} - g^\eta)$. The mean-value theorem for $\nabla w^{N,\eta}$ then gives

$$\nabla w^{N,\eta}(x) - \nabla w^{N,\eta}(x-y) = \mathrm{D}^2 w^{N,\eta}(\cdot + (-1+c^*)y)\cdot y = \int_{-1}^0 \mathrm{D}^2 w^{N,\eta}(\cdot + ry)dr \cdot y,$$

for some $c^* \in (0,1)$, where $\mathrm{D}^2 w^{N,\eta} = (\partial_{ij}^2 w^{N,\eta})_{i,j=1}^d$ denotes the Hessian matrix of $w^{N,\eta}$. The symmetry of Z^η allows us to apply Lemma 4.10:

$$\begin{aligned}
M_{131}(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \langle \mu_{N,\eta} - \bar{u}^\eta, |Z^\eta * \nabla(f^{N,\eta} - g^\eta)|^2 \\
&\quad - Z^\eta * (\nabla(f^{N,\eta} - g^\eta)\cdot \nabla w^{N,\eta})\rangle \mathrm{d}s\bigg) \\
&= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \Big\langle \mu_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y)\mathrm{d}y \cdot \nabla w^{N,\eta}(\cdot) \\
&\quad - \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y)\cdot \nabla w^{N,\eta}(\cdot - y)\mathrm{d}y\Big\rangle \mathrm{d}s\bigg) \\
&= 2\mathbb{E}\bigg(\sup_{0<t<\tau}\int_0^t \Big\langle \mu_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y)^T \\
&\quad \times \Big\{\int_{-1}^0 \mathrm{D}^2 w^{N,\eta}(\cdot + ry)\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg),
\end{aligned}$$

where we used the mean-value theorem in the last step. We expand $\partial_{ij}^2 w^{N,\eta}$ by adding and subtracting the empirical measure $\overline{\mu}_{N,\eta}$ of the intermediate system (4.6):

$$\partial_{ij}^2 w^{N,\eta} = \partial_{ij}^2 V^\eta * (\mu_{N,\eta} - \bar{u}^\eta) = \partial_{ij}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta}) + \partial_{ij}^2 V^\eta * (\mu_{N,\eta} - \bar{u}^\eta).$$

Then $M_{131} \le P_1 + P_2$, where

$$\begin{aligned}
P_1(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \Big\langle \mu_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y) \\
&\quad \times \Big\{\int_{-1}^0 \mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})(\cdot + ry)\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg), \\
P_2(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \Big\langle \mu_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y) \\
&\quad \times \Big\{\int_{-1}^0 \mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\cdot + ry)\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg).
\end{aligned} \tag{4.70}$$

Step 3.1: Estimation of P_1. We use the mean-field estimate in probability (4.16), which is assumed to hold for this theorem. Since $\bar{u}^\eta$ is nonnegative by Theorem 4.4 and $U^\eta(x) := Z^\eta(x)|x|$ is symmetric, by Lemma 4.10

$$P_1(T) \le 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \|\mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty}$$

$$\times \left\langle \mu_{N,\eta} + \bar{u}^\eta \int_{\mathbb{R}^d} U^\eta(y)|\nabla(f^{N,\eta} - g^\eta)(\cdot - y)|\mathrm{d}y \right\rangle \mathrm{d}s\Big) \tag{4.71}$$
$$\le 2\mathbb{E}\Big(\sup_{0<t<T} \int_0^t \|\mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty} \langle U^\eta * (\mu_{N,\eta} + \bar{u}^\eta), |\nabla(f^{N,\eta} - g^\eta)|\rangle \mathrm{d}s\Big)$$
$$\le 2\mathbb{E} \int_0^T \|\mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty} \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}$$
$$\times \big(\|\mu_{N,\eta} * U^\eta\|_{L^2} + \|\bar{u}^\eta * U^\eta\|_{L^2}\big)\mathrm{d}s,$$

and the last step follows from the Cauchy–Schwarz inequality. We claim that the convolutions with U^η are bounded by $CN^{\beta(d-2)}$. Indeed, by (4.38) and Cauchy-Schwarz ($(\sum_{i=1}^N x_i)^2 \le N\sum_{i=1}^N x_i^2$ for any $x_i \in \mathbb{R}$)

$$\|(\mu_{N,\eta} * U^\eta)(s)\|_{L^2}^2 = \int_{\mathbb{R}^d} \left(\frac{1}{N}\sum_{i=1}^N Z^\eta(x - X_i^{N,\eta}(s))|x - X_i^{N,\eta}(s)|\right)^2 \mathrm{d}x \tag{4.72}$$
$$\le \int_{\mathbb{R}^d} Z^\eta(x)^2|x|^2 \mathrm{d}x \le CN^{\beta(d-2)},$$
$$\|(\bar{u}^\eta * U^\eta)(s)\|_{L^2}^2 \le \|\bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}^2 N^{\beta(d-2)} \le CN^{\beta(d-2)}.$$

Inserting these estimates into (4.71) and splitting Ω into $\mathcal{C}_\alpha(s)$ and $\mathcal{C}_\alpha^c(s)$ (defined in (4.61)) yields $P_1 \le P_{11} + P_{12}$, where

$$P_{11}(T) = CN^{\beta(d-2)/2}\mathbb{E}\left(\int_0^T \mathbb{1}_{\mathcal{C}_\alpha^c(s)}\|\mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty}\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s\right),$$
$$P_{12}(T) = CN^{\beta(d-2)/2}\mathbb{E}\left(\int_0^T \mathbb{1}_{\mathcal{C}_\alpha(s)}\|\mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty}\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s\right).$$

Step 3.1a: Estimation of P_{11}. We compute

$$\big\|\mathrm{D}^2 V^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\big\|_{L^\infty} = \operatorname*{ess\,sup}_{x\in\mathbb{R}^d}\left|\frac{1}{N}\sum_{i=1}^N \mathrm{D}^2 V^\eta(x - X_i^{N,\eta}(s)) - \mathrm{D}^2 V^\eta(x - \overline{X}_i^\eta(s))\right|$$
$$\le \|\mathrm{D}^3 V^\eta\|_{L^\infty}\frac{1}{N}\sum_{i=1}^N |X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)|.$$

Together with $\|\mathrm{D}^3 V^\eta\|_{L^\infty} \le CN^{\beta(d+3)}$ from (4.37), the definition of $\mathcal{C}_\alpha^c(s)$, and Young's inequality, this shows that

$$P_{11}(T) \le CN^{\beta(d-2)/2+\beta(d+3)-\alpha}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2}\mathrm{d}s$$
$$\le \delta\mathbb{E}\int_0^t \|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2}^2\mathrm{d}s + C(\delta,T)N^{\beta(3d+4)-2\alpha}$$
$$\le \delta\mathbb{E}\int_0^t \|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2}^2\mathrm{d}s + C(\delta,T)N^{-1/2-\varepsilon},$$

if we choose $\alpha > 1/4 + \beta(3d+4)/2$ which is equivalent to $\beta(3d+4) - 2\alpha < -1/2$. The choice $\alpha > 1/4 + \beta(3d+4)/2$ is compatible with the assumptions of (4.16) since

$$\frac{1}{4} + \frac{\beta}{2}(3d+4) < \alpha < \frac{1}{2} - \beta(d+1) \tag{4.73}$$

is non-empty. This is guaranteed if $\beta(\frac{5}{2}d+3) < \frac{1}{4}$ which can be written as

$$\beta < \frac{1}{10d+12} = \frac{1}{2(5d+6)}. \tag{4.74}$$

Since we choose β in this way, there exists an $\alpha > 0$, that fulfils (4.73), and we can apply (4.16) and get for such an α that for any $\gamma > 0$ and $T > 0$ there exists a constant $C(\gamma, T)$ such that

$$\mathbb{P}(\mathcal{C}_\alpha(s)) \le C(\gamma, T)N^{-\gamma}, \tag{4.75}$$

which helps us in the next estimate:

Step 3.1b: Estimation of P_{12}. To estimate P_{12}, we apply the assumed convergence in probability (4.16). The assumptions are satisfied since $1/4 + \beta(3d+4)/2 < \alpha < 1/2 - \beta(d+1)$ is fulfilled. Therefore, because of $\mathbb{P}(\mathcal{C}_\alpha(s)) \le C(\gamma, T)N^{-\gamma}$ for any $\gamma > 0$, estimate $\|\mathrm{D}^2 V^\eta\|_{L^\infty} \le CN^{\beta(d+2)}$ from (4.37), and (4.39), we can estimate by Fubini's theorem

$$\begin{aligned} P_{12}(T) &\le CN^{\beta(d-2)/2}N^{\beta(d+2)}\mathbb{E}\int_0^T \mathbb{1}_{\mathcal{C}_\alpha(s)}\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s \\ &\le C(T)N^{\beta(3d+2)/2}N^{\beta(d+2)/2}\sup_{0<s<T}\mathbb{P}(\mathcal{C}_\alpha(s)) \le C(\gamma, T)N^{2\beta(d+1)-\gamma} \le C(T)N^{-1/2-\varepsilon}, \end{aligned}$$

where the last step follows if we choose $\gamma > 2\beta(d+1) + 1/2$. Collecting the estimates for P_{11} and P_{12} gives

$$P_1(T) \le C(\delta, T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2}^2\mathrm{d}s. \tag{4.76}$$

Step 3.2: Estimation of P_2. We continue by estimating $P_2(T)$, defined in (4.70). We add and subtract $\overline{\mu}_{N,\eta}$, giving $P_2 \le P_{21} + P_{22}$, where

$$\begin{aligned} P_{21}(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \Big\langle \mu_{N,\eta} - \overline{\mu}_{N,\eta}, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y) \\ &\qquad \times \Big\{\int_{-1}^0 \mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\cdot + ry)\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg), \\ P_{22}(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \Big\langle \overline{\mu}_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta} - g^\eta)(\cdot - y) \\ &\qquad \times \Big\{\int_{-1}^0 \mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\cdot + ry)\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg). \end{aligned} \tag{4.77}$$

Step 3.2a: Estimation of P_{21}. Using (4.35), we write out the dual bracket, and add and subtract a suitable expression in the second step:

$$\begin{aligned}
P_{21}(T) \le 2\mathbb{E}\Bigg(\sup_{0<t<T} \Bigg| \int_0^t \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} & Z^\eta(y)\nabla(f^{N,\eta}-g^\eta)(X_i^{N,\eta}-y) \\
& \times \left\{ \int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(X_i^{N,\eta}+ry)\mathrm{d}r \right\} y\mathrm{d}y\mathrm{d}s \\
& - \int_0^t \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta}-g^\eta)(\overline{X}_i^\eta - y) \\
& \times \left\{ \int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(\overline{X}_i^\eta+ry)\mathrm{d}r \right\} y\mathrm{d}y\mathrm{d}s \Bigg| \Bigg) \\
\le P_{211}(T) + P_{212}(T),
\end{aligned}$$

where

$$\begin{aligned}
P_{211}(T) &= 2\mathbb{E}\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} Z^\eta(y)\big|\nabla(f^{N,\eta}-g^\eta)(X_i^{N,\eta}-y) - \nabla(f^{N,\eta}-g^\eta)(\overline{X}_i^\eta - y)\big| \\
&\quad \times \left| \int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(X_i^{N,\eta}+ry)\mathrm{d}r \cdot y \right| \mathrm{d}y\mathrm{d}s, \\
P_{212}(T) &= 2\mathbb{E}\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} Z^\eta(y)\big|\nabla(f^{N,\eta}-g^\eta)(\overline{X}_i^\eta - y)\big| \\
&\quad \times \left| \int_{-1}^0 \Big(\mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(X_i^{N,\eta}+ry) \right. \qquad (4.78) \\
&\quad \left. - \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(\overline{X}_i^\eta+ry) \Big) \mathrm{d}r \cdot y \right| \mathrm{d}y\mathrm{d}s.
\end{aligned}$$

Splitting Ω again into the sets $\mathcal{C}_\alpha(s)$ and $\mathcal{C}_\alpha^c(s)$, defined in (4.61), the first term P_{211} can be estimated as

$$\begin{aligned}
P_{211}(T) &\le 4\mathbb{E}\left(\int_0^T \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{C}_\alpha(s)} \|\nabla(f^{N,\eta}-g^\eta)\|_{L^\infty} \|\mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)\|_{L^\infty} Z^\eta(y)|y|\mathrm{d}y\mathrm{d}s \right) \\
&+ 2\mathbb{E}\Bigg(\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{C}_\alpha^c(s)} Z^\eta(y)\big|\nabla(f^{N,\eta}-g^\eta)(X_i^{N,\eta}-y) - \nabla(f^{N,\eta}-g^\eta)(\overline{X}_i^\eta - y)\big| \\
&\times \left| \int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(X_i^{N,\eta}+ry)\mathrm{d}r \cdot y \right| \mathrm{d}y\mathrm{d}s \Bigg) =: Q_1(T) + Q_2(T).
\end{aligned}$$

For the first term, we estimate $\sup_{\omega\in\Omega}\sup_{0<s<T}\|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^\infty} \le CN^{\beta(d+1)}$ by using (4.39), $\sup_{\omega\in\Omega}\sup_{0<s<T}\|\mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta)(s)\|_{L^\infty} \le CN^{\beta(d+2)}$ by using (4.37), $\int_{\mathbb{R}^d} Z^\eta(y)|y|\mathrm{d}y \le CN^{-\beta}$ from (4.38), and finally, we choose $\gamma > 2\beta(d+1) - 1/2$:

$$Q_1(T) \le C(T)N^{\beta(d+1)+\beta(d+2)-\beta} \sup_{0<s<T} \mathbb{P}(\mathcal{C}_\alpha(s))$$

$$\leq C(T)N^{2\beta(d+1)-\gamma} \leq C(T)N^{-1/2-\varepsilon}. \tag{4.79}$$

Next, we set

$$\nabla z_i^{N,\eta}(y) := \nabla(f^{N,\eta} - g^\eta)(X_i^{N,\eta} - y) - \nabla(f^{N,\eta} - g^\eta)(\overline{X}_i^\eta - y)$$

and add and subtract $\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\overline{X}_i^\eta + ry)$ to the integrand of Q_2, leading to $Q_2 \leq Q_{21} + Q_{22}$, where

$$\begin{aligned} Q_{21}(T) = 2\mathbb{E}\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{C}_\alpha^c(s)} Z^\eta(y)|\nabla z_i^{N,\eta}(y)| \bigg| \int_{-1}^0 \big(\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(X_i^{N,\eta} + ry) \\ - \mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\overline{X}_i^\eta + ry)\big)\mathrm{d}r \cdot y \bigg| \mathrm{d}y\mathrm{d}s, \\ Q_{22}(T) = 2\mathbb{E}\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{C}_\alpha^c(s)} Z^\eta(y)|\nabla z_i^{N,\eta}(y)| \\ \times \bigg| \int_{-1}^0 (\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\overline{X}_i^\eta + ry)\mathrm{d}r \cdot y \bigg| \mathrm{d}y\mathrm{d}s. \end{aligned}$$

By definition (4.61) of $\mathcal{C}_\alpha^c(s)$, the estimate $\|\mathrm{D}^3 V^\eta\|_{L^\infty} \leq CN^{\beta(d+3)}$, the mean-value theorem applied to $\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)$, $\|\bar{u}^\eta\|_{L^1} = 1$ and recalling the definition $U^\eta(y) = Z^\eta(y)|y|$, we have

$$\begin{aligned} Q_{21}(T) &\leq C\mathbb{E}\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{C}_\alpha^c(s)} Z^\eta(y)|y||\nabla z_i^{N,\eta}(y)|\|\mathrm{D}^3 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)\|_{L^\infty} \\ &\qquad \times |X_i^{N,\eta} - \overline{X}_i^\eta|\mathrm{d}y\mathrm{d}s \\ &\leq CN^{-\alpha}\mathbb{E}\int_0^T \big\langle U^\eta * (\mu_{N,\eta} + \overline{\mu}_{N,\eta}), |\nabla(f^{N,\eta} - g^\eta)|\big\rangle \|\mathrm{D}^3 V^\eta\|_{L^\infty}\mathrm{d}s \\ &\leq C(T)N^{\beta(d+3)-\alpha}\big(\|\mu_{N,\eta} * U^\eta\|_{L^2} + \|\overline{\mu}_{N,\eta} * U^\eta\|_{L^2}\big)\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s \\ &\leq C(T)N^{\beta(d+3)-\alpha}N^{\beta(d-2)/2}\mathbb{E}\int_0^t \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s, \end{aligned}$$

where in the last step we used the bound (4.72), which is uniform in $\omega \in \Omega$, for the estimates for $\mu_{N,\eta} * U^\eta$ and $\bar{u}^\eta * U^\eta$. Hence, by Young's inequality,

$$\begin{aligned} Q_{21}(T) &\leq \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s + C(\delta,T)N^{\beta(3d+4)-2\alpha} \\ &\leq \delta\mathbb{E}\int_0^t \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s + C(T)N^{-1/2-\varepsilon}, \end{aligned} \tag{4.80}$$

after choosing $\alpha > 1/4 + \beta/2(3d+4)$, which is equivalent to $\beta(3d+4) - 2\alpha < -1/2$. This choice of α is admissible, see (4.73).

