Beginner's guide to aggregation-diffusion equations

David Gómez-Castro[1]

Abstract

The aim of this survey is to serve as an introduction to the different techniques available in the broad field of aggregation-diffusion equations. We aim to provide historical context, key literature, and main ideas in the field. We start by discussing the modelling and famous particular cases: heat equation, Fokker–Plank, Porous medium, Keller–Segel, Chapman–Rubinstein–Schatzman, Newtonian vortex, Caffarelli–Vázquez, McKean–Vlasov, Kuramoto, and one-layer neural networks. In Sect. 4 we present the well-posedness frameworks given as PDEs in Sobolev spaces, and gradient-flow in Wasserstein. Then we discuss the asymptotic behaviour in time, for which we need to understand minimisers of a free energy. We then present some numerical methods which have been developed. We conclude the paper mentioning some related problems.

Contents

✉ David Gómez-Castro
   david.gomezcastro@uam.es

[1] Departamento de Matemáticas, Universidad Autónoma de Madrid, Madrid, Spain

1 Introduction

The aim of this survey is to serve as a general introduction to the many models and techniques used to study the *aggregation-diffusion* family of equations

$$\frac{\partial \rho}{\partial t} = \mathrm{div}\left(\rho \nabla (U'(\rho) + V + W * \rho)\right). \tag{ADE}$$

This survey does not intend to provide an encyclopedic presentation including all possible references and results in maximal generality (this would be impossible), but to give newcomers a general overview of the subject.

2 Modelling

Let us discuss first the modelling, and then we devote Sect. 3 to several modelling problems which belong to this family. On the one hand, diffusion is usually modelled as a conservation law.

2.1 Continuity equations

Let ρ be a density and $\omega \subset \mathbb{R}^d$ any control volume, if $\mathbf{F}$ is the out-going flux

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\omega \rho \, \mathrm{d}x = - \int_{\partial\omega} \mathbf{F} \cdot \mathbf{n} \, \mathrm{d}S = - \int_\omega \mathrm{div} \, \mathbf{F} \, \mathrm{d}x \tag{2.1}$$

Hence we arrive at the *continuity equation*

$$\frac{\partial \rho}{\partial t} = -\mathrm{div} \, \mathbf{F}. \tag{2.2}$$

For the transport of heat, we use *Fourier's law* to model the flux $\mathbf{F} = -D\nabla\rho$ and yields the heat equation

$$\frac{\partial \rho}{\partial t} = D\Delta\rho. \tag{HE}$$

A generalisation of this problem which is used to the flow a gas through porous media. We derive the model as mass balance by writing the flux in terms of a velocity $\mathbf{v}$, i.e., $\mathbf{F} = -\rho\mathbf{v}$; we use *Darcy's law* to relate the velocity with the pressure p as $\mathbf{v} = -\frac{k}{\mu}\nabla p$; and lastly we use a general *state equation* $p = \phi(\rho)$, where for perfect gases $\phi(\rho) = p_0\rho^\gamma$. Eventually, we recover that $\mathbf{F} = -\nabla\Phi(\rho)$ for some non-decreasing $\Phi : \mathbb{R} \to \mathbb{R}$, and this yields the so-called *porous medium equation*

$$\frac{\partial \rho}{\partial t} = \Delta\Phi(\rho). \tag{PME$_\Phi$}$$

It is common to select $\Phi(\rho) = \rho^m$ for $m > 0$. The most complete reference for this equation is [171]. Notice $\Delta\Phi(\rho) = \mathrm{div}(\Phi'(\rho)\nabla\rho)$ so we take

$$U''(\rho) = \frac{\Phi'(\rho)}{\rho}. \tag{2.3}$$

The classical modelling also allows for a transport velocity field $\mathbf{v}(t, x)$ to be included. For example, if a pollutant is being transported by water, then $\mathbf{v}(t, x)$ is the velocity of the water. The flux of mass being moved is recovered from the density and the velocity as $\rho\mathbf{v}$. A general flux then looks like

$$\mathbf{F} = -\nabla\Phi(\rho) + \rho\mathbf{v}.$$

Remark 2.1 We do not discuss the case of sign-changing solutions. In that setting u^m must be replaced by $|u|^{m-1}u$.

2.2 Particle systems

We can also understand transport from the particle perspective. Consider a known velocity field $\mathbf{v}(t, x)$. Consider N particles with positions $X_i(t)$ of equal masses $1/N$ moving in the velocity field

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \mathbf{v}(t, X_i(t)), \quad i = 1, \dots, N.$$

Notice that the evolution of the particles is decoupled. To recover a PDE we consider the so-called *empirical distribution* defined as

$$\mu_t^N = \sum_{j=1}^{N} \frac{1}{N}\delta_{X_j(t)}, \tag{2.4}$$

where δ_x is the Dirac delta at a point x. First, notice that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} \varphi(t, x)\,\mathrm{d}\mu_t^N(x) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{N}\sum_j \varphi(t, X_j(t))$$

$$= \frac{1}{N}\sum_j \left(\frac{\partial\varphi}{\partial t}(t, X_j(t)) + \nabla\varphi(t, X_j(t)) \cdot \frac{\mathrm{d}X_j}{\mathrm{d}t}\right)$$

$$= \frac{1}{N}\sum_j \left(\frac{\partial\varphi}{\partial t}(t, X_j(t)) + \nabla\varphi(t, X_j(t)) \cdot \mathbf{v}(t, X_j(t))\right)$$

$$= \int_{\mathbb{R}^d} \left(\frac{\partial\varphi}{\partial t}(t, x) + \nabla\varphi(t, x) \cdot \mathbf{v}(t, x)\right)\mathrm{d}\mu_t^N(x)$$

If $\varphi \in C_c^1((0, \infty) \times \mathbb{R}^d)$ (i.e., $\varphi(0, x) = 0$ and for T large enough $\varphi(t, x) = 0$ for $t \geq T$) we integrate in time to recover

$$\int_0^\infty \int_{\mathbb{R}^d} \left(\frac{\partial\varphi}{\partial t} + \mathbf{v} \cdot \nabla\varphi\right)\mathrm{d}\mu_t^N\,\mathrm{d}t = 0.$$

Integrating by parts, we observe that this is the distributional formulation of the *continuity equation*

$$\partial_t\mu^N + \mathrm{div}(\mu^N\mathbf{v}) = 0.$$

Aggregation and confinement can be modelled by considering *interacting particles* where the velocity field is aware of other particles

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = -\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{1}{N} \nabla W(X_i - X_j) - \nabla V(X_i), \quad i = 1, \ldots, N.$$

The first term models the interactions between particles, with attract/repel each other as a function of their relative position, and the second term models a field attracting/repelling every particle. Noticing that

$$\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{1}{N} \nabla W(x - X_j(t)) = \int_{\mathbb{R}^d} \nabla W(x - y)\, \mathrm{d}\mu_t^N(y)$$

we deduce from the previous reasoning that the empirical measure μ^N is a distributional solution of the *aggregation-confinement equation*

$$\partial_t \mu = \mathrm{div}(\mu \nabla(W * \mu + V)). \tag{2.5}$$

Linear diffusion can be added to the particle system by introducing noise and working with stochastic ODEs and a mean-field approximation (see Sect. 3.8).

As we will see below, special attention has been paid to the case $V = 0$ which is simply called *aggregation equation*

$$\frac{\partial \rho}{\partial t} = \mathrm{div}(\rho \nabla W * \rho). \tag{AE}$$

As a toy model we will also discuss the *confinement problem*

$$\frac{\partial \rho}{\partial t} = \mathrm{div}(\rho \nabla V). \tag{CE}$$

Joining the many-particle approximation for aggregation with the porous medium diffusion we recover (ADE). If we look for solutions defined in the whole space, then it is natural to consider only the case of finite mass. Hence, we assume $\rho \in L^1(\mathbb{R}^d)$ and, in some sense,

$$\rho_t(x) \to 0 \text{ as } |x| \to \infty. \tag{2.6}$$

2.3 Boundary conditions in bounded domains of $\mathbb{R}^d$

The problems above are often posed in the whole space $\mathbb{R}^d$. However, sometimes it is convenient to restrict them to bound domains, specially for the numerical analysis. Here and below we will denote by Ω an open and bounded set of $\mathbb{R}^d$ with a smooth boundary. Although it is possible to study these kinds of problems in less smooth domain (Lipschitz boundary, interior point conditions, etc...) this introduces additional difficulties we will not discuss. Since with this kind of equation we intend to model systems that conserve mass, there are two natural approaches.