For the term Q_{22}, we introduce the set

$$\mathcal{E}_\theta(\overline{X}_i^\eta(s), y, r, s) := \Big\{\omega \in \Omega : \Big|\frac{1}{N}\sum_{i=1}^N \mathrm{D}^2 V^\eta(\overline{X}_i^\eta(s) + ry - \overline{X}_j^\eta(s)) - (\mathrm{D}^2 V^\eta * \bar{u}^\eta)(s, \overline{X}_i^\eta(s) + ry)\Big| > N^{-\theta}\Big\}$$
$$= \big\{\omega \in \Omega : \big|\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\overline{X}_i^\eta(s) + ry)\big| > N^{-\theta}\big\},$$

which corresponds to $\mathcal{B}_{\theta,\psi}^N(s)$ with $\psi_\eta(x) = \mathrm{D}^2 V^\eta(x + ry)$ in Lemma 4.2. As we integrate in Q_{22} over $\mathcal{C}_\alpha^c(s)$, we can use $\max_{i=1,\dots,N} |X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)| \le N^{-\alpha}$. Therefore, applying the mean-value theorem to $\nabla z_i^{N,\eta}$ and using $\sup_{\omega\in\Omega}\sup_{0<s<T} \|\mathrm{D}^2(f^{N,\eta} - g^\eta)(s)\|_{L^\infty} \le CN^{\beta(d+2)}$, $\sup_{\omega\in\Omega}\sup_{0<s<T} \|\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(s)\|_{L^\infty} \le CN^{\beta(d+2)}$ which can be proved similarly as (4.39), we find that

$$Q_{22}(T) \le CN^{-\alpha}\mathbb{E}\Big(\int_0^T \|\mathrm{D}^2(f^{N,\eta} - g^\eta)\|_{L^\infty} \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} Z^\eta(y)|y|$$
$$\times \int_{-1}^0 \big|\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\overline{X}_i^\eta + ry)\big|\big(\mathbb{1}_{\mathcal{E}_\theta(\overline{X}_i^\eta(s),y,r,s)} + \mathbb{1}_{\mathcal{E}_\theta^c(\overline{X}_i^\eta(s),y,r,s)}\big)\mathrm{d}r\mathrm{d}y\mathrm{d}s\Big)$$
$$\le CN^{-\alpha+2\beta(d+2)}\frac{1}{N}\sum_{i=1}^N \int_0^T \int_{\mathbb{R}^d}\int_{-1}^0 Z^\eta(y)|y|\mathbb{P}(\mathcal{E}_\theta^N(\overline{X}_i^\eta(s), y, r, s))\mathrm{d}r\mathrm{d}y\mathrm{d}s$$
$$+ C(T)N^{-\alpha+\beta(d+2)-\beta-\theta},$$

where we also used $\int_{\mathbb{R}^d} Z^\eta(y)|y|\mathrm{d}y \le CN^{-\beta}$ from (4.38) in the second term and Fubini's theorem in the first one. We deduce from Lemma 4.2 that for any $m \in \mathbb{N}$

$$\mathbb{P}(\mathcal{E}_\theta(\overline{X}_i^\eta(s), y, r, s)) \le C(m)\|\mathrm{D}^2 V^\eta\|_{L^\infty}^{2m} N^{2m(\theta-1/2)+1} \le CN^{2m(\theta-1/2+\beta(d+2))+1},$$

where we note that the constant $C > 0$ is independent of y, r and s. This leads to

$$Q_{22}(t) \le CN^{-\alpha+2\beta(d+2)-\beta}N^{2m(\theta-1/2+\beta(d+2))+1} + C(T)N^{\beta(d+1)-\alpha-\theta}$$

To bound the above right-hand side by $N^{-1/2-\varepsilon}$, we need the following conditions on α and θ:

(i) $\theta < 1/2 - \beta(d+2)$: Then $\theta - 1/2 + \beta(d+2)$ is negative and we can choose m large enough to obtain $2m(\theta - 1/2 + \beta(d+2)) + 1 < -1/2 - (\beta(2d+1) - \alpha)$. Note that we choose $m \in \mathbb{N}$ after choosing α.

(ii) $\alpha + \theta > 1/2 + \beta(d+2)$: Then $\beta(d+2) - \alpha - \theta < -1/2$.

We need to ensure that both conditions are compatible with the conditions on α (imposted in the estimation of P_{11}, see (4.73)) and β (imposed in the theorem):

$$0 < \beta < \frac{1}{10d+12} \quad \text{and} \quad \frac{1}{4} + \frac{\beta}{2}(3d+4) < \alpha < \frac{1}{2} - \beta(d+1).$$

We infer from (i), (ii), and $\alpha < 1/2 - \beta(d+1)$ that

$$\frac{1}{2} + \beta(d+1) < \alpha + \theta < 1 - \beta(2d+3).$$

This chain of inequalities is non-empty under the constraint $\beta < 1/(6d+8)$. Choosing α close to $1/4 + \beta(3d+4)/2$ and taking into account the smallness condition on β, we can always find an admissible value for $\theta > 0$. Hence, conditions (i) and (ii) can be fulfilled under the given condition on β.
We combine the inequality $Q_{22}(T) \leq C(T)N^{-1/2-\varepsilon}$ with estimate (4.80) for Q_{21} and estimate (4.79) for Q_1:

$$P_{211}(T) = (Q_1 + Q_{21} + Q_{22})(T) \leq C(T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s. \quad (4.81)$$

We turn to the term $P_{212}(T)$, defined in (4.78), and split Ω in $\mathcal{C}_\alpha(s)$ and $\mathcal{C}_\alpha^c(s)$, defined in (4.61). First, we observe the following two estimates: First, we obtain from the mean-value theorem and Lemma 4.7:

$$\begin{aligned}
&\big|\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(X_i^{N,\eta} + ry) - \mathrm{D}^2 V^\eta(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\overline{X}_i^\eta + ry)\big| \\
&\quad \leq |X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)| \sup_{\omega\in\Omega} \sup_{0<s<T} \|\mathrm{D}^3 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)\|_{L^\infty} \\
&\quad \leq CN^{\beta(d+3)}|X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)|.
\end{aligned}$$

This estimate is used on the set $\mathcal{C}_\alpha^c(s)$. Furthermore, we use on the set $\mathcal{C}_\alpha(s)$:

$$\sup_{\omega\in\Omega} \sup_{0<s<T} \|\mathrm{D}^2 V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(s)\|_{L^\infty} \leq CN^{\beta(d+2)}.$$

Recalling the definition $U^\eta(y) = Z^\eta(y)|y|$, this yields

$$\begin{aligned}
P_{212}(T) &\leq CN^{\beta(d+2)}\mathbb{E}\bigg(\int_0^T \mathbb{1}_{\mathcal{C}_\alpha(s)}\langle U^\eta * \overline{\mu}_{N,\eta}, |\nabla(f^{N,\eta} - g^\eta)|\rangle \mathrm{d}s\bigg) \\
&\quad + CN^{\beta(d+3)-\alpha}\mathbb{E}\bigg(\int_0^T \mathbb{1}_{\mathcal{C}_\alpha^c(s)}\langle U^\eta * \overline{\mu}_{N,\eta}, |\nabla(f^{N,\eta} - g^\eta)|\rangle \mathrm{d}s\bigg) \\
&\leq CN^{\beta(d+2)}\mathbb{E}\Big(\|U^\eta * \overline{\mu}_{N,\eta}\|_{L^2}\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\int_0^T \mathbb{1}_{\mathcal{C}_\alpha(s)}\mathrm{d}s\Big) \\
&\quad + CN^{\beta(d+3)-\alpha}\mathbb{E}\bigg(\|U^\eta * \overline{\mu}_{N,\eta}\|_{L^2}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s\bigg).
\end{aligned}$$

In view of $\|U^\eta * \overline{\mu}_{N,\eta}\|_{L^2} \leq CN^{\beta(d-2)/2}$ and $\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2} \leq CN^{\beta(d+2)/2}$ uniformly in $[0,T]$ and Ω (see (4.72) and (4.39), respectively) and using the Cauchy–Schwarz inequality,

$$\begin{aligned}
P_{212}(T) &\leq C(T)N^{\beta(d+2)+\beta(d-2)/2+\beta(d+2)/2} \sup_{0<s<T} \mathbb{P}(\mathcal{C}_\alpha(s)) \\
&\quad + CN^{\beta(d+3)+\beta(d-2)/2-\alpha}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s
\end{aligned}$$

$$
\begin{aligned}
&\leq C(T)N^{2\beta(d+1)-\gamma} + C(\delta)N^{\beta(3d+4)-2\alpha} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s \\
&\leq C(T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s,
\end{aligned}
$$

choosing as before $\alpha > 1/4 + \beta(3d+4)/2$ and $\gamma > 0$ sufficiently large.
We combine the estimates for P_{211} and P_{212} (see (4.81)):

$$
P_{21}(T) \leq C(T)N^{-1/2-\varepsilon} + 2\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s, \tag{4.82}
$$

which finishes Step 3.2a.

Step 3.2b: Estimation of P_{22}. To estimate P_{22}, defined in (4.77), we split Ω into the sets $\mathcal{F}_\theta(\cdot, y, r, s)$ and $\mathcal{F}_\theta^c(\cdot, y, r, s)$, where

$$
\begin{aligned}
\mathcal{F}_\theta(x,y,r,s) :=& \Bigg\{\omega \in \Omega : \bigg|\frac{1}{N}\sum_{i=1}^N \mathrm{D}^2V^\eta(x+ry-\overline{X}_i^\eta(s)) \\
&\qquad - (\mathrm{D}^2V^\eta * \bar{u}^\eta)(s, x+ry)\bigg| > N^{-\theta}\Bigg\}. \\
=& \Big\{\omega \in \Omega : \Big|\mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(x+ry)\Big| > N^{-\theta}\Big\}
\end{aligned}
$$

Then $P_{22} \leq P_{221} + P_{222}$, where

$$
\begin{aligned}
P_{221}(T) = 2\mathbb{E}\bigg(&\sup_{0<t<T}\int_0^t \Big\langle \overline{\mu}_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta}-g^\eta)(\cdot - y) \\
&\times \Big\{\int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\cdot + ry)\mathbb{1}_{\mathcal{F}_\theta^c(\cdot,y,r,s)}\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg), \\
P_{222}(T) = 2\mathbb{E}\bigg(&\sup_{0<t<T}\int_0^t \Big\langle \overline{\mu}_{N,\eta} - \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)\nabla(f^{N,\eta}-g^\eta)(\cdot - y) \\
&\times \Big\{\int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\cdot + ry)\mathbb{1}_{\mathcal{F}_\theta(\cdot,y,r,s)}\mathrm{d}r\Big\} y\mathrm{d}y\Big\rangle \mathrm{d}s\bigg).
\end{aligned}
$$

We estimate similarly as before, using the definition of $\mathcal{F}_\theta^c(\cdot, y, r, s)$ and $U^\eta(y) = Z^\eta(y)|y|$:

$$
\begin{aligned}
P_{221}(T) &\leq CN^{-\theta}\mathbb{E}\int_0^T \Big\langle \overline{\mu}_{N,\eta} + \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)|y||\nabla(f^{N,\eta}-g^\eta)|\mathrm{d}y\Big\rangle \mathrm{d}s \qquad (4.83)\\
&\leq CN^{-\theta}\mathbb{E}\int_0^T \big(\|U^\eta * \overline{\mu}_{N,\eta}\|_{L^2} + \|U^\eta * \bar{u}^\eta\|_{L^2}\big)\|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}\mathrm{d}s \\
&\leq CN^{\beta(d-2)/2-\theta}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}\mathrm{d}s \\
&\leq C(\delta,T)N^{\beta(d-2)-2\theta} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^2}^2\mathrm{d}s,
\end{aligned}
$$

where we used (4.72) and Young's inequality. Furthermore, recalling that $\|\mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)\|_{L^\infty} \le CN^{\beta(d+2)}$ uniformly in $[0,T]$ and Ω (see Lemma 4.7),

$$\begin{aligned} P_{222}(T) &\le 2\mathbb{E}\int_0^T \Big\langle \overline{\mu}_{N,\eta} + \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)|\nabla(f^{N,\eta} - g^\eta)(\cdot - y)| \\ &\quad \times \Big| \int_{-1}^0 \mathrm{D}^2V^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(\cdot + ry)\mathbb{1}_{\mathcal{F}_\theta(\cdot,y,r,s)}\mathrm{d}r \cdot y\Big|\mathrm{d}y\Big\rangle\mathrm{d}s \\ &\le CN^{\beta(d+2)}\mathbb{E}\int_0^T \Big\langle \overline{\mu}_{N,\eta} + \bar{u}^\eta, \int_{\mathbb{R}^d} Z^\eta(y)|\nabla(f^{N,\eta} - g^\eta)(\cdot - y)| \\ &\quad \times \int_{-1}^0 \mathbb{1}_{\mathcal{F}_\theta(\cdot,y,r,s)}\mathrm{d}r\mathrm{d}y\Big\rangle\mathrm{d}s. \end{aligned}$$

We apply Fubini's theorem and use the definition of $\overline{\mu}_{N,\eta}$ as well as the uniform bound $\|\nabla(f^{N,\eta} - g^\eta)\|_{L^\infty} \le CN^{\beta(d+1)}$ from (4.39). Writing out the dual bracket then leads with $\bar{u}^\eta \ge 0$ to

$$\begin{aligned} P_{222}(T) &\le CN^{\beta(d+2)}\mathbb{E}\Big(\int_{-1}^0\int_0^T \frac{1}{N}\sum_{i=1}^N \int_{\mathbb{R}^d} Z^\eta(y)|y|\mathbb{1}_{\mathcal{F}_\theta(\overline{X}_i^\eta(s),y,r,s)} \\ &\quad \times |\nabla(f^{N,\eta} - g^\eta)(\overline{X}_i^\eta(s) - y)|\mathrm{d}y\mathrm{d}s\mathrm{d}r\Big) \\ &\quad + CN^{\beta(d+2)}\mathbb{E}\Big(\int_{-1}^0\int_0^T\int_{\mathbb{R}^d} \bar{u}^\eta(x)\int_{\mathbb{R}^d} Z^\eta(y)|y|\mathbb{1}_{\mathcal{F}_\theta(x,y,r,s)} \\ &\quad \times |\nabla(f^{N,\eta} - g^\eta)(x - y)|\mathrm{d}y\mathrm{d}x\mathrm{d}s\mathrm{d}r\Big) \\ &\le CN^{\beta(2d+3)}\int_{-1}^0\int_0^T \frac{1}{N}\sum_{i=1}^N Z^\eta(y)|y|\mathbb{P}(\mathcal{F}_\theta^N(\overline{X}_i^\eta(s), y, r, s))\mathrm{d}y\mathrm{d}s\mathrm{d}r \\ &\quad + CN^{\beta(2d+3)}\int_{-1}^0\int_0^T\int_{\mathbb{R}^d} \bar{u}^\eta(x)\int_{\mathbb{R}^d} Z^\eta(y)|y|\mathbb{P}(\mathcal{F}_\theta(x,y,r,s))\mathrm{d}y\mathrm{d}x\mathrm{d}s\mathrm{d}r. \end{aligned} \tag{4.84}$$

Again, we wish to apply Lemma 4.2 to estimate the probability of $\mathcal{F}_\theta(\cdot, y, r, s)$. For fixed $x \in \mathbb{R}^d$, the set $\mathcal{F}_\theta(x, y, r, s)$ corresponds to $\mathcal{A}^N_{\theta,\phi_\eta}(s)$ with $\phi_\eta(z) = \mathrm{D}^2V^\eta(x + ry - z)$, while $\mathcal{F}_\theta(\overline{X}_i^\eta(s), y, r, s))$ corresponds to $\mathcal{B}^N_{\theta,\psi_\eta}$ with $\psi_\eta(z) = \mathrm{D}^2V^\eta(z + ry)$. By Lemma 4.2, for any $m \in \mathbb{N}$, there exists $C(m,T) > 0$ such that

$$\begin{aligned} \mathbb{P}(\mathcal{F}_\theta(\overline{X}_i^\eta(s), y, r, s)) &\le C(m,T)\|\mathrm{D}^2V^\eta(\cdot + ry)\|_{L^\infty}^{2m} N^{2m(\theta-1/2)+1} \\ &\le C(m,T)N^{2m(\theta-1/2+\beta(d+2))+1}, \\ \mathbb{P}(\mathcal{F}_\theta(x,y,r,s)) &\le C(m,T)\|\mathrm{D}^2V^\eta(x + ry - \cdot)\|_{L^\infty}^{2m} N^{2m(\theta-1/2)} \\ &\le C(m,T)N^{2m(\theta-1/2+\beta(d+2))}. \end{aligned}$$

Then, in view of (4.38) and $\|\bar{u}^\eta(s)\|_{L^1} = 1$, we deduce from (4.84) that

$$P_{222}(T) \le C(m,T)N^{\beta(2d+3)}N^{-\beta}N^{2m(\theta-1/2+\beta(d+2))+1}.$$

Together with estimate (4.83) for P_{221}, we obtain

$$P_{22}(T) \le C(m,T)N^{2m(\theta-1/2)+\beta(d+2))+2\beta(d+1)+1} \\ + C(\delta,T)N^{\beta(d-2)-2\theta} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s.$$

To finish the estimate for P_{22}, we need to choose θ for the set $\mathcal{F}_\theta(\cdot,y,r,s)$: We choose θ in this step such that

$$1/4+\beta(d-2)/2 < \theta < 1/2-\beta(d+2).$$

This is possible since $1/4+\beta(d-2)/2 < 1/2-\beta(d+2)$ is equivalent to $\beta < 1/(6d+4)$, and this is fulfilled by our assumptions. With this choice, $\beta(d-2)-2\theta < -1/2$ and $\theta-1/2+\beta(d+2)$ is negative such that, for sufficiently large $m \in \mathbb{N}$, $2m(\theta-1/2+\beta(d+2))+2\beta(d+1)+1 < -1/2$. We infer that

$$P_{22}(T) \le C(\delta,T)N^{-1/2-\varepsilon} + \delta\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s,$$

which finishes Step 3.2b.