2.3.1 No-flux conditions

We can establish that no mass will cross the boundary of Ω, $\partial\Omega$. Going back to (2.1), a natural way is to establish no-flux conditions $\mathbf{F} \cdot \mathbf{n} = 0$ on $\partial\Omega$. With our choice of flux this means

$$\rho\nabla(U'(\rho) + V + W * \rho) \cdot \mathbf{n} = 0 \quad \text{for } t > 0 \text{ and } x \in \partial\Omega \tag{BC}$$

where again ρ_0 is known, and we have to be a little careful with the notion of convolution when ρ is only defined in Ω. For this, we consider the extended notion

$$W * \rho \equiv W * \widetilde{\rho}, \quad \text{where } \widetilde{\rho} = \begin{cases} \rho \text{ in } \Omega, \\ 0 \text{ elsewhere.} \end{cases}$$

In fact, the problem (ADE) with general V and W in a bounded domain is rather delicate in terms of a priori estimates (see Sect. 4.2.3). It is sometimes preferable to discuss the generalisation

$$\frac{\partial\rho}{\partial t} = \text{div}\left(\rho\nabla(U'(\rho) + V + \mathcal{K}\rho)\right), \quad \text{where } \mathcal{K}\rho(x) = \int_\Omega K(x, y)\rho(y)\,\mathrm{d}y, \quad \text{(ADE*)}$$

and with the corresponding boundary conditions

$$\rho\nabla(U'(\rho) + V + \mathcal{K}\rho) \cdot \mathbf{n} = 0, \quad \text{on } \partial\Omega. \tag{BC*}$$

If Ω is bounded we will choose

$$\nabla V \cdot \mathbf{n} = 0 \text{ in } \partial\Omega, \quad \text{and} \quad K \in C_c(\Omega \times \Omega). \tag{2.7}$$

A very typical example is

$$K(x, y) = \eta(x)W(x - y)\eta(y) \tag{2.8}$$

with $\eta \in C_c^\infty(\Omega)$.

2.3.2 Periodic boundary conditions

Another approach is to stablish that any mass that crosses the boundary will "come back" on the opposite side. This happens, for example, if all data are periodic. Some papers have studied periodic solutions defined over the unit cube $[0, 1]^d$ (or, equivalently, the torus $\mathbb{T}^d$). In $d = 1$ this is equivalent to studying solutions defined over the unit circle $\mathbb{S}^1$. This is the case of the Kuramoto model described below. An example of the case $d = 2$ can be found in [64].

2.4 Free energy

We conclude this section by introducing a functional which will be very useful below. We will show how (ADE) is related to the free energy

$$\mathcal{F}(\rho) = \int_\Omega U(\rho) + \int_\Omega V\rho + \frac{1}{2}\iint_{\Omega\times\Omega} W(x - y)\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y, \tag{FE}$$

whereas (ADE*) is related to the more general free energy

$$\mathcal{F}(\rho) = \int_\Omega U(\rho) + \int_\Omega V\rho + \frac{1}{2}\iint_{\Omega\times\Omega} K(x, y)\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y. \tag{FE*}$$

For convenience, let us define the function

$$\frac{\delta \mathcal{F}}{\delta \rho}[\rho] := U'(\rho) + V + \int_\Omega K(\cdot, y)\rho(y)\,dy. \tag{2.9}$$

This is called *first variation of the free energy*, and we will explain its importance below in Sect. 4.2.4. Hence, (ADE*) can be re-written as

$$\frac{\partial \rho}{\partial t} = \operatorname{div}\left(\rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}\right) \tag{2.10}$$

for $\mathcal{F}$ given by (FE).

Occasionally, it will be convenient for us to use the free energies of each of the terms. Hence, we define

$$\mathcal{U}[\rho] = \int_{\mathbb{R}^d} U(\rho), \qquad \mathcal{V}[\rho] = \int_{\mathbb{R}^d} V\rho,$$

$$\mathcal{W}[\rho] = \frac{1}{2} \iint_{\mathbb{R}^d} W(x - y)\rho(x)\rho(y)\,dx\,dy. \tag{2.11}$$

2.5 A comment on mass conservation

Our evolution problem need, of course, an initial state ρ_0. Throughout this survey we will usually restrict ourselves non-negative unit mass initial data, i.e.,

$$\int_{\mathbb{R}^d} \rho_0(x) = 1. \tag{U_0}$$

All the equations presented above are of "conservation type". This means we expect preservation of mass, i.e., that provided unit initial mass (U_0) then there will be non-negative unit initial mass for all times, i.e.,

$$\int_{\mathbb{R}^d} \rho_t = 1, \quad \forall t > 0. \tag{U}$$

3 Famous particular cases

3.1 Heat equation and Fokker–Plank

As explained above, the heat Eq. (HE) is the most classical example in our family of equations. It corresponds to (ADE) with the choices $U = \rho \log \rho$ and $V = W = 0$. In $\mathbb{R}^d$ It is well known that (HE) admits the Gaussian solution

$$K_t(x) = (4\pi t)^{-\frac{d}{2}} \exp\left(-\frac{|x|^2}{4t}\right).$$

All solutions (with suitable initial data) are precisely of the form

$$\rho_t = K_t * \rho_0,$$

and it is not difficult to check that

$$\|\rho_t\|_{L^1} = M, \quad \|\rho_t - MK_t\|_{L^1} \to 0 \quad as\ t \to \infty.$$

A good survey on the matter can be found in [172].

The fundamental solution K_t is of the so-called *self-similar* type

$$K_t(x) = A(t)F\left(\frac{x}{\sigma(t)}\right). \tag{SS}$$

In this expression, F is usually called self-similar profile and $\sigma(t)$, $A(t)$ the scaling parameters. Due to mass conservation the natural choice is $A(t) = \sigma(t)^{-d}$. Plugging (SS) into (HE) and setting the total mass $\int K_t = 1$ yields the Gaussian. Since (HE) is linear, MK_t is a solution of mass M. Usually, the self-similar solution has $\sigma(t)$ increasing, and it only makes sense to consider $\sigma(t) \geq 0$. Hence, we make the choice $\sigma(0) = 0$ that corresponds to $K_t = \delta_0$, a Dirac delta at 0. This type of solutions with $\rho_0 = \delta_0$ are called *source-type solutions*.

This nice convolutional representation is not available in more general settings. There are several other ways to study the asymptotic behaviour. One of the key tricks is the study of the self-similar change of variable

$$\rho(x, t) = u(y, \tau)(1 + t)^{-\frac{d}{2}}, \quad 2\tau = \log(1 + t), \quad y = (1 + t)^{-\frac{1}{2}}x,$$

which leads to the so-called linear Fokker–Planck equation

$$\frac{\partial u}{\partial \tau} = \Delta_y u + \operatorname{div}(uy). \tag{HE-FP}$$

This equation corresponds to

$$U_1(\rho) = \rho \log \rho,$$

$V = \frac{|x|^2}{2}$ and $W = 0$. Notice that stationary states can be obtained as solutions of the equation for the flux $\log u + \frac{|x|^2}{2} = -h$. This yields the Gaussian profile

$$\widehat{u}(y) = AG(y), \quad G(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{|y|^2}{2}}. \tag{3.1}$$

The constant A is the unique value such that $\int \rho_t = M$. In original variable, we recover MK_t.

3.2 Porous medium equation

As presented in the introduction, the most common example of (PME$_\Phi$) is the porous medium equation

$$\frac{\partial \rho}{\partial t} = \Delta \rho^m. \tag{PME}$$

The range $m \in (0, 1)$ is sometimes called *fast diffusion equation*. The (PME) corresponds, for $m \neq 1$, to

$$U_m(\rho) = \frac{\rho^m}{m - 1}$$

and $V = W = 0$. Notice that as $m \to 1$ we have $\rho^{m-1}/(m-1)$ tends to $\log \rho$ and hence U_m tends to U_1. Notice that, due to the homogeneity of the equation, after rescaling space and time, we can preserve the equation (without introducing constants) and can work with (U_0).

Similarly to the heat equation, a suitable change of variable (detailed below in Sect. 5.1.1) converts (PME) into

$$\frac{\partial u}{\partial t} = \Delta u^m + \text{div}(ux) \tag{PME-FP}$$

This corresponds to $U = U_m$ and $V = \frac{|x|^2}{2}$ and $W = 0$. (PME-FP) admits a stationary solution given by the so-called Barenblatt profiles

$$\hat{u}(x) = F_B(x), \quad F_B(x) := \left(C - \tfrac{m-1}{2m}|x|^2\right)_+^{\frac{1}{m-1}}, \tag{3.2}$$

where $C \geq 0$. For $m > 1$ they have compact support, and for $m \in (0, 1)$ they have power-like tails. To check whether finite-mass states are available we compute

$$\int_{\mathbb{R}^d} F_B \approx c + \int_1^\infty r^{\frac{2}{m-1}} r^{d-1} \, \mathrm{d}r.$$

Hence, finite mass steady states are only when $m > \frac{d-2}{d}$. This number is critical, as well will see below. Undoing the change of variables we arrive at the self-similar solution

$$B_t = t^{-\alpha}(C - k|x|^2 t^{-2\beta})_+^{\frac{1}{m-1}}, \quad \text{where } \alpha = \tfrac{d}{d(m-1)+2}, \ \beta = \tfrac{\alpha}{d}, \ k = \tfrac{\alpha(m-1)}{d}. \tag{ZKB}$$

This solution is usually denoted ZKB after Zel'dovich and Kompaneets, and Barenblatt (see [171] and the references therein). Usually this is shortened to Barenblatt. When $m > \frac{d-2}{d}$ there is a unique constant C so that (U). It can be explicitly computed (see [171, Section 17.5]).