It remains to add estimate (4.82) for P_{21} and the previous estimate to conclude that

$$P_2(T) = P_{21}(T) + P_{22}(T) \le C(\delta,T)N^{-1/2-\varepsilon} + 3\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s,$$

this finishes Step 3.2.
Conclusion of Step 3.1 and 3.2. Combining estimate (4.76) for P_1 and the previous estimate for P_2, we obtain for M_{131}, defined in (4.68):

$$M_{131}(T) = P_1(T) + P_2(T) \le C(\delta,T)N^{-1/2-\varepsilon} + 4\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s. \tag{4.85}$$

Step 3.3. Estimation of M_{132}. We consider the term $M_{132}(T)$, defined in (4.69). The estimation of this expression is similar to the previous steps 3.1 and 3.2, but the estimates are simpler. First, we add and subtract $\overline{\mu}_{N,\eta}(s)$ in M_{132} to split the expression in a mean-field part and a law-of-large-numbers part. Then $M_{132} \le R_1 + R_2$, where

$$R_1(T) = 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \big\langle Z^\eta * (\mu_{N,\eta}-\overline{\mu}_{N,\eta}), \nabla(f^{N,\eta}-g^\eta) \\ \times \big(\nabla Z^\eta * (f^{N,\eta}-g^\eta) - \nabla(f^{N,\eta}-g^\eta)\big)\big\rangle \mathrm{d}s\bigg),$$

$$R_2(T) = 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \big\langle Z^\eta * (\overline{\mu}_{N,\eta}-\bar{u}^\eta), \nabla(f^{N,\eta}-g^\eta) \\ \times \big(\nabla Z^\eta * (f^{N,\eta}-g^\eta) - \nabla(f^{N,\eta}-g^\eta)\big)\big\rangle \mathrm{d}s\bigg).$$

We start with R_1. By Young's convolution inequality,

$$\|\nabla Z^\eta * (f^{N,\eta}-g^\eta)\|_{L^2} \le \|Z^\eta\|_{L^1}\|\nabla(f^{N,\eta}-g^\eta)\|_{L^2} = \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}, \tag{4.86}$$

and splitting Ω into $\mathcal{C}_\alpha(s)$ and $\mathcal{C}_\alpha^c(s)$ (see (4.61) for the definition), we arrive at

$$\begin{aligned} R_1(T) &\le 4\mathbb{E}\bigg(\int_0^T \|Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty} \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \big(\mathbb{1}_{\mathcal{C}_\alpha(s)} + \mathbb{1}_{\mathcal{C}_\alpha^c(s)}\big)\mathrm{d}s\bigg) \\ &\le 4\mathbb{E}\bigg(\int_0^T \|\nabla Z^\eta\|_{L^\infty} \frac{1}{N}\sum_{i=1}^N |X_i^{N,\eta}(s) - \overline{X}_i^\eta(s)| \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathbb{1}_{\mathcal{C}_\alpha^c(s)}\mathrm{d}s\bigg) \\ &\quad + 4\sup_{\omega\in\Omega}\sup_{0<s<T}\big(\|Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})\|_{L^\infty}\|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\big)\mathbb{E}\int_0^T \mathbb{1}_{\mathcal{C}_\alpha(s)}\mathrm{d}s, \end{aligned}$$

where we used in the last step the mean-value theorem in the first integral on the right-hand side. It follows from $\|\nabla Z^\eta\|_{L^\infty} \le CN^{\beta(d+1)}$, (4.39), and $\sup_{\omega\in\Omega}\sup_{0<s<T}\|Z^\eta * (\mu_{N,\eta} - \overline{\mu}_{N,\eta})(s)\|_{L^\infty} \le CN^{\beta d}$ that

$$R_1(T) \le C(T)N^{\beta(d+1)-\alpha}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s + C(T)N^{2\beta(d+1)}\sup_{0<s<T}\mathbb{P}(\mathcal{C}_\alpha(s)).$$

We choose α such that the assumptions of (4.16) are fulfilled. This implies that $\alpha > \beta(d+1)$. Therefore, we have $C(T)N^{\beta(d+2)-\alpha} \le \delta$ for sufficiently large $N \in \mathbb{N}$. Moreover, by (4.16), $\mathbb{P}(\mathcal{C}_\alpha(s)) \le CN^{-\gamma}$ for any $\gamma > 0$. Choosing γ sufficiently large then leads to

$$R_1(T) \le C(T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s. \tag{4.87}$$

For $R_2(T)$, we need the law-of-large-numbers estimate. We split Ω into the sets $\mathcal{D}_\delta(s,x)$ and $\mathcal{D}_\delta^c(s,x)$, where we recall from definition (4.64) that $\mathcal{D}_\delta(s,x) = \{\omega \in \Omega : |Z^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)(s,x)| > \delta\}$. We write $R_2 = R_{21} + R_{22}$, where

$$\begin{aligned} R_{21}(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \big\langle Z^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)\mathbb{1}_{\mathcal{D}_\delta^c(s,\cdot)}, \nabla(f^{N,\eta} - g^\eta) \\ &\qquad \times \big(Z^\eta * \nabla(f^{N,\eta} - g^\eta) - \nabla(f^{N,\eta} - g^\eta)\big)\big\rangle\mathrm{d}s\bigg), \\ R_{22}(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\int_0^t \big\langle Z^\eta * (\overline{\mu}_{N,\eta} - \bar{u}^\eta)\mathbb{1}_{\mathcal{D}_\delta(s,\cdot)}, \nabla(f^{N,\eta} - g^\eta) \\ &\qquad \times \big(Z^\eta * \nabla(f^{N,\eta} - g^\eta) - \nabla(f^{N,\eta} - g^\eta)\big)\big\rangle\mathrm{d}s\bigg). \end{aligned}$$

We infer from the definition of the set $\mathcal{D}_\delta^c(s,\cdot)$ and (4.86) that

$$R_{21}(T) \le 4\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s. \tag{4.88}$$

The second term $R_{22}(T)$ is estimated similarly as $M_{122}(T)$ in Step 2 of the estimation of $M(T)$ (Law-of-large numbers estimate), see (4.65). We use $\|Z^\eta * \bar{u}^\eta\|_{L^\infty} \le \|\bar{u}^\eta\|_{L^1}\|Z^\eta\|_{L^\infty} \le CN^{\beta d}$ and Fubini's theorem:

$$R_{22}(T) \le CN^{\beta d}\int_0^T \|\sup_{\omega\in\Omega}\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \sup_{x\in\mathbb{R}^d}\mathbb{P}(\mathcal{D}_\delta(s,x))\mathrm{d}s.$$

Using Lemma 4.2 with $\theta = 0$ and $\psi_\eta(y) = Z^\eta(x-y)$, for any $m \in \mathbb{N}$

$$\sup_{0<s<T} \sup_{x\in\mathbb{R}^d} \mathbb{P}(\mathcal{D}_\delta(s,x)) \le C(m)\|Z^\eta(x-\bullet)\|_{L^\infty}^{2m} N^{-m} \le C(m)N^{-m(1-2\beta d)}.$$

Using (4.66), we find that $R_{22}(T) \le C(T)N^{\beta d}N^{\beta(d+2)}N^{-m(1-2\beta d)} = CN^{2\beta(d+1)-m(1-2\beta d)}$. Since $1-2\beta d < 0$, we can choose m sufficiently large to obtain $R_{22}(T) \le CN^{-1/2-\varepsilon}$. Together with estimate (4.88) of R_{21}, it follows that

$$R_2(T) \le R_{21}(T) + R_{22}(T) \le CN^{-1/2-\varepsilon} + 4\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s.$$

In view of estimate (4.87) of R_1, this finishes our estimate for M_{132}:

$$M_{132}(T) \le R_1(T) + R_2(T) \le CN^{-1/2-\varepsilon} + 5\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s.$$

Finally, we conclude from (4.85) that

$$M_{13}(T) \le M_{131}(T) + M_{132}(T) \le CN^{-1/2-\varepsilon} + 9\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s. \tag{4.89}$$

Finishing the estimate of $M(T)$. We collect estimate (4.63) of M_{11}, estimate (4.67) of M_{12} and estimate (4.89) of M_{13}:

$$M_1(T) \le (M_{11}+M_{12}+M_{13})(T) \le CN^{-1/2-\varepsilon} + 12\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s.$$

Adding this inequality to estimate (4.59) for M_2, we conclude that

$$M(T) \le M_1(T) + M_2(T) \le CN^{-1/2-\varepsilon} + 14\delta\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s. \tag{4.90}$$

4. Estimation of $L(T)$. An expression like $L(T)$, defined in (4.55), has been estimated in [90] using a Taylor approximation and Fourier estimates in one space dimension. This approach is feasible also in higher space dimensions but it would become very tedious in notation. Additionally, we could reduce the assumption on the potential V^η in comparison to [90] since we do not need any assumption on the Fourier transform of the potential.
Our idea is, as above, to split the integral over Ω in a mean-field part and a law-of-large-numbers part.
We add and subtract the empirical measure $\overline{\mu}_{N,\eta}(s)$ of the intermediate problem (4.6) to $L(T)$, defined in (4.55). Then $|L(T)| \le L_1(T) + L_2(T)$, where

$$\begin{aligned} L_1(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\bigg|\int_0^t \langle \mu_{N,\eta}-\overline{\mu}_{N,\eta}, (\nabla V^\eta * \bar{u}^\eta)\cdot(\nabla Z^\eta * (f^{N,\eta}-g^\eta))\rangle \mathrm{d}s\bigg|\bigg), \\ L_2(T) &= 2\mathbb{E}\bigg(\sup_{0<t<T}\bigg|\int_0^t \langle \overline{\mu}_{N,\eta}-\bar{u}^\eta, (\nabla V^\eta * \bar{u}^\eta)\cdot(\nabla Z^\eta * (f^{N,\eta}-g^\eta))\rangle \mathrm{d}s\bigg|\bigg). \end{aligned} \tag{4.91}$$

The term L_1 can be considered as the mean-field part, while L_2 is the law-of-large-numbers part.

Step 1. Estimation of L_1. We start with $L_1(T)$ and add and subtract the term $(\nabla V^\eta * \bar u^\eta)(X_i^{N,\eta})(\nabla Z^\eta * (f^{N,\eta} - g^\eta))(\overline{X}_i^\eta)$, leading to $L_1(T) \le L_{11}(T) + L_{12}(T)$, where

$$\begin{aligned} L_{11}(T) &= 2\mathbb{E}\int_0^T \bigg| \frac{1}{N}\sum_{i=1}^N \big((\nabla V^\eta * \bar u^\eta)(s, X_i^{N,\eta}(s)) - (\nabla V^\eta * \bar u^\eta)(s,\overline{X}_i^\eta(s))\big) \\ &\qquad \times (\nabla Z^\eta * (f^{N,\eta} - g^\eta))(s, \overline{X}_i^\eta(s))\bigg| \mathrm{d}s, \\ L_{12}(T) &= 2\mathbb{E}\int_0^T \bigg| \frac{1}{N}\sum_{i=1}^N (\nabla V^\eta * \bar u^\eta)(s, X_i^{N,\eta}(s)) \\ &\qquad \times \big((\nabla Z^\eta * (f^{N,\eta} - g^\eta))(s, X_i^{N,\eta}(s)) - (\nabla Z^\eta * (f^{N,\eta} - g^\eta))(s, \overline{X}_i^\eta(s))\big)\bigg| \mathrm{d}s. \end{aligned}$$

For $L_{11}(T)$, we split Ω for each time $0 < s < T$ into the sets $\mathcal{C}_\alpha^c(s)$ and $\mathcal{C}_\alpha(s)$, defined in (4.61), use the definition of $\mathcal{C}_\alpha^c(s)$ in the first term (leading to the factor $N^{-\alpha}$) and Lemma 4.2 in the second term (leading to the factor $N^{-\gamma}$ for any $\gamma > 0$). Then, by the mean-value theorem applied to $\nabla V^\eta * \bar u^\eta$,

$$\begin{aligned} L_{11}(T) &\le CN^{-\alpha}\|V^\eta * \mathrm{D}^2\bar u^\eta\|_{L^\infty}\mathbb{E}\int_0^T \frac{1}{N}\sum_{i=1}^N \Big|(Z^\eta * \nabla(f^{N,\eta} - g^\eta))(s, \overline{X}_i^\eta(s))\Big| \mathrm{d}s \\ &\quad + C(T)\sup_{\omega\in\Omega}\sup_{0<s<T}\|Z^\eta * \nabla(f^{N,\eta} - g^\eta)\|_{L^\infty}\|V^\eta * \nabla\bar u^\eta\|_{L^\infty(0,T;L^\infty)}\sup_{0<s<T}\mathbb{P}(\mathcal{C}_\alpha(s)). \end{aligned}$$

We infer from Young's convolution inequality and Theorem 4.4 that for $k = 1, 2$,

$$\|V^\eta * \mathrm{D}^k\bar u^\eta\|_{L^\infty} \le \|V^\eta\|_{L^1}\|\mathrm{D}^k\bar u^\eta\|_{L^\infty} \le C, \tag{4.92}$$

since $H^s(\mathbb{R}^d) \hookrightarrow W^{1,\infty}(\mathbb{R}^d)$. Moreover, by definition of $\overline{\mu}_{N,\eta}$ and the symmetry of $Z^\eta \ge 0$, we have

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^N \Big|(Z^\eta * \nabla(f^{N,\eta} - g^\eta))(s, \overline{X}_i^\eta(s))\Big| &= \big\langle \overline{\mu}_{N,\eta}(s), \big|Z^\eta * \nabla(f^{N,\eta} - g^\eta)(s)\big|\big\rangle \\ &\le \big\langle \overline{\mu}_{N,\eta}(s), Z^\eta * \big|\nabla(f^{N,\eta} - g^\eta)(s)\big|\big\rangle = \big\langle Z^\eta * \overline{\mu}_{N,\eta}(s), |\nabla(f^{N,\eta} - g^\eta)(s)|\big\rangle \\ &\le \|Z^\eta * \overline{\mu}_{N,\eta}(s)\|_{L^2}\|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2} \le C(T)N^{\beta d/2}\|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2} \end{aligned}$$

uniformly in $s \in [0, T]$, where we used (4.40) in the last step. Therefore, in view of the uniform bound $\|\nabla(f^{N,\eta} - g^\eta)\|_{L^\infty} \le CN^{\beta(d+1)}$ (see (4.39)) and Young's inequality,

$$\begin{aligned} L_{11}(T) &\le C(T)N^{-\alpha+\beta d/2}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s \\ &\quad + C(T)N^{-\gamma}\sup_{\omega\in\Omega}\sup_{0<s<T}\|Z^\eta\|_{L^1}\|\nabla(f^{N,\eta} - g^\eta)\|_{L^\infty} \\ &\le C(T)N^{-\alpha+\beta d/2}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\mathrm{d}s + C(T)N^{\beta(d+1)-\gamma} \end{aligned}$$

$$\leq \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s + C(T,\delta)N^{-2\alpha+\beta d} + C(T)N^{\beta(d+1)-\gamma}.$$

Now, we choose $\alpha > 1/4+\beta d/2$ (which is consistent with the assumptions made in (4.16)) and $\gamma > 0$ sufficiently large to arrive at

$$L_{11}(T) \leq C(T,\delta)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2 \mathrm{d}s. \tag{4.93}$$