As we let $m \searrow \frac{d-2}{d}$ we get $C \nearrow \infty$, so for $m = \frac{d-2}{d}$ we expect $B_t = \delta_0$. In fact, Brezis and Friedman [32] showed that if $m \leq \frac{d-2}{d}$ and $\rho_0^{(j)} \in L^1(\mathbb{R}^d)$ are a sequence of initial data converging to δ_0, then the corresponding solutions converge to $\rho^{(j)}(t, x) \to \delta_0(x) \otimes 1(t)$. Therefore, *very fast diffusion cannot diffuse a δ_0*. This is because the diffusion coefficient is $m\rho^{m-1}$. For $m \ll 1$ this means that small values of ρ are indeed diffused fast (hence the name of "fast diffusion"). However, in regions where ρ is large the diffusion is slow.

It is also worth pointing to the equivalent formulation in the so-called pressure variable $u = U'(\rho)$. We recover the Hamilton-Jacobi type equation

$$\frac{\partial u}{\partial t} = \Phi'(\rho)\Delta u + |\nabla u|^2,$$

where we recall the relation (2.3). Notice that $\Phi'(\rho)$ can be written as a non-linear function of u. This formulation is convenient when $m > 1$ since $u = \frac{m}{m-1}\rho^{m-1}$. This presentation is much better for the study of viscosity solutions (see [31, 40]).

First results of asymptotic behaviour for $m > 1$ were obtained by regularity, self-similarity, and comparison arguments in the late 70 s and early 80 s (see [121] and the references therein). Self-similar analysis and relative entropy arguments were later developed to improve the theoretical understanding (see Sects. 5.1.1 and 5.2.3). For a detailed explanation of the theory see [170, 171].

Remark 3.1 (HE) and (PME) are "diffusive". This can be seen in different directions. For example, at a point (t_0, x_0) of local maximum of ρ we get $\Delta\rho^m \leq 0$ and hence $\frac{\partial\rho}{\partial t} \leq 0$, i.e., *the maximum values decays*. This is strongly related to the *maximum principle*. The converse happens to the minimum if it exists (notice that non-negative solutions in $\mathbb{R}^d$ will tend to 0 as $|x| \to \infty$).

Remark 3.2 Let us recall some "numerological values". The value $m = 1 - \frac{2}{d}$ is the critical value for finite-mass Barenblatt. It is also be the critical value below which there is finite-time extinction (see Sect. 4.2.1). The (3.2) has finite λ-moment, i.e.,

$$\int_{\mathbb{R}^d} |x|^\lambda B_t(x)\,\mathrm{d}x < \infty$$

if and only if $m > 1 - \frac{2}{d+\lambda}$. In Sect. 4.3.4 we will present a third relevant value $m = 1 - \frac{1}{d}$ coming from the Wasserstein gradient-flow analysis.

Remark 3.3 We will not discuss here the limit $m \to 0$. In this direction there are two possibilities. On the one hand, some authors have discussed the rescaled limit $\frac{u^m}{m} \to \log u$, and we arrive at the logarithmic diffusion equation $\partial_t u = \Delta \log(u)$. We point to [170, Section 8.2] for a discussing on this topic and [170, Section 8.4] for historical notes. On the other hand, some authors choose to look at the limit $|u|^{m-1}u \to \mathrm{sign}(u)$ leads to the so-called Sign Fast Diffusion Equation $\partial_t u = \Delta \mathrm{sign}(u)$. Some results in this direction are available [30].

3.3 Transport problem with known potential

To understand the transport terms in the (ADE) family, we will discuss the toy problem

$$\frac{\partial \rho}{\partial t} = \mathrm{div}(\rho \nabla V)$$

where V is known and does not depend on t. Imagine that $\rho_0 \geq 0$ is radially symmetric function. Computing the balance of mass of a ball B_r we get

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{B_r} \rho_t = \int_{B_r} \mathrm{div}(\rho \nabla V) = \int_{\partial B_r} \rho \nabla V \cdot x. \tag{3.3}$$

If $\nabla V \cdot x \leq 0$ all balls B_r are "loosing mass", i.e., then mass if always flowing "away" from 0 and towards infinity. Hence, we can think of these kinds of potentials as "diffusive" (but not as strongly as in the sense of Remark 3.1). We will show below by explicitly solving these problems that potentials such that $\nabla V \cdot x \geq 0$ will attract mass towards 0, asymptotically (or even instantaneously) creating a singularity.

3.4 Keller–Segel model

The Keller–Segel [133] (or Patlak–Keller–Segel model [149]) model for describes the motion of cells by chemotactical attraction by means of the coupled system

$$\begin{cases} \frac{\partial \rho}{\partial t} = \Delta \rho - \mathrm{div}(\rho \nabla V_t) \\ -\Delta V_t = \rho_t. \end{cases} \tag{KSE}$$

This model was first studied for $d = 2$. Some authors replace the second equation by a more general evolutionary problem $\varepsilon \frac{\partial V_t}{\partial t} - \Delta V_t + \alpha V_t = \rho_t$. We will discuss (KSE), which can be thought as the limit $\varepsilon, \alpha \to 0$. This simplification was first introduced by [146] to model the case where the diffusion of the chemical is much faster than its production.

So far, this model is not in the (ADE) family. However, the can analysis the equation for V_t to recover some information. In $\mathbb{R}^d$ we can use the Fourier transform F for the elliptic problem to recover

$$|\xi|^2 F[V_t](\xi) = F[\rho_t]$$

and hence we can solve and recover $V_t = W_N * \rho$, where $W_N = F^{-1}[|\xi|^{-2}]$. In fact, we have the closed expression

$$W_N(x) = \begin{cases} -\frac{1}{2\pi} \log |x| & \text{if } d = 2, \\ \frac{1}{d(d-2)\omega_d} |x|^{2-d} & \text{if } d > 2. \end{cases} \tag{3.4}$$

Notice that $-\Delta W_N = \delta_0$. The so-called *Newtonian potential* corresponds to $N = -W_N$. Eventually, we can write this problem as (ADE) where $U = \rho \log \rho$, $V = 0$ and $W = -W_N$.

We can think about the convolution term $W * \rho_t$ as an aggregation potential. The justification is the following. If ρ_t is non-increasing, then $V_t = W * \rho_t$ is also non-increasing, and hence $\nabla V_t \cdot x \le 0$. Going back to Sect. 3.3, this is what we called aggregation potentials. Unfortunately, we do not have in general the sign $\nabla V_t \cdot x \le 0$ for all t, $x > 0$. In fact, (KSE) does not "prefer" $x = 0$ from other points.

Changing the sign of the aggregation to $\frac{\partial \rho}{\partial t} = \Delta \rho + \mathrm{div}(\rho \nabla V_t)$ makes the problem more regularising. In [24] the author proves global well-posedness of that problem for $d = 3$ in a bounded domain with no-flux condition was shown to have.

Sometimes it is convenient to express the problem in a single equation using operator notation:

$$\frac{\partial \rho}{\partial t} = \Delta \rho - \mathrm{div}(\rho \nabla (-\Delta)^{-1} \rho).$$

When written in this form, the authors usually allow for any initial mass $\|\rho_0\|_{L^1}$. We can rescale to impose (U₀). If we let $\chi = \|\rho_0\|_{L^1}$ and $u(t, x) = \rho(t, x)/\chi$ we arrive at the equation

$$\frac{\partial u}{\partial t} = \Delta u - \chi \, \mathrm{div}(u \nabla (-\Delta)^{-1} u). \tag{KSE$_\chi$}$$

In bounded domains some authors introduce Neumann boundary conditions $\frac{\partial u}{\partial n} = \frac{\partial V}{\partial n} = 0$. The authors of [129] extend the problem to $-\Delta V = \beta u - \mu$ which a few authors later called Jäger–Luckhaus problem (e.g., [156]). This is equivalent to replacing $(-\Delta)^{-1}$ by $(-\Delta + \mathrm{id})^{-1}$.

Some authors later introduced the Keller–Segel model with porous medium diffusion. One of the model arguments is that $m > 1$ would deal with "over-crowding" (i.e., the formation of Dirac deltas). We arrive at the problem

$$\frac{\partial \rho}{\partial t} = \Delta \rho^m - \chi \, \mathrm{div}(\rho \nabla (-\Delta)^{-1} \rho). \tag{KSE$_{m,\chi}$}$$

This model has the advantage of not having finite time blow-up (see [44]). Existence and uniqueness of solutions for $m > 1$ and a family of W covering W_N can be found in [15]. Some authors replace $(-\Delta)^{-1}$ by $(-\Delta + \gamma \mathrm{id})^{-1}$, see, e.g., [142, 163–165].

As we will discuss below, (KSE$_{m,\chi}$) admits global solutions in the sub-critical range $m > 2 - \frac{2}{d}$. However, for $m < 2 - \frac{2}{d}$ there is finite-time blow-up. In the critical case blow-up depends on χ. Curiously enough, the first case studied was $d = 2$ and $m = 1$, which is critical. A good explanation of this dichotomy is the free-energy minimisation argument which we will provide in Sect. 6.3.