For the term $L_{12}(T)$, we split Ω again into $\mathcal{C}_\alpha(s)$ and $\mathcal{C}_\alpha^c(s)$ and we estimate similarly as above. Using the mean-value theorem, estimate $\|V^\eta * \nabla\bar{u}^\eta\|_{L^\infty} \leq C$ from (4.92),

$$\sup_{\omega\in\Omega}\sup_{0<s<T} \|\nabla Z^\eta * (f^{N,\eta}-g^\eta)(s)\|_{L^\infty} \leq \sup_{\omega\in\Omega}\sup_{0<s<T} \|\nabla(f^{N,\eta}-g^\eta)(s)\|_{L^\infty} \leq CN^{\beta(d+1)}$$

due to Young's convolution inequality and (4.39), as well as $\|\nabla Z^\eta\|_{L^2} \leq CN^{\beta(d+1)/2}$ from (4.37), we see that

$$\begin{aligned}
L_{12}(T) &\leq N^{-\alpha}\|V^\eta * \nabla\bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}\mathbb{E}\int_0^T \|\mathrm{D}^2 Z^\eta * (f^{N,\eta}-g^\eta)\|_{L^\infty}\mathrm{d}s \\
&\quad + C(T)\|V^\eta * \nabla\bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}N^{\beta(d+1)-\gamma} \\
&\leq C(T)N^{-\alpha}\mathbb{E}\int_0^T \|\nabla Z^\eta\|_{L^2}\|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}\mathrm{d}s + C(T)N^{\beta(d+1)-\gamma} \\
&\leq \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2\mathrm{d}s + C(T,\delta)N^{-2\alpha+\beta(d+1)} + C(T)N^{\beta(d+1)-\gamma}.
\end{aligned}$$

Again, choosing $\alpha > 1/4+\beta(d+1)/2$ (which is a possible choice in (4.16)) and $\gamma > 0$ sufficiently large, we infer that

$$L_{12}(T) \leq C(T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2\mathrm{d}s.$$

Together with estimate (4.93) for $L_{11}(t)$, we conclude that

$$L_1(T) = L_{11}(T) + L_{12}(T) \leq C(T)N^{-1/2-\varepsilon} + 2\delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2\mathrm{d}s. \tag{4.94}$$

Step 2. Estimation of L_2. The last step for estimating $L(T)$ is to derive suitable estimates for $L_2(T)$ defined in (4.91). To simplify the presentation, we abuse the notation by using an integral notation instead of the dual product in

$$\begin{aligned}
\int_{\mathbb{R}^d} Z^\eta(x-y)\mu_{N,\eta}(y)\nabla V^\eta(y-z)\mathrm{d}y &:= \int_{\mathbb{R}^d} Z^\eta(x-y)\nabla V^\eta(y-z)\mathrm{d}\mu_{N,\eta}(y) \\
&:= \frac{1}{N}\sum_{i=1}^n Z^\eta(x-\overline{X}_i^\eta)\nabla V^\eta(\overline{X}_i^\eta - z).
\end{aligned}$$

In this way, we can easier keep track of the variables. With this notation, we can re-write the integrand of L_2 by exploiting the symmetry of Z^η:

$$\langle\overline{\mu}_{N,\eta} - \bar{u}^\eta, (\nabla Z^\eta * \nabla(f^{N,\eta}-g^\eta)\nabla V^\eta * \bar{u}^\eta\rangle$$

$$= \langle Z^\eta * (\nabla V^\eta * \bar{u}^\eta(\overline{\mu}_{N,\eta} - \bar{u}^\eta)), \nabla(f^{N,\eta} - g^\eta)\rangle.$$

where in integral-notation

$$\begin{aligned}&\langle Z^\eta * (\nabla V^\eta * \bar{u}^\eta(\overline{\mu}_{N,\eta} - \bar{u}^\eta)), \nabla(f^{N,\eta} - g^\eta)\rangle \\ &= \int_{\mathbb{R}^d} (\nabla f^{N,\eta}(x) - \nabla g^\eta(x)) \int_{\mathbb{R}^d} Z^\eta(y-x)\nabla V^\eta * \bar{u}^\eta(y)(\overline{\mu}_{N,\eta}(y) - \bar{u}^\eta(y))\mathrm{d}y\mathrm{d}x.\end{aligned}$$

Writing out the convolution $\nabla V^\eta * \bar{u}^\eta$ and applying the Cauchy–Schwarz inequality for the integral in x, we find that

$$\begin{aligned}L_2(T) \le 2\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{\mathbb{R}^d}\bigg|\int_{\mathbb{R}^d} \bar{u}^\eta(z) \\ \times \int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\mathrm{d}z\bigg|^2\mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg).\end{aligned}$$

To estimate L_2 further, we define for some $\theta > 0$ the set

$$\mathcal{G}_\theta(x,z,s) := \bigg\{\omega \in \Omega : \bigg|\int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\bigg| > N^{-\theta}\bigg\},$$

which corresponds to the set $\mathcal{A}^N_{\theta,\phi_\eta}(s)$ from Lemma 4.2 with $\phi_\eta(y) = Z^\eta(x-y)\nabla V^\eta(y-z)$. We infer from this lemma that for any $m \in \mathbb{N}$, there exists $C(m) > 0$ such that

$$\mathbb{P}(\mathcal{G}_\theta(x,z,s)) \le C(m)\|\nabla V^\eta\|_{L^\infty}^{2m}\|Z^\eta\|_{L^\infty}^{2m} N^{2m(\theta-1/2)} \le C(m)N^{2m(\theta-1/2+\beta(2d+1))}, \tag{4.95}$$

where the last inequality follows from (4.37), and this bound is uniform in (x,z,s). We split the z-integral in $B_1(0)$ and $B_1(0)^c$ and the expectation in $\mathcal{G}_{\theta_k}(x,z,s)$ and $\mathcal{G}^c_{\theta_k}(x,z,s)$ for two different choices of θ_k, where $k = 1,2$. Then $L_2(T) \le L_{21}(T) + L_{22}(T)$, where

$$\begin{aligned}L_{21}(T) &= 2\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{\mathbb{R}^d}\bigg|\int_{B_1(0)} \bar{u}^\eta(z)\big(\mathbb{1}_{\mathcal{G}^c_{\theta_1}(x,z,s)} + \mathbb{1}_{\mathcal{G}_{\theta_1}(x,z,s)}\big) \\ &\quad\times \int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\mathrm{d}z\bigg|^2\mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg) \\ L_{22}(T) &= 2\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{\mathbb{R}^d}\bigg|\int_{B_1(0)^c} \bar{u}^\eta(z)\big(\mathbb{1}_{\mathcal{G}^c_{\theta_2}(x,z,s)} + \mathbb{1}_{\mathcal{G}_{\theta_2}(x,z,s)}\big) \\ &\quad\times \int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\mathrm{d}z\bigg|^2\mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg).\end{aligned}$$

We start with the term $L_{21}(T)$, stressing the fact that we integrate over $z \in B_1(0)$. Since V^η and Z^η have compact support in a ball of radius $\eta = N^{-\beta} < 1$, it is sufficient to integrate in y over $|y| < 2$, as otherwise $|y - z| > 1$ and consequently, $\nabla V^\eta(y-z) = 0$. Then it is sufficient to integrate in x over $|x| < 3$, as otherwise $|x-y| > 1$ and thus $Z^\eta(x-y) = 0$. Hence, with the definition of $\mathcal{G}^c_{\theta_1}(x,z,s)$, we have $L_{21}(T) \le L_{211}(T) + L_{212}(T)$, where

$$L_{211}(T) = CN^{-\theta}\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{B_3(0)}\bigg(\int_{B_1(0)} \bar{u}^\eta(z)\mathrm{d}z\bigg)^2\mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg),$$

$$L_{212}(T) = C\|Z^\eta\|_{L^\infty}\|\nabla V^\eta\|_{L^\infty}\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2} \times \bigg(\int_{B_3(0)}\bigg(\int_{B_1(0)} \bar{u}^\eta(z)\mathbb{1}_{\mathcal{G}_{\theta_1}(x,z,s)}\mathrm{d}z\bigg)^2\mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg).$$

For the first term, we simply use Young's inequality:

$$\begin{aligned} L_{211}(T) &\leq C(\delta,T)N^{-2\theta_1} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s \\ &\leq C(\delta,T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s, \end{aligned}$$

choosing $\theta_1 > 1/4$ (which is possible; see below). For the second term L_{212}, we use estimate (4.37) for Z^η and ∇V^η, Hölder's inequality, estimate (4.39) for $\nabla(f^{N,\eta} - g^\eta)$, the bound $\|\bar{u}^\eta\|_{L^\infty} \leq C$ from Theorem 4.4, and Lemma 4.2:

$$\begin{aligned} L_{212}(T) &\leq CN^{\beta d}N^{\beta(d+1)}N^{\beta(d+2)/2}\mathbb{E}\int_0^T\bigg(\int_{B_3(0)}\int_{B_1(0)}\big(\bar{u}^\eta(z)\mathbb{1}_{\mathcal{G}_{\theta_1}(x,z,s)}\big)^2\mathrm{d}z\mathrm{d}x\bigg)^{1/2}\mathrm{d}s \\ &\leq C(m,T)N^{\beta(5d+4)/2}N^{2m(\theta_1-1/2+\beta(2d+1))} \leq C(m,T)N^{-1/2-\varepsilon}, \end{aligned}$$

where we used the uniform probability estimate (4.95) for $\mathcal{G}_{\theta_1}(x,z,s)$, $\|\bar{u}^\eta(s)\|_{L^2} \leq C$ from Theorem 4.4 as well as the fact that we integrate over a bounded domain in the x-variable. The last inequality is possible since we can choose $\theta_1 > 0$ such that $\theta_1 - 1/2 + \beta(2d+1) < 0$. and $m \in \mathbb{N}$ large enough.
Remember that for L_{211} we need to choose $\theta_1 > 1/4$. Both conditions $1/4 < \theta_1 < 1/2 - \beta(2d+1)$ can be satisfied since $\beta < 1/(8d+4)$. This shows that

$$L_{21}(T) \leq C(T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2\mathrm{d}s, \tag{4.96}$$

which finishes the estimate for $L_{21}(T)$.
Next, we estimate L_{22}. To control the integrals over the far-field $B_1(0)^c$, we take advantage of the boundedness of the $(d+1)$th moment of $\bar{u}^\eta$, stated in Theorem 4.4. Since V^η and Z^η have compact support in a ball of radius $N^{-\beta}$, which is arbitrarily small for sufficiently large N – with similar arguments as for L_{21} – if $|z| > 1$ we can integrate in y over $|y| > 1/2$, as otherwise $|y - z| > 1/2$ and $\nabla V^\eta(y - z) = 0$ for sufficiently large N. Moreover, we can integrate in x over $|x| > 1/3$, as otherwise $|x - y| > 1/6$ and $Z^\eta(x - y) = 0$ for N large enough.
Additionally, due to the compact support $Z^\eta(x - y)\nabla V^\eta(y - z) = 0$ if $|x - y| \geq N^{-\beta}$ or $|y - z| \geq N^{-\beta}$ for sufficiently large N. Thus, it is sufficient to integrate over $|x - z| \leq |x - y| + |y - z| < 2N^{-\beta}$.
With these considerations, we can write $L_{22}(T) \leq L_{221}(T) + L_{222}(T)$, where

$$L_{221}(T) = C\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{B_{1/3}(0)^c}\bigg|\int_{B_1(0)^c} \bar{u}^\eta(z)\big(\mathbb{1}_{\mathcal{G}^c_{\theta_2}(x,z,s)}\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}$$

$$\times \int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\mathrm{d}z\bigg|^2 \mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg),$$

$$L_{222}(T) = C\mathbb{E}\bigg(\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{B_{1/3}(0)^c}\bigg|\int_{B_1(0)^c} \bar{u}^\eta(z)\big(\mathbb{1}_{\mathcal{G}_{\theta_2}(x,z,s)}\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}$$

$$\times \int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\mathrm{d}z\bigg|^2 \mathrm{d}x\bigg)^{1/2}\mathrm{d}s\bigg).$$

It follows from the definition of $\mathcal{G}_{\theta_2}(x,z,s)$ that

$$L_{221}(T)$$
$$\leq CN^{-\theta_2}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}\bigg(\int_{B_{1/3}(0)^c}\bigg|\int_{B_1(0)^c} \bar{u}^\eta(z)\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}\mathrm{d}z\bigg|^2\mathrm{d}x\bigg)^{1/2}\mathrm{d}s.$$

Since $\bar{u}^\eta$ is a probability density function, for fixed $x \in \mathbb{R}^d$, the inner integral can be estimated as

$$\bigg|\int_{B_1(0)^c} \bar{u}^\eta(z)\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}\mathrm{d}z\bigg|^2 \leq \bigg(\int_{B_1(0)^c}\bar{u}^\eta(z)\mathrm{d}z\bigg)\bigg(\int_{B_1(0)^c}\bar{u}^\eta(z)\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}\mathrm{d}z\bigg)$$
$$\leq \int_{B_1(0)^c}\bar{u}^\eta(z)\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}\mathrm{d}z.$$

In view of $|x|/|z| \leq (|x-z|+|z|)/|z| < 2N^{-\beta}+1 \leq C$ for $|x-z|<2N^{-\beta}$ and $|z|>1$, we have

$$\int_{B_{1/3}(0)^c}\bigg|\int_{B_1(0)^c}\bar{u}^\eta(s,z)\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}\mathrm{d}z\bigg|^2\mathrm{d}x \leq \int_{B_{1/3}(0)^c}\int_{B_1(0)^c}\bar{u}^\eta(s,z)\mathbb{1}_{\{|x-z|<2N^{-\beta}\}}\mathrm{d}z\mathrm{d}x$$
$$\leq C\int_{B_{1/3}(0)^c}\frac{\mathrm{d}x}{|x|^{d+1}}\int_{B_1(0)^c}|z|^{d+1}\bar{u}^\eta(s,z)\mathrm{d}z \leq C(d),$$

since $\bar{u}^\eta$ has a bounded $(d+1)$st moment and $\int_{B_{1/3}(0)^c}|x|^{-(d+1)}\mathrm{d}x < \infty$. This estimate allows us to conclude for $L_{221}(T)$ as follows by Young's inequality, choosing $\theta_2 > 1/4$:

$$L_{221}(T) \leq C(d)N^{-\theta_2}\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}\mathrm{d}s$$
$$\leq C(\delta,T)N^{-1/2-\varepsilon} + \delta\mathbb{E}\int_0^T \|\nabla(f^{N,\eta}-g^\eta)\|_{L^2}^2\mathrm{d}s.$$

The remaining term $L_{222}(T)$ is treated in a similar way. First, we notice that for fixed $x, z \in \mathbb{R}^d$

$$\bigg|\int_{\mathbb{R}^d} Z^\eta(x-y)(\overline{\mu}_{N,\eta} - \bar{u}^\eta)(y)\nabla V^\eta(y-z)\mathrm{d}y\bigg| \leq \|Z^\eta\nabla V^\eta\|_{L^\infty}(1+\|\bar{u}^\eta\|_{L^1})$$

and hence by using (4.37) and the uniform estimate (4.39):

$$L_{222}(T) \leq C\|Z^\eta\|_{L^\infty}\|\nabla V^\eta\|_{L^\infty}N^{\beta(d+2)/2}$$

$$\times \mathbb{E}\bigg(\int_0^T \bigg(\int_{B_{1/3}(0)^c} \bigg(\int_{B_1(0)^c} \bar{u}^\eta(z) \mathbb{1}_{\mathcal{G}_{\theta_2}(x,z,s)} \mathbb{1}_{\{|x-z|<2N^{-\beta}\}} \mathrm{d}z \bigg)^2 \mathrm{d}x \bigg)^{1/2} \mathrm{d}s \bigg)$$
$$\leq C(T) N^{\beta(2d+1)} N^{\beta(d+2)/2} \mathbb{E}\bigg(\int_0^T \bigg(\int_{B_{1/3}(0)^c} \frac{\mathrm{d}x}{|x|^{d+1}} \times \int_{B_1(0)^c} |z|^{d+1} \bar{u}^\eta(s,z) \mathbb{1}_{\mathcal{G}_{\theta_2}(x,z,s)} \mathrm{d}z \bigg)^{1/2} \mathrm{d}s \bigg).$$

Thus, Fubini's theorem, Jensen's inequality for $\sqrt{\cdot}$ and using the probability estimate (4.95) for $\mathcal{G}_{\theta_2}(x,z,s)$, which is uniform in (x,z,s), gives for any $m \in \mathbb{N}$

$$L_{222}(T) \leq C(T) N^{\beta(5d+4)/2} N^{m(\theta_2 - 1/2 + \beta(2d+1))} \leq C(T) N^{-1/2-\varepsilon},$$

choosing $1/4 < \theta_2 < 1/2 - \beta(2d+1)$ (such that $\theta_2 - 1/2 + \beta(2d+1) < 0$) and sufficiently large $m \in \mathbb{N}$. Combining the estimates for L_{221} and L_{222} leads to

$$L_{22}(T) \leq L_{221}(T) + L_{222}(T) \leq C(\delta,T) N^{-1/2-\varepsilon} + \delta \mathbb{E} \int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s. \tag{4.97}$$