3.5 Chapman–Rubinstein–Schatzman

In the context of superconductivity [90] introduced a mean-field model for "the motion of rectilinear vortices in the mixed state of a type-II semiconductor". They argue this is the limit where the Ginzburg–Landau parameter κ tends to $+\infty$ and the number of vortices becomes large one recovers the problem

$$\begin{cases} \frac{\partial \rho}{\partial t} = \operatorname{div}(|\rho|\nabla V_t) \\ -\Delta V_t + V_t = \rho_t. \end{cases} \tag{CRSE}$$

The model was first introduced in $\mathbb{R}^2$. Then, the authors also make a particularisation to a bounded domain Ω where they introduce Dirichlet conditions for $V_t(x) = h_0$ on $\partial\Omega$, and flux conditions setting the number of vortices crossing the boundary $|\rho|\nabla v \cdot n$ known.

In fact, when $\rho \geq 0$ it becomes of our class (ADE) (or (ADE*) in bounded domains). On $\mathbb{R}^d$ we can solve as before to recover $V_t = W * \rho_t$ where $W = \mathrm{F}^{-1}[(1 + |\xi|^2)^{-1}]$. As long as $\rho \geq 0$, this model corresponds to (ADE) where $U, V = 0$. If we work in a bounded domain Ω the homogeneous Dirichlet condition we recover (ADE*) where $U, V = 0$ and K is the Green kernel of the equation. Notice that K is not a compactly supported in $\Omega \times \Omega$.

There is extensive literature on this problem. First, there were different works applying PDE theory. First, [90] proved existence of the problem posed in bounded domains. Then in [127] the authors work with non-negative ρ_0 and compactly supported, using generalised characteristics. We also point to [154] where the authors perform a vanishing viscosity analysis in bounded domains.

Later, some authors applied optimal transport methods, in particular the 2-Wasserstein gradient-flow approach. In [5] the authors worked in with non-negative probability measures in a bounded domain, i.e., $\mathcal{P}_2(\overline{\Omega})$. Later, the work was extended in [4] to a bounded domain and signed measures. They extend the Wasserstein distance to signed measures and in fact they minimise a localised energy in the space of signed unit measures in $\mathbb{R}^d$ to recover a problem of the form

$$\frac{\partial \rho}{\partial t} = \operatorname{div}(\mathbf{1}_\Omega |\rho| \nabla V_t), \quad \text{for } t > 0, x \in \mathbb{R}^d.$$

This sparked broader interest in the family of aggregation equations (AE).

3.6 Newtonian vortex problem

In [139] the authors justify that a different limit from Gizburg–Landau equations leads to the problem without zero-order term

$$\begin{cases} \frac{\partial \rho}{\partial t} = \operatorname{div}(\rho \nabla V_t) \\ -\Delta V_t = \rho. \end{cases} \tag{NVE}$$

Again, this can be set in $\mathbb{R}^d$ or in a bounded domain (with suitable boundary conditions).

This problem also attracted several authors, who brought different approaches. [139] the authors construct solutions by characteristics with the so-called "vortex-blob" method. In [143] the authors prove existence of renormalised solutions. The authors of [21] study positive and negative solutions. This is equivalent to studying non-negative solutions of the problems

$$\frac{\partial \rho}{\partial t} = \pm \operatorname{div}(\rho \nabla W_{\mathrm{N}} * \rho).$$

The $W = -W_N$ can also be seen of the problem $W = W_N$ but taken backwards in time. In fact, the authors prove that the $+$ sign leads to asymptotic dispersion whereas $-$ leads to blow-up.

3.7 The Caffarelli–Vázquez problem

In [38], Caffarelli and Vázquez introduced a non-local porous medium-type equation given by

$$\begin{cases} \frac{\partial \rho}{\partial t} = \mathrm{div}(\rho \nabla V_t) \\ (-\Delta)^s V_t = \rho, \end{cases}$$

where $(-\Delta)^s$ denotes the fractional Laplacian. In $\mathbb{R}^d$ this operator is to the operational fractional power of the Laplacian, which can be rigorously defined as the operator with Fourier symbol $|\xi|^{2s}$. This problem is often called *Caffarelli–Vázquez problem* and is often written in the compact form

$$\frac{\partial \rho}{\partial t} = \mathrm{div}(\rho \nabla (-\Delta)^{-s} \rho). \tag{CVE}$$

This is a "diffusive" operator. To solve the fractional Laplace equation we take $W = \mathrm{F}^{-1}[|\xi|^{-2s}]$, and we recover the so-called *Riesz potential*

$$W_R(x) = c_{d,s} \begin{cases} |x|^{2s-d} & \text{if } 2s < d, \\ \log |x| & \text{if } 2s = d \end{cases}$$

The properties convolution properties of the Riesz potential where already well known in harmonic analysis. There has been a lot of analysis for this equation: asymptotic analysis [41], sign-changing case and self-similar solutions [25], regularity theory [37, 39], and in [157] that the limit as $s \to 1$ is (NVE). In [140] the authors construct a solution through the Wasserstein gradient-flow theory described in Sect. 4.3.

3.8 McKean–Vlasov and Kuramoto problems

The McKean–Vlasov process is given by the stochastic differential equation

$$\mathrm{d}X_t = a_t \, \mathrm{d}\mathcal{B}_t + b_t \, \mathrm{d}t$$

where a_t, b_t may depend on X_t and its law, and $\mathcal{B}_t$ is a Wiener process, i.e., we introduce *Brownian motion*. We can extend this idea to a system of N particles driven by coupled McKean–Vlasov equations. If we consider that a_t is constant, say $a_t = \sqrt{2\beta^{-1}}$, and b_t is given by the interaction we can arrive at

$$\mathrm{d}X_t^{(i)} = -V(X_t^{(i)}) - \frac{1}{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \nabla W(X_t^{(i)} - X_t^{(j)}) \, \mathrm{d}t + \sqrt{2\beta^{-1}} \, \mathrm{d}\mathcal{B}_t.$$

Through arguments of mean-field theory (or *propagation of chaos* see, e.g., [128]), one arrives asymptotically as $N \to \infty$ to (ADE) with linear diffusion $U = \beta \rho \log \rho$ (see, e.g., [64]). This is the reason the problem

$$\frac{\partial \rho}{\partial t} = \beta^{-1} \Delta \rho + \mathrm{div}(\rho \nabla (V + W * \rho)) \tag{McKVE}$$

is sometimes called *McKean–Vlasov problem* (see, e.g., [64]). However, this denotation is not universal (see, e.g., [61]).

Kuramoto [135] introduced a model for synchronisation of chemical and biological oscillators, and it has been well-received by the community in neuroscience. It is a particular case of McKean–Vlasov problem in $d = 1$ where $X_t^{(i)}$ models the phase of a coupled oscillator. Therefore, we take $X_t^{(i)} \in [0, 2\pi)$ or, equivalently, periodic boundary condition. The particular choices are $V(x) = -\omega_i$ the natural frequency of the system, and $W(x) = \sin(x)$. The resulting (ADE) is known as the *Kuramoto model*

$$\begin{cases} \frac{\partial \rho}{\partial t} = \beta^{-1} \Delta \rho + \operatorname{div}(\rho(-\omega + \sin *\rho)) & t > 0, x \in (0, 2\pi), \\ \rho_t \text{ is } 2\pi - \text{periodic.} \end{cases} \tag{KE}$$

The case $\beta^{-1} = 0$ is also interesting. An interesting analysis of this problem without diffusion (i.e., $\beta^{-1} = 0$) can be found in [48].

3.9 The neural networks of machine learning

A neural network is no more than a specific type of parametric function accepting an input vector $x \in \mathbb{R}^d$ and returning $y \in \mathbb{R}$. If the output is higher dimensional, then we can create a neural network for each output. In this context "training" corresponds to optimising the parameters to fit a set of prescribed data (supervised learning) or to achieve some objective (unsupervised learning). Neural networks are form by connecting so-called *perceptron*. A perceptron is a parametric function of the form

$$x \longmapsto \sigma(\theta_1 \cdot x + \theta_2)$$

where σ is called the activation function and $\theta_1 \in \mathbb{R}^d, \theta_2 \in \mathbb{R}$. A *one-layer neural network* is the linear combination of the output of N perceptrons, and can be expressed

$$f_N(x|w, \theta) := \sum_{i=1}^{N} \frac{w_i}{N} \sigma(\theta_{1,i} \cdot x + \theta_{2,i})$$

where $w_i \in \mathbb{R}, \theta_{1,i} \in \mathbb{R}^d$, and $\theta_{2,i} \in \mathbb{R}$. The aim is, for N fixed, to find the best parameter to approximate a target function f, in the sense that

$$\min_{(w,\theta)} \int_{\mathbb{R}^d} |f(x) - f_N(x|w, \theta)|^2 \, dx. \tag{3.5}$$

The quadratic error can be replaced by more general functions.

In [118] the authors give a nice presentation of the following *continuous formulation*, i.e., limit as $N \to \infty$. When can let $\xi_i := (w_i, \theta_{1,i}, \theta_{2,i})$, $\Xi(\xi, x) := w\sigma(\theta_{1,i}x + \theta_{2,i})$, and define the empirical distribution

$$\mu_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i}. \tag{3.6}$$

For convenience, define Ω as the set of admissible ξ. Then, the value of the neural network can be computed as

$$f_N(x|w, \theta) = \int_{\Omega} \Xi(\xi, x) \, d\mu_N(\xi).$$

We can rewrite the (3.5) to a minimisation problem over the set of empirical distributions given by (3.6). The limit as $N \to \infty$ is now clear, we pass to the minimisation problem over $\mathcal{P}(\Omega)$. To smoothen the problem the authors allow for a penalisation $\widehat{V}$ (of type $|\xi|^2/2$), which can be understood as trying to prevent *overfitting*. We land on the problem

$$\min_{\mu \in \mathcal{P}(\Omega)} \frac{1}{2} \int_{\mathbb{R}^d} \left| f(x) - \int_{\Omega} \Xi(x, \xi) \, d\mu(\xi) \right|^2 dx + \int_{\Omega} \widehat{V}(\xi) \, d\mu(\xi).$$

Expanding the square, this is precisely (FE*) for $U = 0$ and

$$K(x, y) = \frac{1}{2} \int_{\Omega} \Xi(x, z) \Xi(y, z) \, dz,$$

$$V(x) = - \int_{\Omega} \Xi(x, y) f(y) \, dy + \widehat{V}(x).$$

To find local minimums of this energy, it is natural over $\mathcal{P}(\Omega)$, it is natural to consider the associated 2-Wasserstein gradient flow. As will show below, this leads to (ADE*).

In [94] the authors generalise this result to problems with more than one layer.

3.10 Granular flow equation

Another popular example of the (ADE) family is the so-called granular flow equation. This phenomenom is usually modelled in phase space (t, x, v) where v is the velocity. In that model $\rho(t, x, v)$ represents the phase-space distribution. We arrive at (ADE) if the $\rho(0, x, v)$ does not depend on x. In this setting, U models the random interactions of the granules with their environment (a fluid or heat bath), V models friction, and W models inelastic collisions between granules with different velocities. For more details see [16, 77, 176].

3.11 The power-type family of aggregation-diffusion equations

Many authors have devoted their attention to the power-type family of non-linearities given by $U = U_m$,

$$V_\lambda(x) = \begin{cases} \frac{|x|^\lambda}{\lambda} & \text{if } \lambda \neq 0 \\ \log |x| & \text{if } \lambda = 0, \end{cases} \qquad W_k(x) = \begin{cases} \frac{|x|^k}{k} & \text{if } k \in (-d, d) \setminus \{0\} \\ \log |x| & \text{if } k = 0, \end{cases}$$

Notice that

$$W_N = -C_d W_{2-d} \quad \text{and} \quad W_R = -C_{d,s} W_{2s-d} \quad \text{with } C_d > 0, C_{d,s} > 0$$

This family covers most of the relevant examples above. We introduce the associated free energies

$$\mathcal{U}_m[\rho] = \int_{\mathbb{R}^d} U(\rho), \qquad \mathcal{V}_\lambda[\rho] = \int_{\mathbb{R}^d} V_\lambda \rho,$$

$$\mathcal{W}_k[\rho] = \frac{1}{2} \iint_{\mathbb{R}^d} W_k(x - y) \rho(x) \rho(y) \, dx \, dy. \tag{3.7}$$

4 Well-posedness frameworks

After setting a model like the ones described in Sect. 3, and before any other analysis can take place, the mathematician immediately aims to solve the question of existence and uniqueness. For this, one needs a suitable framework in which to define what is a "solution".

The very first approach to a PDE is to consider the existence of so-called *classical* solutions. These are solutions that are smooth enough so that all the derivatives in the equation can be taken, and the equation is satisfied is *all* points in the domain.

The hope for classical solutions quickly vanishes for many singular problems (e.g., (PME) when $m > 1$). To deal with this difficulty, many authors in 20th century dedicated their efforts to the notion of weak solutions, with their derivatives taken in *distributional sense*. So the equation could simply be satisfied in *almost all points*, and additional reasonable conditions. We will make some comments on this in Sect. 4.1.

We devote Sect. 4.2 to present some good and bad expected properties in different setting. In particular, we will also make a brief stop in Sect. 4.1.6 to discuss stationary solutions (i.e., solutions no evolving in time). When not even weak solutions may exist (like in the transport equation), the community turned towards *Optimal Transport* techniques, which we will discuss in Sect. 4.3.

4.1 The PDE approach

First, we discuss the more classical approach to understand "solutions" of (ADE).

4.1.1 Local PDEs

Non-linear diffusion problems of the general type

$$\frac{\partial \rho}{\partial t} = \Delta \Phi(\rho) + \mathrm{div}(\mathbf{E}(t, x, \rho(t, x))) \tag{4.1}$$

were widely studied in the 20-th century. Through the work of many authors, different types of solutions have been studied: classical, weak, entropy, viscosity,...

Let us start by discussing the nicer problem given by non-degenerate parabolic problems. When Φ, E are smooth and we assume Φ is uniformly elliptic, in the sense that there exist constants such that

$$0 < c_1 \le \Phi'(u) \le c_2 < \infty, \tag{4.2}$$

existence, uniqueness, and maximum principle hold from the classical theory. The literature is extensive: in $\mathbb{R}^d$ this issue was solved at the beginning of the twentieth century (see [136]). This book already contains several results for bounded domains (see [136, Chapter V.6 and V.7]). However, the development of regular solutions would leave room for later improvement. A good reference for Dirichlet boundary conditions is [108]. There is a series of works by [1] for Neumann/Robin boundary conditions, although they provide very general ellipticity conditions which written for systems and are not easy to apply to specific case. There are later references which are more concise, e.g., [180].

The theory of weak solution is based on the notion of weak (or distributional) derivative, and hence solutions are found in Sobolev spaces. This is typical of parabolic problems where $\Phi \ne 0$. There are more general theory that allow for well-posedness even when $\Phi = 0$:

entropy solutions (see [47, 132] and the references therein) and viscosity solutions (e.g., [97] and the references therein)

The theory of weak and energy solutions for (PME) is well presented in the book [171]. Notice that for $m > 1$ we have $\Phi'(0) = 0$, and we say that the diffusion is degenerate, and for $m < 1$ we have $\Phi'(0) = \infty$, and we say the diffusion is singular. The first order problems $U = 0$ (or, equivalently, $\Phi = 0$) are purely of the first order and hence may yield discontinuous solutions.

Remark 4.1 Occasionally, Hölder spaces are used in addition to Sobolev. For this, one can use the method of intrinsic scaling (see [108, 169] and the references therein)

Remark 4.2 Purely-diffusive problems like (PME) admit semigroup-type solutions. In the linear theory this is called Hille–Yosida theorem (see, e.g., [33, Chapter 7]) and in the non-linear setting Crandall–Liggett theorem (see [96]). The semigroup is constructed as the limit $\tau \to 0$ of solutions of the implicit Euler scheme $u_0 = \rho$ and

$$u_{k+1} + \tau(-\Delta u_{k+1}^m) = u_k.$$

To be precise, we take $t_k = k\tau$ and for $t \in [t_k, t_{k+1})$ then we define

$$\rho_t^{(\tau)} = (1 - \tfrac{t-t_k}{\tau})u_k + \tfrac{t-t_k}{\tau}u_{k+1}.$$

Remark 4.3 Boundary conditions of homogeneous Dirichlet (i.e., $u = 0$) and Neumann (i.e., $\nabla u \cdot n = 0$) are fairly well understood in terms of existence and uniqueness (see, e.g., [136]). In linear problems, even so-called Robin type boundary conditions $\nabla u \cdot n + \alpha u = 0$. The operator $Au = -\Delta u - \text{div}(u\nabla V)$ with no-flux condition is monotone (i.e., $\langle Au, u \rangle \geq 0$) if $\|\nabla V\|_{L^\infty}$ is small enough, so well-posedness follows from the Hille–Yosida theory. Non-linear diffusion problems are more difficult. Works for (KSE) and related problem usually take advantage of an auxiliary boundary condition for V_t.

When there is no diffusion (i.e., $U, W = 0$) and $V = -|x|^2$ leads to mass trying to exit any ball B_r. If work on a ball B_R and we set no-flux condition $\rho \nabla V \cdot n = 0$ on ∂B_R, the result will be the formation of Dirac deltas on ∂B_R. This can be properly understood by the optimal transport theory below. It points to a subtle trade-of between diffusion and aggregation.

Existence and uniqueness for (ADE*) is only understood with (BC*) if we assume (2.7) (see [57] for $U = U_m$ with $m \in (0, 1)$). The general case when (2.7) does not hold seems to be open. We provide a tricky example with no-diffusion in Remark 4.14.