Finally, we collect estimate (4.96) for L_{21} and estimate (4.97) for L_{22},

$$L_2(T) \leq L_{21}(T) + L_{22}(T) \leq C(\delta,T) N^{-1/2-\varepsilon} + 2\delta \mathbb{E} \int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s$$

and add this inequality to estimate (4.94) for L_1:

$$L(T) \leq L_1(T) + L_2(T) \leq C(\delta,T) N^{-1/2-\varepsilon} + 4\delta \mathbb{E} \int_0^T \|\nabla(f^{N,\eta} - g^\eta)\|_{L^2}^2 \mathrm{d}s. \tag{4.98}$$

5. Conclusion. We insert estimates (4.57) for K_1, (4.58) for K_6, (4.90) for M, and (4.98) for L into (4.29) to obtain

$$\mathbb{E}\bigg(\sup_{0<t<T} \|(f^{N,\eta} - g^\eta)(t)\|_{L^2}^2 \bigg) + 2(\sigma - 14\delta) \mathbb{E} \int_0^T \|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2}^2 \mathrm{d}s$$
$$\leq \mathbb{E}\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 + C(T) N^{\beta(d+2)-1} + \frac{C(\sigma)}{N} + C(T,\delta) N^{-1/2-\varepsilon}.$$

Since $\beta < 1/(10d+12) < 1/(2d+4)$, we have $\beta(d+2) - 1 < -1/2$. If (4.18) holds, i.e. $\mathbb{E}\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 \leq C N^{-1/2-\varepsilon_0}$, we obtain, after taking $\delta \leq \sigma/28$,

$$\mathbb{E}\bigg(\sup_{0<t<T} \|f^{N,\eta}(t) - g^\eta(t)\|_{L^2}^2 \bigg) + \sigma \mathbb{E} \int_0^T \|\nabla(f^{N,\eta} - g^\eta)(s)\|_{L^2}^2 \mathrm{d}s \leq C(T,\sigma) N^{-1/2-\varepsilon},$$

for some $\varepsilon > 0$ which proves the desired estimate.
It remains to verify (4.18). For this, we can argue similarly as in the beginning of the proof, see (4.44) and below. We write

$$\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 = \|f^{N,\eta}(0)\|_{L^2}^2 - 2\langle f^{N,\eta}(0), g^\eta(0)\rangle + \|g^\eta(0)\|_{L^2}^2$$

$$= \frac{1}{N^2}\sum_{i,j=1}^{N} V^\eta(X_i^{N,\eta}(0) - X_j^{N,\eta}(0)) - \frac{2}{N}\sum_{i=1}^{N}(V^\eta * u_0)(X_i^{N,\eta}(0)) + \langle u_0, V^\eta * u_0\rangle.$$

Since $X_i^{N,\eta}(0) = \zeta_i$ and since $\zeta_1, \ldots, \zeta_N$ are independent with common density function u_0, we infer that

$$\begin{aligned}
\mathbb{E}\|(f^{N,\eta} - g^\eta)(0)\|_{L^2}^2 &= \frac{1}{N^2}\sum_{i,j=1,\, i\neq j}^{N}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} V^\eta(x-y)u_0(x)u_0(y)\mathrm{d}x\mathrm{d}y \\
&\quad + \frac{1}{N}N^{\beta d}V(0) - \frac{2}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^d}(V^\eta * u_0)(x)u_0(x)\mathrm{d}x + \int_{\mathbb{R}^d}(V^\eta * u_0)(x)u_0(x)\mathrm{d}x \\
&= \frac{N(N-1)}{N^2}\int_{\mathbb{R}^d}(V^\eta * u_0)(x)u_0(x)\mathrm{d}x + N^{\beta d-1}V(0) - \int_{\mathbb{R}^d}(V^\eta * u_0)(x)u_0(x)\mathrm{d}x \\
&\le CN^{\beta d-1} + N^{-1}\|V^\eta * u_0\|_{L^\infty}\|u_0\|_{L^1} \\
&\le CN^{\beta d-1} + N^{-1}\|V^\eta\|_{L^1}\|u_0\|_{L^\infty}\|u_0\|_{L^1} \le CN^{-1/2-\varepsilon_0},
\end{aligned}$$

where we used Young's convolution inequality, $\|V^\eta\|_{L^1} = 1$, $\|u_0\|_{L^\infty} \le C$, and we have set $\varepsilon_0 = 1/2 - \beta d > 0$. This finishes the proof of Theorem 4.1. □

4.A Comments on Assumption (C1)

In the appendix of this chapter we discuss Assumption (C1) and show a proof of convergence in probability for interaction kernels approximating Coulomb interactions, which is partly done in a joint work with Li Chen, Veniamin Gvozdik and Yue Li, [28]. This shows that Assumption (C1) can be met by approximations of singular potentials. In Section 4.A.2 we point out what technical difficulties which occur if one wants to adapt the techniques used for Coulomb interactions in order to give a rigorous proof of Assumption (C1).

4.A.1 Convergence in probability for Coulomb interactions

In order to discuss Assumption (C1) (see (4.16)), we show a convergence result in probability, see Theorem 4.12 below, for a diffusion-aggregation model with Coulomb-type kernels:

For $d \ge 3$, we consider the following diffusion system on $\mathbb{R}^d$ with Coulomb-type aggregation ($\kappa = 1$) or Coulomb-type repulsion ($\kappa = -1$)

$$\partial_t \rho = \sigma\Delta\rho - \kappa\,\mathrm{div}\,\big(\rho\nabla\Phi * \rho\big), \tag{4.99}$$

where $\Phi = \frac{C_d}{|x|^{d-2}}$, denotes the fundamental solution of the Laplace equation in dimension $d \ge 3$ with a constant $C_d > 0$. Additionally, we assume $\rho(0) = \rho_0$ for a probability density function $\rho_0 \in L^\infty(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. Since the aim of this appendix is to illustrate ideas in order to show a mean-field convergence result in probability, we additionally assume that $\rho_0 \in C_c^\infty(\mathbb{R}^d)$, which can be reduced by using suitable approximating sequences, [29].

For approximating the Keller-Segel-type model (4.99) by a system of interacting particles, we first introduce an approximating sequence

$$V^\eta_{coul} := \chi^\eta * \Phi^\eta, \tag{4.100}$$

where $\chi^\eta = \eta^{-d}\chi(|x|/\eta)$ for a normalized $\chi \in C^2_c(\mathbb{R}^d)$ which fulfils $\chi = \xi * \xi$ for some $\xi \in C^2_c(\mathbb{R}^d)$ and Φ^η is a sequence of approximating kernels such that $\Phi^\eta \to \Phi$ point-wise for $\eta \to 0$; we comment on the choice of approximating kernels after introducing the particle systems:
The mean-field particle system for N interacting particles reads as follows

$$\begin{aligned} \mathrm{d}Y_i^{N,\eta}(t) &= \frac{\kappa}{N}\sum_{j=1}^N \nabla V^\eta_{coul}(Y_i^{N,\eta}(t) - Y_j^{N,\eta}(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ Y_i^{N,\eta}(0) &= \zeta_i \quad \text{in } \mathbb{R}^d, \quad i = 1,\dots,N, \end{aligned} \tag{4.101}$$

where ζ_i are i.i.d. random variables with common density function ρ_0 and $(W_i)_{i=1}^N$ denotes a family of independent d-dimensional Brownian motions.
Using standard ideas for moderately interacting particles, for fixed $\eta > 0$, we introduce an intermediate system of size N starting with the same initial condition as the mean-field particle system:

$$\begin{aligned} \mathrm{d}\overline{Y}_i^\eta(t) &= \kappa(\nabla V^\eta_{coul} * \bar\rho^\eta)(\overline{Y}_i^\eta(t))\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_i(t), \\ \overline{Y}_i^\eta(0) &= \zeta_i \quad \text{in } \mathbb{R}^d, \quad i = 1,\dots,N, \end{aligned} \tag{4.102}$$

where the particles $\overline{Y}_i^\eta$ are already independent with common density function $\bar\rho^\eta$, which is the solution to the smoothed version of the Keller-Segel model (4.99)

$$\begin{aligned} \partial_t \bar\rho^\eta &= \sigma\Delta\bar\rho^\eta - \kappa\,\mathrm{div}(\bar\rho^\eta \nabla V^\eta_{coul} * \bar\rho^\eta) \\ \bar\rho^\eta(0) &= \rho_0. \end{aligned} \tag{4.103}$$

This can be seen by using Itô's formula, which has been pointed out several times in this thesis. For a complete existence theory of (4.99) and (4.103) we refer to the work [29].

In this section of the thesis – for proving a mean-field result with respect to convergence in probability – we need the following regularity of the solution of the intermediate system:

Lemma 4.11. *For any $T > 0$, there exists a unique solution $\bar\rho^\eta \in L^\infty(0,T;H^s(\mathbb{R}^d))$ with $s > d/2 + 2$ to (4.103) such that $\|\bar\rho^\eta(t)\|_{L^1} = 1$ for all $0 < t < T$.*

For a proof of Lemma 4.11 we refer to [29, Theorem 2] where even a more general setting is considered.
The approximating sequence Φ^η of the Coulomb-type kernel Φ is chosen according to the work of Lazarovici and Pickl [72] such that the *Lipschitz*[6] constant of the mean-field force only diverges logarithmically in η^{-1}, i.e.

$$\||D^2 V^\eta_{coul}| * \bar\rho^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C\log(\eta^{-1})(\|\bar\rho^\eta\|_\infty + \|\bar\rho^\eta\|_{L^1}). \tag{4.104}$$

[6] With the absolute value inside it is not exactly the Lipschitz constant, but of same order

This can be assured by taking the sequence Φ^η according to [72] with additional *cut-off* in a ball around the origin, see [72, Lemma 6.1].

In the following, we prove a mean-field result in probability for particle system (4.101) to the intermediate system (4.102) for the choice $\eta = N^{-\beta}$. This corresponds to Assumption (C1) and (4.16) but with interaction potentials approximating the singular Coulomb kernel instead of a Dirac distribution. It follows mainly techniques developed in [71] and [72].

Theorem 4.12 (Convergence in probability for Coulomb potential)**.** *Let $\eta = N^{-\beta}$. We assume that $0 < \beta < 1/4d$ and $\beta(d+1) < \alpha < 1/2 - \beta(d-1)$. Let $(Y_i^{N,\eta})_{i=1}^N$ and $(\overline{Y}_i^\eta)_{i=1}^N$ be the solutions to systems (4.101) and (4.102), respectively. Then, for any $\gamma > 0$ and $T > 0$, there exists $C(\gamma, T) > 0$ such that for all $0 < t < T$,*

$$\mathbb{P}\Big(\max_{i=1,\dots,N} |Y_i^{N,\eta}(t) - \overline{Y}_i^\eta(t)| > N^{-\alpha}\Big) \le C(\gamma, T) N^{-\gamma}. \tag{4.105}$$

In order to prove Theorem 4.12, we need an equivalent result to Lemma 4.2 in the case that the interaction potentials approximate the Coulomb potential:

Lemma 4.13 (Law of large numbers)**.** *Let $(\overline{Y}_i^\eta)_{i=1}^N$ be the solution to system (4.102) and let $\bar\rho^\eta$ be the density function associated to $\overline{Y}_i^\eta$. Given $\theta \ge 0$ and $\phi_\eta \in L^\infty(\mathbb{R}^d)$, $\psi_\eta \in L^\infty(\mathbb{R}^d; \mathbb{R}^n)$ with $n \in \{1, d, d\times d\}$, we define the sets*

$$\mathcal{A}^N_{\theta,\phi_\eta}(t) := \left\{\omega \in \Omega : \left|\frac{1}{N}\sum_{i=1}^N \phi_\eta(\overline{Y}_i^\eta(t)) - \int_{\mathbb{R}^d} \phi_\eta(x)\bar\rho^\eta(t,x)\mathrm{d}x\right| > N^{-\theta}\right\}, \tag{4.106}$$

$$\mathcal{B}^N_{\theta,\psi_\eta}(t) := \bigcup_{i=1}^N \left\{\omega \in \Omega : \left|\frac{1}{N}\sum_{j=1}^N \psi_\eta\big(\overline{Y}_i^\eta(t) - \overline{Y}_j^\eta(t)\big) - (\psi_\eta * \bar\rho^\eta)(\overline{Y}_i^\eta(t))\right| > N^{-\theta}\right\}. \tag{4.107}$$

Then, for every $m \in \mathbb{N}$ and $T > 0$, there exists $C(m) > 0$ such that for all $0 < t < T$,

$$\mathbb{P}(\mathcal{A}_{\theta,\phi_\eta}(t)) \le C(m)\|\phi_\eta\|_{L^\infty}^{2m} N^{2m(\theta-1/2)},$$
$$\mathbb{P}(\mathcal{B}_{\theta,\psi_\eta}(t)) \le C(m)\|\psi_\eta\|_{L^\infty}^{2m} N^{2m(\theta-1/2)+1}.$$

Since all $\overline{Y}_i^\eta$ are already independent, the proof can be done in an analogous way as the proof of Lemma 4.2.

Idea of the proof of Theorem 4.12: For the proof of Theorem 4.12, we use a combination of Markov inequality and a stopping time argument. First, we estimate the probability using Markov's inequality according to

$$\mathbb{P}\Big(\max_{i=1,\dots,N} |Y_i^{N,\eta}(t) - \overline{Y}_i^\eta(t)| > N^{-\alpha}\Big) \le \mathbb{E}(S_\alpha^k(t)),$$

where $S_\alpha^k(t) = (N^\alpha \max_{i=1,\dots,N}(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t\wedge\tau_\alpha)|)^k$, τ_α is a suitable stopping time such that $S_\alpha^k(t)$ is bounded, and $k \in \mathbb{N}$ is an arbitrary number. To bound the expectation of $S_\alpha^k(t)$ by $N^{-\gamma}$ (up to a constant), we make use Lemma of 4.13 (law-of-large numbers), Taylor's expansion and a Gronwall argument. As mentioned in the introduction of this appendix, the main ideas follow techniques developed in [71] and [72]. Furthermore, for the proof of Theorem 4.12 we need the following auxiliary lemma:

Lemma 4.14 (Estimates for V^η_{coul})**.** *For V^η_{coul} defined in (4.100) with $\eta = N^{-\beta}$, there exists a constant $C > 0$ such that for $k = 1, 2, 3$ and $f \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$*

$$\|\mathrm{D}^k V^\eta_{coul}\|_{L^\infty} \le C N^{\beta(d-2+k)}, \qquad \|V^\eta_{coul} * f\| \le C. \tag{4.108}$$

Additionally, if $\bar\rho^\eta$ denotes the weak solution to (4.103), then the following estimate holds

$$\|\mathrm{D}^2 V^\eta_{coul} * \bar\rho^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le C, \tag{4.109}$$

for a constant $C > 0$ which does not depend on N.

Proof. For fixed $x \in \mathbb{R}^d$, we get by recalling that $\eta = N^{-\beta}$ and by the fact that the cut-off fulfils $\Phi^\eta \le \Phi$

$$\begin{aligned}|\mathrm{D}^k V^\eta_{coul}(x)| &\le \int_{\mathbb{R}^d} |\mathrm{D}^k \chi^\eta(x-y)|\Phi_{|y|\le\eta}(y)\mathrm{d}y + \int_{\mathbb{R}^d} |\mathrm{D}^k \chi^\eta(x-y)|\Phi_{|y|\ge\eta}(y)\mathrm{d}y \\ &\le \|\mathrm{D}^k\chi^\eta\|_{L^\infty}\|\Phi_{|y|\le\eta}\|_{L^1} + \|\mathrm{D}^k\chi^\eta\|_{L^1}\|\Phi_{|y|\ge\eta}\|_{L^\infty} \le CN^{\beta(d-2+k)},\end{aligned}$$

since a simple computation shows

$$\|\mathrm{D}^k\chi^\eta\|_{L^\infty} \le CN^{\beta(d+k)}, \quad \|\Phi_{|y|\le\eta}\|_{L^1} \le C\eta^2 = CN^{-2\beta},$$

and

$$\|\mathrm{D}^k\chi^\eta\|_{L^1} = N^{k\beta}, \quad \|\Phi_{|y|\ge\eta}\|_{L^\infty} \le C\eta^{-(d-2)} = N^{\beta(d-2)}.$$

This shows the first claim in (4.108), where we remark that we do not need the cut-off in the definition of Φ^η for this part of the proof.
In order to show the second claim in (4.108), we see that for fixed $x \in \mathbb{R}^d$, we have

$$\begin{aligned}V^\eta_{coul} * f(x) &= \int_{\mathbb{R}^d} \chi^\eta * \Phi^\eta(y) f(x-y)\mathrm{d}y \\ &\le \int_{|y|\le 1} \chi^\eta * \Phi^\eta(y) f(x-y)\mathrm{d}y + \int_{|y|\ge 1} \chi^\eta * \Phi^\eta(y) f(x-y)\mathrm{d}y \\ &\le C\|\chi^\eta\|_{L^1}\|\Phi\|_{L^1(B_1)}\|f\|_{L^\infty} + C\|\chi^\eta\|_{L^1}\|\Phi\|_{L^\infty(\mathbb{R}^d\setminus B_1)}\|f\|_{L^1} \le C.\end{aligned}$$

For the proof of (4.109) we need to be a bit more careful, since $D^2\Phi$ is not integrable in a ball around zero.
First, for $\|\mathrm{D}^2 V^\eta_{coul} * \bar\rho^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}$ we can put one derivative on the solution $\bar\rho^\eta$ and arrive at

$$\begin{aligned}\|\mathrm{D}^2 V^\eta_{coul} * \bar\rho^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} &= \|\nabla V^\eta_{coul} * \nabla\bar\rho^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \\ &\le C\|\nabla V^\eta_{coul} * \nabla\bar\rho^\eta\|_{L^\infty(0,T;W^{s-1,p}(\mathbb{R}^d))},\end{aligned}$$

for the choice $p = 2d/(d+2)$, where we used the Sobolev embedding for $p(s-1) > d$, which is possible if we take $s > d/2+2$. Young's convolutional inequality with $q = d/(d+1)$ then implies

$$\begin{aligned}\|\nabla V^\eta_{coul} * \nabla\bar\rho^\eta\|_{L^\infty(0,T;W^{s-1,p}(\mathbb{R}^d))} &\le C\|\chi^\eta\|_{L^1(\mathbb{R}^d)}\|\nabla\bar\rho^\eta\|_{L^\infty(0,T;H^{s-1}(\mathbb{R}^d))} \\ &\le C\|\bar\rho^\eta\|_{L^\infty(0,T;H^s(\mathbb{R}^d))} \le C,\end{aligned}$$

where we have used the Hardy-Littlewood-Sobolev inequality, $\|\chi^\eta\|_{L^1} = 1$ and the fact that $\|\bar\rho^\eta\|_{L^\infty(0,T;H^s(\mathbb{R}^d))}$ is uniformly bounded, see Lemma 4.11. This finishes the proof. □

Proof of Theorem 4.12

Proof. **1. Preparations.** We start with some definitions. Let $\alpha > 0$ be given as in the theorem and let $k \in \mathbb{N}$. We define the stopping time

$$\tau_\alpha(\omega) := \inf\left\{t \in (0,T) : \max_{i=1,\dots,N} |(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t)| \geq N^{-\alpha}\right\}$$

and the random variable

$$S_\alpha^k(t) := \left(N^\alpha \max_{i=1,\dots,N} |(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t\wedge\tau_\alpha)|\right)^k \leq 1.$$

Additionally, we define the set

$$B_\alpha(t) := \{\omega \in \Omega : S_\alpha^k(t) = 1\},$$

which includes for fixed $t > 0$ all $\omega \in \Omega$ such that the first time $s > 0$ of $\max_{i=1,\dots,N} |(Y_i^{N,\eta} - \overline{Y}_i^\eta)(s,\omega)| \geq N^{-\alpha}$ fulfils that $s \leq t$, i.e. $\tau_\alpha(\omega) \leq t$.
Note that this set does not depend on k, since $S_\alpha^k(t) = 1$ is equivalent to $S_\alpha^k(t)^{1/k} = 1$, and $S_\alpha^k(t)^{1/k}$ does not depend on k.
It follows from the continuity of the paths of $Y_i^{N,\eta}$ and $\overline{Y}_i^\eta$ and the fact that if $\max_{i=1,\dots,N} |(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t,\omega)| > N^{-\alpha}$ for a fixed $t > 0$ then $t > \tau_\alpha(\omega)$ that

$$\begin{aligned}\mathbb{P}\Big(\max_{i=1,\dots,N} |(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t)| > N^{-\alpha}\Big) &\leq \mathbb{P}\Big(\max_{i=1,\dots,N} |(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t\wedge\tau_\alpha)| = N^{-\alpha}\Big)\\ &= \mathbb{P}(B_\alpha(t)) = \mathbb{P}(S_\alpha^k(t) = 1) \leq \mathbb{E}(S_\alpha^k(t)),\end{aligned}$$

where the last estimate follows from Markov's inequality.
Now, if we show that for every $\gamma > 0$ and $T > 0$, there exists $k \in \mathbb{N}$ and $C = C(\gamma, k, T) > 0$ such that

$$\mathbb{E}(S_\alpha^k(t)) \leq CN^{-\gamma},$$

the proof is finished.
To prove this claim, we insert the integral formulations of (4.1) and (4.6) and add

$$\pm\nabla V_{coul}^\eta(\overline{Y}_i^\eta(s) - \overline{Y}_j^\eta(s))$$

in the last step: For every $i = 1,\dots,N$ it holds that

$$\begin{aligned}&\left|(Y_i^{N,\eta} - \overline{Y}_i^\eta)(t\wedge\tau_\alpha)\right|^k \qquad (4.110)\\ &\leq C(k,T)\int_0^{t\wedge\tau_\alpha}\left|\frac{1}{N}\sum_{j=1}^N \nabla V_{coul}^\eta(Y_i^{N,\eta}(s) - Y_j^{N,\eta}(s)) - (\nabla V_{coul}^\eta * \bar\rho^\eta)(s,\overline{Y}_i^\eta(s))\right|^k \mathrm{d}s\\ &\leq C(k,T)(I_{1,i}(t) + I_{2,i}(t)),\end{aligned}$$

where

$$I_{1,i}(t) = \int_0^{t\wedge\tau_\alpha}\left|\frac{1}{N}\sum_{j=1}^N\left[\nabla V_{coul}^\eta(Y_i^{N,\eta}(s) - Y_j^{N,\eta}(s)) - \nabla V_{coul}^\eta(\overline{Y}_i^\eta(s) - \overline{Y}_j^\eta(s))\right]\right|^k \mathrm{d}s,$$

$$I_{2,i}(t) = \int_0^{t\wedge\tau_\alpha} \left| \frac{1}{N}\sum_{j=1}^N \nabla V^\eta_{coul}(\overline{Y}^\eta_i(s) - \overline{Y}^\eta_j(s)) - (\nabla V^\eta_{coul} * \bar{\rho}^\eta)(s, \overline{Y}^\eta_i(s)) \right|^k \mathrm{d}s.$$

In the following, we estimate both terms.

2. Estimate for $I_{2,i}(t)$. The term $I_{2,i}(t)$ can be estimated by a law-of-large numbers argument. We wish to apply Lemma 4.13 with $\psi_\eta = \nabla V^\eta_{coul}$ and $\theta > 0$ which will be chosen later in the proof. In order to shorten notation, we abbreviate the integrand of $I_{2,i}(t)$ as

$$\widetilde{I}_i(s) = \left| \frac{1}{N}\sum_{j=1}^N \nabla V^\eta_{coul}(\overline{Y}^\eta_i(s) - \overline{Y}^\eta_j(s)) - (\nabla V^\eta_{coul} * \bar{\rho}^\eta)(s, \overline{Y}^\eta_i(s)) \right|.$$

We have, with the notation of Lemma 4.13, $\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s) = \bigcup_{i=1}^N \{\widetilde{I}_i(s) > N^{-\theta}\}$. Keeping in mind that we want to estimate $\mathbb{E}(S^k_\alpha(t))$, we compute the expectation of

$$N^{\alpha k} \max_{i=1,\dots,N} I_{2,i}(t)$$

by splitting Ω into the two sets $\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s)$ and its complement $\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s)^c$. First, we observe that $\widetilde{I}_i(s) \le N^{-\theta}$ for all $i = 1,\dots,N$ on $\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s)^c$. This yields

$$\begin{aligned}
\mathbb{E}\Big(N^{\alpha k} \max_{i=1,\dots,N} I_{2,i}(t)\Big) &\le \mathbb{E}\Big(N^{\alpha k} \int_0^t \max_{i=1,\dots,N} \widetilde{I}_i(s)^k \, \mathbb{1}_{\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s)^c} \mathrm{d}s\Big) \\
&\quad + \mathbb{E}\Big(N^{\alpha k} \int_0^t \max_{i=1,\dots,N} \widetilde{I}_i(s)^k \, \mathbb{1}_{\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s)} \mathrm{d}s\Big) \\
&\le T N^{\alpha k} N^{-\theta k} + C(T) N^{\alpha k} \|\nabla V^\eta_{coul}\|^k_{L^\infty} \sup_{0<s<T} \mathbb{P}(\mathcal{B}^N_{\theta,\nabla V^\eta_{coul}}(s)).
\end{aligned}$$

Then, because of $\|\nabla V^\eta_{coul}\|_{L^\infty} \le C N^{\beta(d-1)}$ (see (4.108)) and after an application of Lemma 4.13, for any $m \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}\Big(N^{\alpha k} \max_{i=1,\dots,N} I_{2,i}(t)\Big) &\le C(k,m,T) N^{\alpha k}\big(N^{-\theta k} + N^{\beta(d-1)k} N^{2m\beta(d-1)} N^{2m(\theta-1/2)+1}\big) \\
&= C(k,m,T)\big(N^{(\alpha-\theta)k} + N^{\alpha k + \beta(d-1)(k+2m) + m(2\theta-1)+1}\big).
\end{aligned} \tag{4.111}$$

This finishes the estimate for $I_{2,i}$.

3. Estimate for $I_{1,i}(t)$. Similar as for $I_{2,i}(t)$, in order to shorten notation, we define

$$\widehat{I}_i(s) = \left| \frac{1}{N}\sum_{j=1}^N \nabla V^\eta_{coul}(Y^{N,\eta}_i(s) - Y^{N,\eta}_j(s)) - \nabla V^\eta_{coul}(\overline{Y}^\eta_i(s) - \overline{Y}^\eta_j(s)) \right|.$$

The estimate for $I_{1,i}(t)$ is more technical than the law-of-large numbers estimate for $I_{2,i}(t)$. We perform a Taylor expansion of ∇V^η_{coul} around $(\overline{Y}^\eta_i - \overline{Y}^\eta_j)(s)$ with a linear term and a quadratic remainder:

$$\mathbb{E}\Big(N^{\alpha k} \max_{i=1,\dots,N} I_{1,i}(t)\Big) \le \mathbb{E}\Big(N^{\alpha k} \int_0^{t\wedge\tau_\alpha} \max_{i=1,\dots,N} \widehat{I}_i(s)^k \mathrm{d}s\Big) \tag{4.112}$$

$$\begin{aligned}
&\leq C(k)\mathbb{E}\bigg(N^{\alpha k}\int_0^{t\wedge\tau_\alpha}\max_{i=1,\dots,N}\bigg|\frac{1}{N}\sum_{j=1}^N \mathrm{D}^2V_{coul}^\eta(\overline{Y}_i^\eta(s)-\overline{Y}_j^\eta(s)) \\
&\quad\times\big((Y_i^{N,\eta}-Y_j^{N,\eta})-(\overline{Y}_i^\eta-\overline{Y}_j^\eta)\big)(s)\bigg|^k\mathrm{d}s\bigg) \\
&\quad + C(k)\|\mathrm{D}^3V_{coul}^\eta\|_{L^\infty}^k\mathbb{E}\bigg(N^{\alpha k}\int_0^{t\wedge\tau_\alpha}\max_{i=1,\dots,N}\frac{1}{N}\sum_{j=1}^N\big|\big((Y_i^{N,\eta}-\overline{Y}_i^\eta)(s\wedge\tau_\alpha) \\
&\quad -(Y_j^{N,\eta}-\overline{Y}_j^\eta)(s\wedge\tau_\alpha)\big)\big|^{2k}\mathrm{d}s\bigg)\leq C(k)(I_{11}+I_{12}+I_{13})(t),
\end{aligned}$$

where

$$\begin{aligned}
I_{11}(t)&=\mathbb{E}\bigg(N^{\alpha k}\int_0^{t\wedge\tau_\alpha}\max_{i=1,\dots,N}\bigg|\frac{1}{N}\sum_{j=1}^N\mathrm{D}^2V_{coul}^\eta(\overline{Y}_i^\eta-\overline{Y}_j^\eta)(Y_i^{N,\eta}-\overline{Y}_i^\eta)(s)\bigg|^k\mathrm{d}s\bigg),\\
I_{12}(t)&=\mathbb{E}\bigg(N^{\alpha k}\int_0^{t\wedge\tau_\alpha}\max_{i=1,\dots,N}\bigg|\frac{1}{N}\sum_{j=1}^N\mathrm{D}^2V_{coul}^\eta(\overline{Y}_i^\eta-\overline{Y}_j^\eta)(Y_j^{N,\eta}-\overline{Y}_j^\eta)(s)\bigg|^k\mathrm{d}s\bigg),\\
I_{13}(t)&=\|\mathrm{D}^3V_{coul}^\eta\|_{L^\infty}^k\mathbb{E}\bigg(N^{\alpha k}\int_0^t\max_{i=1,\dots,N}\big|(Y_i^{N,\eta}-\overline{Y}_i^\eta)(s\wedge\tau_\alpha)\big|^{2k}\mathrm{d}s\bigg).
\end{aligned}$$

We start with $I_{13}(t)$. It follows from Fubini's theorem, $\|\mathrm{D}^3V_{coul}^\eta\|_{L^\infty}\leq CN^{\beta(d+1)}$ (see (4.108)), the definition of $S_\alpha^k(s)$ and $S_\alpha^k(s)^2\leq S_\alpha^k(s)$ (since $S_\alpha^k(s)\leq 1$) that

$$I_{13}(t)\leq C(k)N^{\beta(d+1)k}\int_0^t\mathbb{E}\big(N^{-\alpha k}S_\alpha^k(s)^2\big)\mathrm{d}s\leq C(k)N^{\beta(d+1)k-\alpha k}\int_0^t\mathbb{E}(S_\alpha^k(s))\mathrm{d}s. \quad (4.113)$$

Note that we need the definition of the stopping time τ_α, which guarantees that $S_\alpha^k(t)\leq 1$. Next, we estimate $I_{11}(t)\leq I_{111}(t)+I_{112}(t)$ by adding and subtracting $(\mathrm{D}^2V_{coul}^\eta*\bar{\rho}^\eta)(\overline{Y}_i^\eta)$:

$$\begin{aligned}
I_{111}(t)&=\mathbb{E}\bigg(\int_0^{t\wedge\tau_\alpha}S_\alpha^k(s)\max_{i=1,\dots,N}\bigg|\frac{1}{N}\sum_{j=1}^N\mathrm{D}^2V_{coul}^\eta(\overline{Y}_i^\eta(s)-\overline{Y}_j^\eta(s))\\
&\qquad-(\mathrm{D}^2V_{coul}^\eta*\bar{\rho}^\eta)(s,\overline{Y}_i^\eta(s))\bigg|^k\mathrm{d}s\bigg),\\
I_{112}(t)&=\mathbb{E}\bigg(\int_0^{t\wedge\tau_\alpha}S_\alpha^k(s)\max_{i=1,\dots,N}\big|(\mathrm{D}^2V_{coul}^\eta*\bar{\rho}^\eta)(s,\overline{Y}_i^\eta(s))\big|^k\mathrm{d}s\bigg).
\end{aligned}$$