4.1.2 Weak vs very weak solution

If we work in $\mathbb{R}^d$ we can take $\varphi \in C_c^\infty((0, T) \times \mathbb{R}^d)$, multiply the equation and integrate one to recover

$$-\int_0^\infty \int_{\mathbb{R}^d} \rho_t \frac{\partial \varphi}{\partial t} = -\int_0^\infty \int_{\mathbb{R}^d} \rho \nabla(U'(\rho) + V + W * \rho) \cdot \nabla\varphi$$

Functions ρ that satisfy this equation are so-called *weak solutions* of the problem. Here, $U(\rho)$ must have a gradient in some sense. In $\mathbb{R}^d$ we can integrate by parts again in the diffusion term to recover

$$-\int_0^\infty \int_{\mathbb{R}^d} \rho_t \frac{\partial \varphi}{\partial t} = \int_0^\infty \int_{\mathbb{R}^d} \Phi(\rho)\Delta\varphi - \int_0^\infty \int_{\mathbb{R}^d} \rho \nabla(V + W * \rho) \cdot \nabla\varphi.$$

These are the so-called *very weak* solutions to the problem.

If we work in a bounded domain Ω and consider the no-flux condition we can recover weak solutions. If we want to recover the no-flux condition from this weak formulation, we need to take φ not vanishing in the boundary, for example $C_c^\infty((0, T); C^\infty(\overline{\Omega}))$. To integrate by parts again we would need that $\partial_n \Phi(\rho) = 0$ on the boundary. This is only possible if we assume that V and K do not interfere with the boundary (2.7).

4.1.3 Solutions of (ADE*) by approximation

We will not make any detailed analysis of Sobolev spaces let us make some comments that the equation would have if the solutions where smooth. We define $\Phi'(s) = sU''(s)$ and $\Phi(0) = 0$. To give a hint on the procedure:

1. For uniformly elliptic Φ_ε, e.g.,

$$\Phi'_\varepsilon \sim \varepsilon + (\Phi' \wedge \varepsilon^{-1})$$

   and $E_t(x)$ known, existence is done through the classical theory above.
2. Letting $\varepsilon \to 0$ yields solutions of the degenerate/singular diffusion term. This is the approach of [171].
3. Take V_κ, K_κ smooth approximations of V and K. Then replace $E_t(x)$ by $\rho\nabla(V + \mathcal{K}\rho)$ by a fixed-point argument.
4. Pass to the limit in κ.

The order of steps 2 and 3 can be suitable exchanged depending on convenience. They depend on delicate a priori estimates. For two particular examples for global-in-time existence and uniqueness theory using this approach we point to [66] for $m > 1$, [61] for $m = 1$, and [57] for $m \in (0, 1)$ (the arguments in [171] make $m > 1$ simpler). For (AE) we point the reader to [137] for a vanishing viscosity argument.

4.1.4 Push-forward solutions

When the velocity field v is known, the transport can be written as

$$\frac{\partial \rho}{\partial t} + \text{div}(\rho v_t) = 0.$$

If v is smooth problems, then can be solved by characteristics. This problem can also be recovered trying to find solutions as generalised characteristics $\rho_t(X_t(y)) = A_t\rho_0(y)$. This leads the set of ODEs

$$\begin{cases} \dfrac{\partial X_t}{\partial t} = v_t(X_t) & t > 0, \\ X_0(y) = y. \end{cases} \tag{4.3}$$

If v_t is Lipschitz in x, the field of characteristics is the unique solution of this problem. We do not concern ourselves with A_t for now.

A more elegant approach is to look for solutions of the Lagrangian formulation of the problem.

$$\int_{X_t(A)} \rho_t(x)\, dx = \int_A \rho_0(x)\, dx. \tag{4.4}$$

This is equivalent to

$$\int_{\mathbb{R}^d} \rho_t(x)\varphi(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \rho_0(x)\varphi(X_t(x))\,\mathrm{d}x.$$

For the details we recommend [2, Lecture 16]. This taps into a well-understood theory coming from the *optimal transport* world (we will come back to this in Sect. 4.3). Given a measure $\mu \in \mathcal{P}(\Omega)$ and a measurable map $T : \Omega \to \Omega$ we can define the push-forward of μ by T, which we denote $T_\sharp \mu \in \mathcal{P}(\Omega)$, as

$$T_\sharp \mu(B):=\mu(T^{-1}(B)) \quad \text{for all } B \text{ measurable.}$$

In particular, we can rewrite as (4.4)

$$\rho_t = (X_t)_\sharp \rho_0.$$

There are many properties of ρ_t that be derived directly from this structure and the analysis of the ODEs (4.3).

4.1.5 Solutions by characteristics when $U = 0$

If $U, W = 0$, then we can simply solve the de-coupled PDEs

$$\begin{cases} \dfrac{\partial X_t}{\partial t} = -\nabla V(X_t) & t > 0, \\ X_0(y) = y, \end{cases}$$

and set $\rho_t = (X_t)_\#\rho_0$. When V is of class C^2, then a unique solution holds by the Picard–Lindelöf theorem. Furthermore, coming back to Sect. 3.3 if we have $\nabla V(x) \cdot x \geq 0$ then

$$\frac{\mathrm{d}}{\mathrm{d}t}|X_t|^2 = -\nabla V(X_t) \cdot X_t \leq 0$$

so all the mass is moving towards 0. Since $X_t = 0$ is a solution, when $V \in C^2$ mass will concentrate in infinite time. A posteriori, for (ADE) when $U = 0$ one can always define

$$v_t = \nabla(V + W * \rho)$$

and arrive at a transport problem.

The problem (AE) is more involved. The specific case (CRSE) was done [127], where the author arrive at closed-form formulas for the characteristics. The author of [138] proves global existence of solutions when W is good enough by a fixed-point argument with suitable a priori estimates in Sobolev spaces. This result later applied to construct local-in-time solutions by approximation [18, 19, 22], where a key difficulty is to then decide if there is global existence or finite-time blow-up. Later works use characteristics to study blow-up behaviours (see, e.g., [20]).

4.1.6 Stationary solutions

There is a particular type of solutions which will be of interest for the asymptotics. A *stationary solution* is a solution that does not evolve in time. We recall the equivalent writing (2.10). If $\widehat{\rho}$ is a stationary solution, using $\varphi = \frac{\delta \mathcal{F}}{\delta \rho}[\widehat{\rho}]$ as a test function yields

$$\int_\Omega \widehat{\rho}\left|\nabla\frac{\delta \mathcal{F}}{\delta \rho}[\widehat{\rho}]\right|^2 = 0.$$

Hence, we arrive at the equation for stationary states

$$\widehat{\rho} \, \nabla \frac{\delta \mathcal{F}}{\delta \rho}[\widehat{\rho}] = 0, \qquad \text{a.e. in } \Omega. \tag{ADE-S}$$

This means that in open connected sets where $\hat{\rho} > 0$ then $\frac{\delta \mathcal{F}}{\delta \rho}[\widehat{\rho}]$ is constant.

Usually it is asked also that $\rho \in L^1 \cap L^\infty$ that $\Phi(\rho) \in W_{loc}^{1,1}$ (where $\Phi'(s) = sU'(s)$ and $\Phi(0) = 0$) and that $\nabla V, \nabla W * \rho \in L_{loc}^1$. This regularity is not always achievable.

Remark 4.4 When they are available, these kinds of solutions will typically be recovered from a minimisation problem. We will discuss particular examples in Sect. 6.3.

Remark 4.5 Notice that solving this problem when $U = U_m$ and $V = V_2$ the Barenblatt profile (3.2) (i.e., (ZKB) in re-scaled variables) is a solution of (ADE-S).

4.2 A priori estimates for smooth solutions

In this section we make some formal computations to understand properties of the solution of (ADE*). We will mostly assume that ρ is smooth. However, due to weak lower semi-continuity of norms, the a priori estimates still usually hold after approximation.

4.2.1 Conservation of mass

Let us consider (ADE*), and take a smooth set $\omega \subset \mathbb{R}^d$. We can formally compute the balance of mass

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\omega \rho \, \mathrm{d}x = \int_\omega \frac{\partial \rho}{\partial t} = \int_\omega \mathrm{div} \left(\rho \nabla (U'(\rho) + V + \mathcal{K}\rho) \right) \mathrm{d}x$$
$$= - \int_{\partial \omega} \rho \nabla (U'(\rho) + V + \mathcal{K}\rho) \cdot n \, \mathrm{d}S.$$

Hence, if we have the boundary conditions (BC) we have conservation of mass. In $\mathbb{R}^d$ we will need that the decay as $|x| \to \infty$ is fast enough. This is not always the case, as shown by the next result (see [17]).

Proposition 4.6 Let ρ be a solution of (PME) in $\mathbb{R}^d$ with $m \in (0, \frac{d-2}{d})$ and $d \geq 3$. Then

1. If $\rho_0 \in L^q(\mathbb{R}^d)$ with $q = \frac{(1-m)d}{2}$ then there is finite-time extinction.
2. If $\rho_0 \in L^1(\mathbb{R}^d)$ then there is (at least), infinite-time extinction.