For $I_{111}(t)$, we apply Lemma 4.13 for $m\in\mathbb{N}$ (which will be chosen later in the proof) with $\psi_\eta=\mathrm{D}^2V_{coul}^\eta$ and $\theta=0$ and split Ω into $B_{0,\mathrm{D}^2V_{coul}^\eta}(s)$ and $B_{0,\mathrm{D}^2V_{coul}^\eta}(s)^c$. Fubini's theorem then leads to

$$\begin{aligned}
I_{111}(t)&\leq\mathbb{E}\bigg(\int_0^{t\wedge\tau_\alpha}S_\alpha^k(s)\max_{i=1,\dots,N}\bigg|\frac{1}{N}\sum_{j=1}^N\mathrm{D}^2V_{coul}^\eta(\overline{Y}_i^\eta(s)-\overline{Y}_j^\eta(s))\\
&\qquad-(\mathrm{D}^2V_{coul}^\eta*\bar{\rho}^\eta)(s,\overline{Y}_i^\eta(s))\bigg|^k\mathbb{1}_{B_{0,\mathrm{D}^2V_{coul}^\eta}(s)^c}\mathrm{d}s\bigg)
\end{aligned}$$

$$+ C(T)\|\mathrm{D}^2 V^\eta_{coul}\|^k_{L^\infty} \sup_{0<s<T} \mathbb{P}(B_{0,\mathrm{D}^2 V^\eta_{coul}}(s))$$
$$\leq \int_0^t \mathbb{E}(S^k_\alpha(s))\mathrm{d}s + C(m,T)N^{\beta dk} N^{2m\beta d} N^{2m(0-1/2)+1},$$

using again $\|\mathrm{D}^2 V^\eta_{coul}\|_{L^\infty} \leq CN^{\beta d}$ (see (4.108)) and the construction of $B_{0,\mathrm{D}^2 V^\eta_{coul}}(s)$. The estimate for $I_{112}(t)$ simply follows from Fubini's theorem and (4.108):

$$I_{112}(t) \leq \|\mathrm{D}^2 V^\eta_{coul} * \bar{\rho}^\eta\|^k_{L^\infty} \int_0^t \mathbb{E}(S^k_\alpha(s))\mathrm{d}s \leq C(k) \int_0^t \mathbb{E}(S^k_\alpha(s))\mathrm{d}s,$$

where we used estimate (4.109). We conclude that

$$I_{11}(t) \leq C(k) \int_0^t \mathbb{E}(S^k_\alpha(s))\mathrm{d}s + C(k,m,T)N^{\beta d(k+2m)-m+1}, \tag{4.114}$$

where the constant $C(k) > 0$ depends on the $L^\infty(0,T;W^{2,\infty}(\mathbb{R}^d))$ norm of $\bar{\rho}^\eta$, which is bounded uniformly in η.
Finally, we estimate $I_{12}(t)$ by similar techniques as for $I_{11}(t)$, however, since we sum over $j = 1,\dots,N$, we have to put the modulus inside the sum in order to put $S^k_\alpha(s)$ out of the expression. First, we put the modulus inside the sum:

$$I_{12}(t) \leq \mathbb{E}\bigg(N^{\alpha k} \int_0^{t\wedge\tau_\alpha} \max_{i=1,\dots,N} \bigg(\frac{1}{N} \sum_{j=1}^N \big|\mathrm{D}^2 V^\eta_{coul}(\overline{Y}^\eta_i(s) - \overline{Y}^\eta_j(s))\big| \big|Y^{N,\eta}_j(s) - \overline{Y}^\eta_j(s)\big| \bigg)^k \mathrm{d}s \bigg)$$
$$\leq \mathbb{E}\bigg(\int_0^{t\wedge\tau_\alpha} S^k_\alpha(s) \max_{i=1,\dots,N} \bigg(\frac{1}{N} \sum_{j=1}^N \big|\mathrm{D}^2 V^\eta_{coul}(\overline{Y}^\eta_i(s) - \overline{Y}^\eta_j(s))\big| \bigg)^k \mathrm{d}s \bigg),$$

where we used the definition of $S^k_\alpha(s)$ for $s < \tau_\alpha$. Similarly as in the estimate for $I_{11}(t)$, we add and subtract $|\mathrm{D}^2 V^\eta_{coul}| * \bar{\rho}^\eta(s, \overline{Y}^\eta_i(s))$, which yields $I_{12}(t) \leq C(k)(I_{121}(t) + I_{122}(t))$, where

$$I_{121}(t) = \mathbb{E}\bigg(\int_0^{t\wedge\tau_\alpha} S^k_\alpha(s) \max_{i=1,\dots,N} \bigg| \frac{1}{N} \sum_{j=1}^N \big|\mathrm{D}^2 V^\eta_{coul}(\overline{Y}^\eta_i(s) - \overline{Y}^\eta_j(s))\big|$$
$$- \big(|\mathrm{D}^2 V^\eta_{coul}| * \bar{\rho}^\eta\big)(s, \overline{Y}^\eta_i(s)) \bigg|^k \mathrm{d}s \bigg),$$
$$I_{122}(t) = \mathbb{E}\bigg(\int_0^{t\wedge\tau_\alpha} S^k_\alpha(s) \max_{i=1,\dots,N} \big((|\mathrm{D}^2 V^\eta_{coul}| * \bar{\rho}^\eta)(s, \overline{Y}^\eta_i(s)) \big)^k \mathrm{d}s \bigg).$$

By Lemma 4.13 with $\psi_\eta = |\mathrm{D}^2 V^\eta_{coul}|$ and $\theta = 0$, using $S^k_\alpha(s) \leq 1$ yields for any $m \in \mathbb{N}$

$$I_{121}(t) \leq \int_0^t \mathbb{E}(S^k_\alpha(s))\mathrm{d}s + C(T)\|\mathrm{D}^2 V^\eta_{coul}\|^k_{L^\infty} \sup_{0<s<T} \mathbb{P}(B^N_{0,|\mathrm{D}^2 V^\eta_{coul}|}(s))$$
$$\leq \int_0^t \mathbb{E}(S^k_\alpha(s))\mathrm{d}s + C(m,T)N^{\beta dk} N^{2m\beta d} N^{2m(0-1/2)+1}, \tag{4.115}$$

where the estimates follow the ones for $I_{111}(t)$.
We obtain for $I_{122}(t)$:

$$\begin{aligned} I_{122}(t) &\leq C \big\| |\mathrm{D}^2 V_{coul}^{\eta}| * \bar{\rho}^{\eta} \|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}^k \int_0^t \mathbb{E}(S_\alpha^k(s))\mathrm{d}s \\ &\leq C(k)\log(N) \int_0^t \mathbb{E}(S_\alpha^k(s))\mathrm{d}s, \end{aligned} \tag{4.116}$$

where we used that our approximating sequence fulfils (4.104).
Together with (4.115), we infer that

$$I_{12}(t) \leq C(k)(1+\log(N)) \int_0^t \mathbb{E}(S_\alpha^k(s))\mathrm{d}s + C(k,m,T)N^{\beta d(k+2m)-m+1}. \tag{4.117}$$

We insert the estimates for $I_{11}(t)$ in (4.114), $I_{12}(t)$ in (4.117), and $I_{13}(t)$ in (4.113) into (4.112):

$$\begin{aligned} \mathbb{E}\Big(N^{\alpha k} \max_{i=1,\dots,N} I_{1,i}(t)\Big) \leq C(k)\big(1+\log(N) + N^{\beta(d+1)k-\alpha k}\big) \int_0^t \mathbb{E}(S_\alpha^k(s))\mathrm{d}s \\ + C(k,m,T)N^{\beta d(k+2m)-m+1}. \end{aligned}$$

Combining this estimate with (4.111), we conclude from (4.110) that

$$\begin{aligned} \mathbb{E}(S_\alpha^k(t)) &= \mathbb{E}\Big(N^{\alpha k} \max_{i=1,\dots,N} \big|Y_i^{N,\eta}(t\wedge\tau_\alpha) - \overline{Y}_i^{\eta}(t\wedge\tau_\alpha)\big|^k\Big) \\ &\leq C(k)\big(1+N^{\beta(d+1)k-\alpha k} + \log(N)\big) \int_0^t \mathbb{E}(S_\alpha^k(s))\mathrm{d}s \\ &\quad + C(k,m,T)\big(N^{k(\alpha-\theta)} + N^{\alpha k+\beta(d-1)(k+2m)+m(2\theta-1)+1} + N^{\beta d(k+2m)-m+1}\big). \end{aligned}$$

Since $\alpha \geq \beta(d+1)$ by assumption, the factor $N^{\beta(d+3)k-\alpha k}$ is bounded for all N. We claim that for any given $\gamma > 0$ and (β,α) chosen according to the theorem, we can choose k, θ and m such that the remaining terms are bounded by $N^{-\gamma}$.
Indeed, let $\theta \in (\alpha, 1/2-\beta(d-1))$. Then we choose $k \in \mathbb{N}$ so large that $k(\alpha-\theta) \leq -\gamma$. Furthermore, we choose $m \in \mathbb{N}$ sufficiently large such that

- ▷ $\beta d(k+2m) - m + 1 \leq -\gamma$ (which is possible because of $\beta < 1/4d$) and
- ▷ $\alpha k + \beta(d-1)(k+2m) + m(2\theta-1) + 1 \leq -\gamma$ (which is possible since $\theta < 1/2 - \beta(d-1)$).

We infer that

$$\mathbb{E}(S_\alpha^k(t)) \leq C(k)(1+\log(N)) \int_0^t \mathbb{E}(S_\alpha^k(s))\mathrm{d}s + C(k,m,T)N^{-\gamma},$$

and an application of Gronwall's lemma implies that $\mathbb{E}(S_\alpha^k(t)) \leq C(k,m,T)N^{-\gamma+1}$. Since $\gamma > 0$ was arbitrary, this concludes the proof. □

4.A.2 Comments on the proof in the moderate regime

An interesting and natural question is whether we can adapt the proof of Theorem 4.12 for moderately interacting particles such that we do not need to assume Assumption (C1); see (4.16). If we perform the proof of Theorem 4.12 with V^η defined in (4.2) for $\eta = N^{-\beta}$ instead of V^η_{coul}, two terms are of particular interest (with $\bar{u}^\eta$ we denote the weak solution to the intermediate PDE (4.5)):

(i) $\|\mathrm{D}^2 V^\eta * \bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}$ used for estimates of term I_{112}, and

(ii) $\||\mathrm{D}^2 V^\eta| * \bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))}$ used for estimates of term I_{122}.

For the first term, we can use the regularity of $\bar{u}^\eta$ in order to estimate

$$\|\mathrm{D}^2 V^\eta * \bar{u}^\eta\|_{L^\infty(0,T;L^\infty(\mathbb{R}^d))} \le \|V^\eta\|_{L^1(\mathbb{R}^d)}\|\mathrm{D}^2\bar{u}^\eta\|_{L^\infty} \le C.$$

The situation becomes more delicate for $|\mathrm{D}^2 V^\eta| * \bar{u}^\eta$: Due to the absolute value, we can not put the derivate on the solution of the intermediate partial differential equation $\bar{u}^\eta$. Choosing a purely convex (or concave) potential V is also not possible, since we need at least integrability on the whole space $\mathbb{R}^d$. Interestingly, in [93] Oelschläger showed a mean-field convergence result for the porous media equation without additional diffusion in one dimension by using a singular potential V such that $V''(x) \ge 0$ for $x \neq 0$, see [93, Forumla (2.8)]. It is still an open problem whether we can do a similar trick since the case $x = 0$, which corresponds to particles being exactly at the same place, has to be treated in a careful way. Another way to treat the difficulties could be to use a different strategy in the proof in order to avoid the absolute value inside the convolution. Future work will go in those two directions.

At the end of this appendix, the author wants to remark that for logarithmic scaling of $\eta > 0$ with respect to the number of particles, the proof of Assumption (C1) can be done exactly as for Theorem 4.12. The main reason lies in the fact that for $\eta^{-1} \sim \log(N)$ the norm of $|\mathrm{D}^2 V^\eta| * \bar{u}^\eta$ scales only logarithmically in N, which is the same situation as in the case of Coulomb interaction in Theorem 4.12. However – in terms of fluctuations around the mean-field limit — we are interested in an algebraic rate of $\eta > 0$, i.e. $\eta = N^{-\beta}$ for some $\beta > 0$.

Bibliography

[1] L. Ambrosio and S. Serfaty. A gradient flow approach to an evolution problem arising in superconductivity. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences,* 61 (2008), 1495–1539.

[2] J. Baladron, D. Fasoli, O. Faugeras, and J. Touboul. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *The Journal of Mathematical Neuroscience*, 2, 1-50, 2012.

[3] M. Bertsch, M. Gurtin, D. Hilhorst, and L. Peletier. On interacting populations that disperse to avoid crowding: preservation of segregation. *Journal of Mathematical Biology*, 23 (1985), 1–13.

[4] P. Biler, C. Imbert, and G. Karch. Barenblatt profiles for a non local porous medium equation. *Comptes Rendus Mathematique*, 349 (2011), 641–645.

[5] P. Biler, C. Imbert, and G. Karch. The nonlocal porous medium equation: Barenblatt profiles and other weak solutions. *Archive for Rational Mechanics and Analysis*, 215 (2015), 497–529.

[6] N. Boers, and P. Pickl. On mean field limits for dynamical systems. *Journal of Statistical Physics*, 164, 1-16, 2016.

[7] L. Boltzmann. Weitere Studien über das Wärmegleichgewicht unter Gas-Molekülen. *Wissenschaftliche Abhandlungen*, 316-402, 1872.

[8] M. Bonforte, Y. Sire, and J. L. Vázquez. Existence, uniqueness and asymptotic behaviour for fractional porous medium equations on bounded domains. *Discrete and Continuous Dynamical Systems* 35 (2015), 5725–5767.

[9] W. Braun, and K. Hepp. The Vlasov dynamics and its fluctuations in the 1/N limit of interacting classical particles. *Communications in mathematical physics*, 56(2), 101-113, 1977.

[10] D. Bresch, P.-E. Jabin, and Z. Wang. Mean-field limit and quantitative estimates with singular attractive kernels. *arXiv preprint arXiv:2011.08022*, 2020.

[11] H. Brézis. Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, New York, 2011.

[12] M. Burger, J. A. Carrillo, J.-F. Pietschmann, and M. Schmidtchen. Segregation effects and gap formation in cross-diffusion models. *Interfaces Free Bound,* 22 (2020), 175–203.

[13] M. Burger, J.-F. Pietschmann, H. Ranetbauer, C. Schmeiser, and M.-T. Wolfram. Mean-field models for segregation dynamics. To appear in *European Journal of Applied Mathematics*, 2020. arXiv:1808.04069.

[14] L. Caffarelli, M. Gualdani, and N. Zamponi. Existence of weak solutions to a continuity equation with space time nonlocal Darcy law. *Communications in Partial Differential Equations* 45 (2020), 1799–1819.

[15] L. Caffarelli, F. Soria, F., and J. L. Vázquez. Regularity of solutions of the fractional porous medium flow. *Journal of the European Mathematical Society* 15 (2013), 1701–1746.

[16] L. Caffarelli and J. L. Vázquez. Asymptotic behaviour of a porous medium equation with fractional diffusion. *Discrete and Continuous Dynamical Systems*, 29 (2011), 1393–1404.

[17] L. Caffarelli and J. L. Vázquez. Nonlinear porous medium flow with fractional potential pressure. *Archive for Rational Mechanics and Analysis*, 202 (2011), 537–565.

[18] J. A. Carrillo, Y. P. Choi, and M. Hauray. The derivation of swarming models: mean-field limit and Wasserstein distances. *Collective dynamics from bacteria to crowds*, 553, 1-46, 2014.

[19] J. A. Carrillo, Y. P. Choi, and S. Salem. Propagation of chaos for the Vlasov–Poisson–Fokker–Planck equation with a polynomial cut-off. *Communications in Contemporary Mathematics*, 21(04), 1850039, 2019.

[20] J. A. Carrillo, Y. Huang, M. C. Santos, and J. L. Vazquez. Exponential convergence towards stationary states for the 1D porous medium equation with fractional pressure. *Journal of Differential Equations*, 258 (2015), 736–763.

[21] L. P. Chaintron, and A. Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *arXiv preprint arXiv:2203.00446*, 2022.

[22] L. P. Chaintron, and A. Diez. Propagation of chaos: a review of models, methods and applications. II. Applications. *arXiv preprint arXiv:2106.14812*, 2021.

[23] N. Champagnat, R. Ferrière, and S. Méléard. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theoretical population biology*, 69(3), 297-321, 2006.

[24] J. Chapman, J. Rubinstein, and M. Schatzman. A mean-field model for superconducting vortices. *European Journal of Applied Mathematics*, 7 (1996), 97–111.

[25] L. Chen, E. S. Daus, A. Holzinger, and A. Jüngel. Rigorous derivation of population cross-diffusion systems from moderately interacting particle systems. *Journal of Nonlinear Science*, 31, 1-38, 2021.

[26] L. Chen, E. S. Daus, and A. Jüngel. Rigorous mean-field limit and cross-diffusion. *Zeitschrift für angewandte Mathematik und Physik,* 70, no. 122, 2019.

[27] L. Chen, S. Göttlich, and S. Knapp. Modeling of a diffusion with aggregation: rigorous derivation and numerical simulation, *ESAIM: Mathematical Modelling and Numerical Analysis*, 52(2), 567-593, 2018.

[28] L. Chen, V. Gvozdik, A. Holzinger, and Y. Li. Rigorous Derivation of the Degenerate Parabolic-Elliptic Keller-Segel System from a Moderately Interacting Stochastic Particle System. Part II Propagation of Chaos, *arXiv preprint arXiv:2302.08763*, 2023.

[29] L. Chen, V. Gvozdik, and Y. Li. Rigorous Derivation of the Degenerate Parabolic-Elliptic Keller-Segel System from a Moderately Interacting Stochastic Particle System. Part I Partial Differential Equation, *arXiv preprint arXiv:2301.07640*, 2023.

[30] L. Chen, A. Holzinger, A. Jüngel, and N. Zamponi. Analysis and mean-field derivation of a porous-medium equation with fractional diffusion. *Communications in Partial Differential Equations*, 1-53, 2022.