Proof. First let $u_0 \in L^q$. Then, we compute for sufficiently good solutions that

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{q} \int_{\mathbb{R}^d} \rho^q = -C \int_{\mathbb{R}^d} |\nabla \rho^{\frac{m+q}{2}}|^2 \leq -C \left(\int_{\mathbb{R}^d} \rho^{\frac{m+q}{2} 2^*} \right)^{\frac{1}{2^*}} \tag{4.5}$$

using Sobolev's inequality where $2^* = \frac{1}{2} - \frac{1}{d}$. The equation $\frac{\mathrm{d}}{\mathrm{d}t} X = -C X^\alpha$ where $\alpha < 1$ has finite-time extinction in the sense $X_T = 0$. Thus, we have the finite-time extinction

$$\int_{\mathbb{R}^d} \rho^q \to 0, \qquad t \nearrow T.$$

Let $\rho_0 \in L^1(\mathbb{R}^d)$. Take $\varepsilon > 0$ and an approximating sequence $\rho_0^{(\varepsilon)} \in L^q$ with $\|\rho_0 - \rho_0^{(\varepsilon)}\|_{L^1} \le \varepsilon$. For $t \ge T_\varepsilon^*$

$$
\begin{aligned}
\|\rho(t)\|_{L^1} &\le \|\rho(t) - \rho^{(\varepsilon)}(t)\|_{L^1} + \|\rho^{(\varepsilon)}(t)\|_{L^1} \\
&\le \|\rho_0 - \rho_0^{(\varepsilon)}\|_{L^1} + 0 \\
&\le \varepsilon. \qquad\qquad\qquad\qquad\qquad\qquad\qquad \square
\end{aligned}
$$

This effect is known as "loss of mass through infinity", and is typical to the *very fast diffusion*. In some settings, this can be offset by the aggregation term. To prevent this loss of mass it suffices that ρ_t is a *tight* family (a notion we discuss in the next section).

4.2.2 Mass escaping through infinity: tightness

In the previous section we have discussed the problem of mass escaping through infinity "as time passes". This phenomenon is also related to a difficult technical problem when working on $\mathcal{P}(\mathbb{R}^d)$. We recall some classical results for working with probability measures.

First, let ρ_n be a sequence of functions in $\mathcal{P}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. If they have an a.e. limit we know, at most, that

$$
\int_{\mathbb{R}^d} \hat{\rho} \le \liminf_n \int_{\mathbb{R}^d} \rho_n = 1.
$$

But the equality needs to hold. Furthermore, let us take a sequence in $\mu_n \in \mathcal{P}(\mathbb{R}^d)$. Since they are uniformly bounded measures, and $\mathcal{M}(\mathbb{R}^d)$ is the dual of $C_b(\mathbb{R}^d)$ by the Banach-Alaoglu theorem we have that the sequence μ_n is weak-$\star$ pre-compact, and take $\widehat{\mu}$ an accumulation point. It needs not lie in $\mathcal{P}(\mathbb{R}^d)$.

To ensure that the limit is a measure we need to add some information. Prokhorov's theorem (see, e.g., [3, Theorem 5.1.3]) states that a family of measures is pre-compact in $\mathcal{P}(\mathbb{R}^d)$ if and only it is *tight*.

Definition 4.7 A family of probability measures $(\mu_a)_{a \in A}$ is tight if, for every $\varepsilon > 0$ there exists $K_\varepsilon \subset \mathbb{R}^d$ compact such that

$$
\mu_a(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon, \quad \forall a \in A.
$$

This means, informally speaking, that the *tails* of μ_n hold "uniformly small" mass.

Remark 4.8 To show that a family of probability measures $(\mu_a)_{a \in A}$ is tight it suffices to show that it is uniformly integrable against a suitable function. Following [3, Remark 5.1.5], it suffices to find $\varphi : \mathbb{R}^d \to [0, \infty]$ such that its sublevel sets (i.e., $\{x \in \mathbb{R}^d : \varphi(x) \le c\}$ for $c \in \mathbb{R}$) are compact, and we have

$$
\sup_{a \in A} \int_{\mathbb{R}^d} \varphi(x) \, \mathrm{d}\mu_a(x) < \infty. \tag{4.6}
$$

Very typical examples is $\varphi(x) = |x|^p$ with $p \ge 1$. The quantity $\int_{\mathbb{R}^d} |x|^p \, \mathrm{d}\mu$ is called the pth order moment of μ.

It is clear that (4.6) can follow for minimisers (FE) if V is good enough. We point out the following estimate that allows to use $\mathcal{W}$ for the same purpose.

Fig. 1 Sets used of the proof of Lemma 4.9
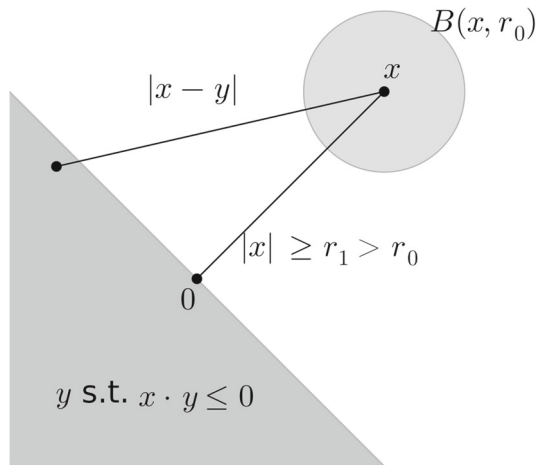
Lemma 4.9 [84, 145] *Let ρ be radially symmetric, $\Omega = \mathbb{R}^d$, w non-decreasing such that $w(r_0) \geq 0$ for some r_0, and let $r_1 > r_0$. Then*

$$\int_{|x| \geq r_1} w(|x|)\rho(x)dx \leq \frac{\iint_{\mathbb{R}^d \times \mathbb{R}^d} |w(|x-y|)|\rho(x)\rho(y)\,\mathrm{d}x\,\mathrm{d}y}{\frac{1}{2}\int_{\mathbb{R}^d} \rho(x)\,\mathrm{d}x}$$

Proof. Taking into account the geometry shown in Fig. 1 we can estimate

$$\int_{|x| \geq r_1} \int_{|x-y| \geq r_0} w(|x-y|)\rho(x)\rho(y)\,\mathrm{d}y\,\mathrm{d}x$$

$$\geq \int_{|x| \geq r_1} \int_{y \cdot x \leq 0} w(|x-y|)\rho(x)\rho(y)\,\mathrm{d}y\,\mathrm{d}x$$

Since $x \cdot y \leq 0$ then $|x-y|^2 = |x|^2 + |y|^2 - 2x \cdot y \geq |x|^2$ and since w is non-decreasing

$$\geq \int_{|x| \geq r_1} \int_{y \cdot x \leq 0} w(|x|)\rho(x)\rho(y)\,\mathrm{d}y\,\mathrm{d}x$$

$$\geq \int_{|x| \geq r_1} w(|x|)\rho(x) \int_{y \cdot x \leq 0} \rho(y)\,\mathrm{d}y\,\mathrm{d}x.$$

Since ρ is radially symmetric, for any $x \in \mathbb{R}^d$ we have

$$\int_{y \cdot x \leq 0} \rho(y)\,\mathrm{d}y = \frac{1}{2}\int_{\mathbb{R}^d} \rho(y)\,\mathrm{d}y.$$

$\square$

4.2.3 An L^p estimate

We can proceed similarly to the approach in (4.5) in more general settings. If we drop the term coming from the diffusion, then we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{p}\int_\omega \rho^p = \int_\omega \rho^{p-1}\frac{\partial\rho}{\partial t}$$

$$= \int_\omega \rho^{p-1} \operatorname{div} \rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}$$

$$= -(p-1) \int_\omega \rho^{p-1} \nabla \rho \nabla (U'(\rho) + V + \mathcal{K}\rho) + \int_{\partial \omega} \rho^p \nabla \frac{\delta \mathcal{F}}{\delta \rho} \cdot n$$

$$\leq -\frac{p-1}{p} \int_\omega \nabla \rho_t^p \nabla (V + \mathcal{K}\rho) + \int_{\partial \omega} \rho^p \nabla \frac{\delta \mathcal{F}}{\delta \rho} \cdot n$$

Integrating by parts one final time leads to

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{p} \int_\omega \rho^p \leq \frac{p-1}{p} \int_\omega \rho^p \Delta (V + \mathcal{K}\rho)$$
$$- \frac{p-1}{p} \int_{\partial \omega} \rho^p \nabla (V + \mathcal{K}\rho) \cdot n(x) + \int_{\partial \omega} \rho^p \nabla \frac{\delta \mathcal{F}}{\delta \rho} \cdot n. \qquad (4.7)$$

If we work in $\mathbb{R}^d$ with $V, W \in W^{2,\infty}$, we assume good decay of $\rho^p \nabla (V + \mathcal{K}\rho)$ and $\rho^p \nabla \frac{\delta \mathcal{F}}{\delta \rho}$ then the boundary term vanish and we can write an inequality for equation of the form

$$\|\rho_t\|_{L^p(\mathbb{R}^d)} \leq \|\rho_0\|_{L^p(\mathbb{R}^d)} e^{Ct}.$$

However, when we work in Ω with (BC*) and we take $\omega = \Omega$, then the last term vanishes. However, we cannot relate the boundary term $\rho^p \nabla (V + \mathcal{K}\rho) \cdot n(x)$ to the L^p norm of ρ. In some cases it can be compensated by the negative diffusion term, but not in general. This is one of the reasons we assume (2.7) on the boundary.