[31] L. Chen, X. Li, P. Pickl, and Q. Yin. Combined mean field limit and non-relativistic limit of Vlasov–Maxwell particle system to Vlasov–Poisson system. *Journal of Mathematical Physics*, 61(6), 061903, 2020.

[32] X. Chen, E. S. Daus, and A. Jüngel. Global existence analysis of cross-diffusion population systems for multiple species. *Archive for Rational Mechanics and Analysis,* 227 (2018), 715–747.

[33] Y.-P. Choi and I.-J. Jeong. Classical solutions for fractional porous medium flow. Submitted for publication, 2021. arXiv:2102.01816.

[34] Y.-P. Choi and I.-J. Jeong. Relaxation to fractional porous medium equation from Euler–Riesz system. Submitted for publication, 2021. arXiv:2102.01817.

[35] S. Cifani and E. Jakobsen. Entropy solution theory for fractional degenerate convection-diffusion equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire* 28 (2011), 413–441.

[36] E. S. Daus, L. Desvillettes, and H. Dietert. About the entropic structure of detailed balanced multi-species cross-diffusion equations. *Journal of Differential Equations,* 266 (2019), 3861–3882.

[37] E. S. Daus, M. Gualdani, and N. Zamponi. Long time behavior and weak-strong uniqueness for a nonlocal porous media equation. *Journal of Differential Equations,* 268 (2020), 1820–1839.

[38] E. S. Daus, M. Ptashnyk, and C. Raithel. Derivation of a fractional cross-diffusion system as the limit of a stochastic many-particle system driven by Lévy noise. Submitted for publication, 2020. arXiv:2006.00277.

[39] D. A. Dawson. Critical dynamics and fluctuations for a mean-field model of cooperative behavior. *Journal of Statistical Physics*, 31(1), 29-85, 1983.

[40] L. Desvillettes, T. Lepoutre, A. Moussa, and A. Trescases. On the entropic structure of reaction-cross-diffusion systems. *Communications in Partial Differential Equations*, 40 (2015), 1705–1747.

[41] R. M. Dudley. Central limit theorems for empirical measures. *The Annals of Probability*, 899-929, 1978.

[42] M. Duerinckx. Mean-field limits for some Riesz interaction gradient flows. *SIAM Journal on Mathematical Analysis* 48 (2016), 2269–2300.

[43] M. Duerinckx, and S. Serfaty. Mean field limit for coulomb-type flows. *Duke Mathematical Journal*, 169(15), 2887-2935, 2020.

[44] L. Evans. An Introduction to Stochastic Differential Equations. *American Mathematical Soc.*, Providence, USA, 1997.

[45] E. Feireisl, and A. Novotný. *Singular Limits in Thermodynamics of Viscous Fluids*. Birkhäuser, Basel 2009.

[46] B. Fernandez, and S. Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. *Stochastic processes and their applications*, 71(1), 33-53, 1997.

[47] A. Figalli, and R. Philipowski. Convergence to the viscous porous medium equation and propagation of chaos. *ALEA Lat. Am. J. Probab. Math. Stat*, 4, 185-203, 2008.

[48] F. Flandoli, and M. Leocata. A particle approach to aggregation phenomena. *Journal of Applied Probability*, 56.1, 282–306, 2019.

[49] F. Flandoli, M. Leimbach, and C. Olivera. Uniform convergence of proliferating particles to the FKPP equation. *Journal of Mathematical Analysis and Applications*, 473, 27–52, 2019.

[50] F. Flandoli, C. Olivera and M. Simon. Uniform Approximation of 2 Dimensional Navier–Stokes Equation by Stochastic Interacting Particle Systems. *SIAM Journal on Mathematical Analysis*, 52.6, 5339-5362, 2020.

[51] J. Fontbona and S. Méléard. Non local Lotka-Volterra system with cross-diffusion in an heterogeneous medium. *Journal of mathematical biology*, 70 (2015), 829–854.

[52] F. Golse. Hydrodynamic limits. In: *European Congress of Mathematics*. Zürich: Eur Math Soc, S. 699-717, 2005.

[53] F. Golse. On the dynamics of large particle systems in the mean field limit. In: *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, pp. 1-144. Springer, Cham, 2016.

[54] F. Golse. The mean-field limit for the dynamics of large particle systems. *Journées Equations aux dérivées partielles* (2003), 1–47.

[55] S. Guo, and D. Luo. Scaling limit of moderately interacting particle systems with singular interaction and environmental noise. Submitted for publication, *arXiv preprint arXiv:2207.03616*, 2022.

[56] M. Hauray, and S. Mischler, S. On Kac's chaos and related problems. *Journal of Functional Analysis*, 266(10), 6055-6157, 2014.

[57] D. Hilbert. Mathematische Probleme. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse*. Vandenhoeck & Ruprecht, 3, 253-297, 1900.

[58] R. A. Holley and D. W. Stroock. Generalized Ornstein-Uhlenbeck processes and infinite particle branching Brownian motions. *Publications of the Research Institute for Mathematical Sciences*, 14(3), 741-788, 1978.

[59] K. Ichikawa, M. Rouzimaimaiti, and T. Suzuki. Reaction diffusion equation with nonlocal term arises as a mean field limit of the master equation. *Discrete Contin. Dyn. Syst.* 5 (2012), 115–126.

[60] C. Imbert, R. Tarhini, and F. Vigneron. Regularity of solutions of a fractional porous medium equation. *Interfaces Free Bound.* 22 (2020), 401–442.

[61] P. E. Jabin, and Z. Wang. Mean field limit and propagation of chaos for Vlasov systems with bounded forces. *Journal of Functional Analysis*, 271(12), 3588-3627, 2016.

[62] P. E. Jabin, and Z. Wang. Mean field limit for stochastic particle systems. *Active Particles, Volume 1: Advances in Theory, Models, and Applications*, 379-402, 2017.

[63] P. E. Jabin, and Z. Wang. Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels. *Inventiones mathematicae*, 214, 523-591, 2018.

[64] B. Jourdain. Convergence of moderately interacting particle systems to a diffusion-convection equation. *Stochastic processes and their applications,* 73, 247–270, 1998.

[65] B. Jourdain, and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Annales de l'Institut Henri Poincare (B) Probability and Statistics,* 34, 727–766, 1998.

[66] M. Kac. Foundations of kinetic theory. In: *Proceedings of The third Berkeley symposium on mathematical statistics and probability*, Vol. 3, pp. 171-197, 1956.

[67] I. Karatzas, and S. Shreve. *Brownian Motion and Stochastic Calculus.* Second edition. Springer, New York, 1991.

[68] P. Kloeden, and E. Platen. *Numerical Solution of Stochastic Differential Equations.* Springer, Berlin, 1992.

[69] N. Kusolitsch. Maß– und Wahrscheinlichkeitstheorie: Eine Einführung. Springer-Verlag, 2014.

[70] C. Lancellotti. On the fluctuations about the Vlasov limit for N-particle systems with mean-field interactions. *Journal of Statistical Physics,* 136 (2009), 643–665.

[71] D. Lazarovici. The Vlasov-Poisson dynamics as the mean field limit of extended charges. *Communications in Mathematical Physics*, 347, 271-289, 2016.

[72] D. Lazarovici, and P. Pickl. A mean-field limit for the Vlasov–Poisson system. *Archive for Rational Mechanics and Analysis* 225, 1201–1231, 2017.

[73] T. Lepoutre, and A. Moussa. Entropic structure and duality for multiple species cross-diffusion systems. *Nonlinear Analysis,* 159 (2017), 298–315.

[74] P. Lewicki. Infinite interacting particle systems: the fluctuation phenomenon. *Demonstratio Mathematica*, 25(3), 623-640, 1992.

[75] E. H. Lieb, and M. Loss. Analysis. Second edition. Amer. Math. Soc., Providence, 2001.

[76] G. M. Lieberman. Second Order Parabolic Differential Equations. World Scientific, Singapore, 1996.

[77] S. Lisini. Fractional high order thin film equation: gradient flow approach. Submitted for publication, 2020. arXiv:2007.00459.

[78] S. Lisini, E. Mainini, and A. Segatti. A gradient flow approach to the porous medium equation with fractional pressure. *Archive for Rational Mechanics and Analysis,* 277 (2018), 567–606.

[79] E. Luçon, and W. Stannat. Transition from Gaussian to non-Gaussian fluctuations for mean-field diffusions in statial interaction. *The Annals of Applied Probability,* 26 (2016), 3840–3909.

[80] A. Majda. *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*. Springer, New York, 1984.

[81] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic processes and their applications*, 95.1 (2001), 109-132.

[82] H. P. McKean. Propagation of chaos for a class of non-linear parabolic equations, In: *Lecture Series in Differential Equations*, Volume 2 (ed. A. K. Aziz), no. 19 in Van Nostrand Mathematical Studies, Van Nostrand Reinhold Company, 177–194, 1969.

[83] S. Méléard. Asymptotic behavior of some interacting particle systems; McKean-Vlasov and Boltzmann models. In: *Probabilistic Models for Nonlinear Partial Differential Equations (Lecture Notes in Mathematics)*, Vol. 1627, Springer, 1996.

[84] S. Méléard, and S. Roelly-Coppoletta. A propagation of chaos result for a system of particles with moderate interaction. *Stochastic processes and their applications,* 26 (1987), 317–332.

[85] S. Mischler, and C. Mouhot. Kac's program in kinetic theory. *Inventiones Mathematicae*, 193 (1), pp.1-147. hal-01067318, 2013.

[86] D. Morale, V. Capasso, and K. Oelschläger. An interacting particle system modelling aggregation behavior: from individuals to populations. *Journal of mathematical biology*, 50 (2005), 49–66.

[87] A. Moussa. From non-local to classical SKT systems: Triangular case with bounded coefficients. *SIAM Journal on Mathematical Analysis* 52 (2020), 42–64.

[88] Q.-H. Nguyen and J. L. Vázquez. Porous medium equation with nonlocal pressure in a bounded domain. *Communications in Partial Differential Equations* 43 (2018), 1502–1539.

[89] D. Nualart. The Malliavin Calculus and Related Topics. Springer, Berlin, 2006.

[90] K. Oelschläger. A fluctuation theorem for moderately interacting diffusion processes. *Probability theory and related fields*, 74(4), 591-616, 1987.

[91] K. Oelschläger. A law of large numbers for moderately interacting diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 69, 279–322, 1985.

[92] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *The Annals of Probability*, 12 (1984), 458–479.

[93] K. Oelschläger. Large systems of interacting particles and the porous medium equation. *Journal of differential equations*, 88 (1990), 294–346.

[94] K. Oelschläger. On the derivation of reaction-diffusion equations as limit dynamics of systems of moderately interacting stochastic processes. *Probability Theory and Related Fields,* 82 (1989), 565–586.

[95] B. Oksendal. Stochastic differential equations: An introduction with applications. Springer Science & Business Media, 2013.

[96] C. Olivera, A. Richard, and M. Tomašević. Quantitative particle approximation of nonlinear Fokker–Planck equations with singular kernel. Submitted for publication, 2020. arXiv:2011.00537.

[97] A. de Pablo, F. Quiros, A. Rodriguez, and J. L. Vázquez. A general fractional porous medium equation. *Communications on Pure and Applied Mathematics,* 65 (2012), 1242–1284.

[98] R. Pinnau, C. Totzeck, O. Tse, and S. Martin. A consensus-based model for global optimization and its mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 27(01), 183-204, 2017.

[99] L. Płociniczak. Derivation of the nonlocal pressure form of the fractional porous medium equation in the hydrological setting. *Communications in Nonlinear Science and Numerical Simulation*, 76 (2019), 66–70.

[100] H. Rost. Hydrodynamik gekoppelter Diffusionen: Fluktuationen im Gleichgewicht. In: *Dynamics and Processes: Proceedings of the Third Encounter in Mathematics and Physics, held in Bielefeld, Germany Nov. 30–Dec. 4*, pp. 97-107. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981.

[101] T. Roubíček. *Nonlinear Partial Differential Equations with Applications.* Birkhäuser, Basel, 2013.

[102] I. Seo. Scaling limit of two-component interacting Brownian motions. *The Annals of Probability*, 46 (2018), 2038–2063.

[103] S. Serfaty. Mean-field limits of the Gross–Pitaevskii and parabolic Ginzburg–Landau equations. *Journal of the American Mathematical Society,* 30 (2017), 713–768.

[104] N. Shigesada, K. Kawasaki, and E. Teramoto. Spatial segregation of interacting species. *Journal of theoretical biology* 79, 83–99, 1979.

[105] J. Sirignano, and K. Spiliopoulos. Mean field analysis of neural networks: A central limit theorem. *Stochastic Processes and their Applications*, 130(3), 1820-1852, 2020.

[106] J. Sirignano, and K. Spiliopoulos. Mean field analysis of neural networks: A law of large numbers. *SIAM Journal on Applied Mathematics*, 80(2), 725-752, 2020.

[107] D. Stan, F. del Teso, and J. L. Vázquez. Finite and infinite speed of propagation for porous medium equations with fractional pressure. *Comptes Rendus Mathematique*, 352 (2014), 123–128.

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

[109] A. Stevens. The derivation of chemotaxis equations as limit dynamics of moderately interacting stochastic many-particle systems. *SIAM Journal on Applied Mathematics,* 61 (2000), 183-212. (Erratum: 61 (2000), 2200-2200.)

[110] P. R. Stinga. User's guide to the fractional Laplacian and the method of semigroups. In: A. Kochubei and Y. Luchko (eds.), *Handbook of Fractional Calculus with Applications*, vol. 2, pp. 235–266. De Gruyter, Berlin, 2019.

[111] A. S. Sznitman. A fluctuation result for nonlinear diffusions. In: *Infinite-Dimensional Analysis and Stochastic Processes* (Bielefeld, 1983), Res. Notes Math. 124, pp. 145–160. Pitman, Boston, 1985.

[112] A. S. Sznitman. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of functional analysis*, 56(3), 311-336, 1984.

[113] A. S. Sznitman. Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX—1989*, 1464, 165-251, 1991.

[114] H. Tanaka. Limit theorems for certain diffusion processes with interaction. In: K. Itô (ed.), *Stochastic Analysis*, Proceedings of the Taniguchi Symposium (Katata 1982), pp. 469–488. North-Holland, Amsterdam, 1984.

[115] J. L. Vázquez. Nonlinear diffusion with fractional Laplacian operators. In: H. Holden and K. Karlsen (eds.). *Nonlinear Partial Differential Equations*, Abel Symposia, vol. 7, pp. 1–27. Springer, Berlin, 2012.

[116] J. L. Vázquez. The mathematical theories of diffusion. Nonlinear and fractional diffusion. In: M. Bonforte and G. Grillo (eds.), *Nonlocal and Nonlinear Diffusions and Interactions: New Methods and Directions*. Lect. Notes Math. 2186, pp. 205–278. Springer, Cham, 2017.

[117] Z. Wang, X. Zhao, and R. Zhu. Gaussian fluctuations for interacting particle systems with singular kernels. *arXiv preprint arXiv:2105.13201*, 2021.

[118] E. Weinan. Dynamics of vortex liquids in Ginzburg–Landau theories with applications in superconductivity. *Physical Review B,* 50 (1994), 1126–1135.

[119] N. Zamponi and A. Jüngel. Analysis of degenerate cross-diffusion population models with volume filling. *Annales de l'Institut Henri Poincaré C, Analyse Non Linéaire* 34 (2017), 1–29.

Curriculum Vitae

Personal data

Name	**Alexandra Holzinger**
Place of birth	Vienna, Austria
Working address	Institute for Analysis and Scientific Computing, Wiedner Hauptstraße 8-10, 1040 Vienna, Austria
E-mail	`alexandra.holzinger@tuwien.ac.at`

Academic Education

since 02/2023	Project-Assistant (Forschungsgruppe Analysis nichtlinearer PDEs), TU Wien, Austria
04/2019–02/2023	University-Assistant (Prae-Doc), TU Wien, Austria.
06/2016–03/2019	MSc Technical Mathematics with focus on Applied Mathematics, TU Wien, Austria
10/2012 - 06/2016	BSc Technical Mathematics, TU Wien, Austria.
06/2012	High school diploma (Matura), Gymnasium Erlgasse XII. Wien, Austria.

Publications

Dvorak, D., Bäuml, T., Holzinger, A., Popp, H. (2017). A comprehensive algorithm for estimating lithium-ion battery parameters from measurements. IEEE Transactions on Sustainable Energy, 9(2), 771-779.

Chen, L., Daus, E. S., Holzinger, A., Jüngel, A. (2021). Rigorous derivation of population cross-diffusion systems from moderately interacting particle systems. Journal of Nonlinear Science, 31(6), 1-38.

Chen, L., Holzinger, A., Jüngel, A., Zamponi, N. (2022). Analysis and mean-field derivation of a porous-medium equation with fractional diffusion. Communications in Partial Differential Equations, 1-53.

Wien, am 8. Mai 2023

Alexandra Holzinger