4.2.4 Free-energy dissipation

Definition 4.10 (*First variation*) Let X be a normed space, $\mathcal{F} : X \to \mathbb{R}$ and fix $\rho_0 \in X$. We say $\mathcal{F}$ it is Gateaux differentiable at X if for all $\varphi \in X$ there exists

$$\mathrm{d}\mathcal{F}[\rho_0, \varphi] := \lim_{\varepsilon \to 0} \frac{\mathcal{F}[\rho_0 + \varepsilon \varphi] - \mathcal{F}[\rho_0]}{\varepsilon}.$$

The function $\mathrm{d}\mathcal{F}$ is usually called *first variation of* $\mathcal{F}$. Some authors extend this definition and require just that the limit exists for φ in a dense subset of X. Functionals that admit first variation are called *Gateaux differentiable*.

When $X \subset L^1(\Omega)$, the first variation can sometimes be represented by a function $f_0 \in L^1(\Omega)$ the sense that

$$\int_\Omega f_0(x)\varphi(x) \, \mathrm{d}x = \mathrm{d}\mathcal{F}[\rho_0, \varphi], \qquad \forall \varphi \in L^\infty(\Omega).$$

If this is the case, we denote $\frac{\delta F}{\delta \rho}[\rho_0] := f_0$. It is easy to check formally that, when $\mathcal{F}$ is given by (FE*) and $K(x, y) = K(y, x)$, then (2.9). In many of the examples the first variation can be rigorously computed.

If $K(x, y) = K(y, x)$ then the solutions of (ADE*) admits (FE*) as Lyapunov functional. This functional is usually called *free energy*. For smooth solutions of (ADE*) (with sufficiently good decay as $|x| \to \infty$ in unbounded domains) we can compute

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}[\rho] = -\int_\Omega \rho \left| \nabla \frac{\delta \mathcal{F}}{\delta \rho} \right|^2 + \int_{\partial \Omega} \rho \nabla \frac{\delta \mathcal{F}}{\delta \rho} \cdot n. \qquad (4.8)$$

When $\Omega = \mathbb{R}^d$ (with sufficiently good decay of $\rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}[\rho_t]$) or we have the no-flux condition (BC*), we recover the decay of the free energy, i.e., along solutions of (ADE*) we have decay of free-energy dissipation

$$\mathcal{F}[\rho_t] - \mathcal{F}[\rho_s] = -\int_s^t \int_\Omega \rho \left| \nabla \frac{\delta \mathcal{F}}{\delta \rho}[\rho] \right|^2. \tag{4.9}$$

In particular if $s < t$ then $\mathcal{F}[\rho_s] > \mathcal{F}[\rho_t]$.

Remark 4.11 Often the free energy dissipation allows us to prove tightness. One approach is to recall Remark 4.8 and notice that free energy dissipation includes a term $\int_\Omega V \rho$. The term $\iint W(x - y)\rho(x)\rho(y)$ can also be useful due to Lemma 4.9.

4.2.5 Negative Sobolev spaces

Given a set Ω, the negative Sobolev spaces can be defined by distributions with finite norm

$$\|f\|_{W^{-1,1}(\Omega)} = \inf_{f = \mathrm{div}\, F} \|F\|_{L^1(\Omega)}$$

Using the free-energy dissipation

$$\left\| \frac{\partial \rho_t}{\partial t} \right\|_{W^{-1,1}(\Omega)} = \inf_{\frac{\partial \rho_t}{\partial t} = \mathrm{div}\, F} \|F\|_{L^1(\Omega)} \le \left\| \rho \nabla \frac{\delta \mathcal{F}}{\delta \rho} \right\|_{L^1}$$

$$\le \|\rho_t\|_{L^1}^{\frac{1}{2}} \left(\int_\Omega \rho_t \left| \nabla \frac{\delta \mathcal{F}}{\delta \rho} \right|^2 \right)^{\frac{1}{2}} = \|\rho_t\|_{L^1}^{\frac{1}{2}} \left(-\frac{d}{dt} \mathcal{F}[\rho_t] \right)^{\frac{1}{2}}$$

Also $\|\rho_t\|_{L^1} \le \|\rho_0\|_{L^1}$. In the smooth setting, we can integrate in time

$$\|\rho_t - \rho_s\|_{W^{-1,1}} \le C|t - s|^{\frac{1}{2}} (\mathcal{F}[\rho_s] - \mathcal{F}[\rho_t])^{\frac{1}{2}}.$$

Since L^1 is compactly embedded in $W^{-1,1}$, when $\mathcal{F}$ is bounded below, the sequence

$$\rho_t^{[n]} = \rho_{n+t}$$

is relatively compact in $C([0,1]; W^{-1,1}(B_R))$ due to Ascoli–Arzelá theorem. Let $\widehat{\rho} \in W^{-1,1}$ be the limit of a sub-sequence. Since $\mathcal{F}[\rho_t]$ is bounded below and non-increasing, there is a limit $\widehat{\mathcal{F}}$. Then

$$\|\widehat{\rho_t} - \widehat{\rho_s}\|_{W^{-1,1}} \le C|t - s|^{\frac{1}{2}} (\widehat{\mathcal{F}} - \widehat{\mathcal{F}}) = 0.$$

So $\widehat{\mu}$ does not depend on t. In the cases where one can prove stronger convergences, it can be shown that it is a stationary solution, i.e.,

$$\mathrm{div} \left(\widehat{\rho} \nabla \frac{\delta \mathcal{F}}{\delta \rho}[\widehat{\rho}] \right) = 0. \tag{4.10}$$

4.3 Optimal-transport approach: Wasserstein spaces and the JKO scheme

As we already mentioned in Sect. 4.1.4, there is a clear connection between some of our equations and *Optimal Transport* problems. Otto realised in [147] (see also [148]) that the (PME) in $\mathbb{R}^d$ corresponds, in fact, to the gradient flow of a $\mathcal{F}$ with respect to the 2-Wasserstein distance. This idea led to the so-called "Otto calculus" that was later perfected.

The following pages are meant as an extremely brief introduction. For the reader interested in deepening their understanding of these spaces and their connection to PDEs we suggest reading the lecture notes [2] first, and then the very detailed books [3, 177]. A very nice presentation with an emphasis on the examples can be found in [153].

4.3.1 The Wasserstein distances

The Wasserstein metrics were a tool already used by people studying optimal transport between probability measures. We give a brief definition. Let X be a metric space with a Borel algebra, and recall the definition of push-forward given in Sect. 4.1.4. If we are given $\mu, \nu \in \mathcal{P}(X)$, a transport map is a measurable function T such that $\nu = T_\sharp \mu$. We can formally think about the p-Wasserstein distance between probability distributions are defined for $p \geq 1$ as

$$\left(\inf_{T:\nu=T_\sharp\mu} \int_X |x - T(x)|^p \, \mathrm{d}\mu(x) \right)^{\frac{1}{p}} \tag{4.11}$$

This is the so-called Monge optimisation problem. The infimum is not always achieved, and sometimes no valid T exists. Kantorovich improved this idea by introducing transport plans. A transport plan is a probability distribution in $\mathcal{P}(X \times X)$ that have μ, ν, i.e.,

$$\mu(A) = \pi(A \times X), \quad \nu(A) = \pi(X \times A), \quad \text{for all } A \subset X \text{ measurable.}$$

This set is denoted by $\Pi(\mu, \nu)$, and is never empty because of the measure $\mu \otimes \nu \in \Pi(\mu, \nu)$, the unique measure such that

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B).$$

The *Wasserstein distances* are defined as

$$\mathfrak{W}_p(\mu, \nu) := \left(\min_{\pi \in \Pi(\mu,\nu)} \iint_{X \times X} |x - y|^p \, \mathrm{d}\pi(x, y) \right)^{\frac{1}{p}}. \tag{4.12}$$

These distances are sometimes called also *Kantorovich–Rubinstein distance*. The distance between two measures in $\mathcal{P}(X)$ can be infinite, unless the finite pth order moment is finite. Hence, we define the p-Wasserstein space

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) : \int_X |x|^p \, \mathrm{d}\mu(x) < \infty \right\}.$$

The pair $(\mathcal{P}_p(X), \mathfrak{W}_p)$ is a metric space. If X is complete, then so this is space is also complete and it is equivalent to the narrow convergence (see [3, Proposition 7.1.5]). This is why we pick $X = \overline{\Omega}$.

As it usually happens in these families of spaces, there are three highlighted cases: $p = 1, 2, \infty$. The case $p = 2$ will appear in the next section, and it is the one directly related to (ADE). The advantage of $p = 1$ is the so-called Kantorovich–Rubinstein duality

$$\mathfrak{W}_1(\mu, \nu) = \sup \left\{ \int \psi \, \mathrm{d}\mu - \int \psi \, \mathrm{d}\nu : \text{for } \psi \text{ such that } \sup_{x \neq y} \frac{|\psi(x) - \psi(y)|}{|x - y|} \leq 1 \right\}. \tag{4.13}$$

A very interesting property of the Wasserstein distance is that it admits a dynamic characterisation through the so-called *Benamou–Brenier formula* which is stated easiest in $\mathbb{R}^d$

$$\mathfrak{W}_2(\mu_0, \mu_1) = \inf_{(\mu,v)\in\mathcal{A}(\mu_0,\mu_1)} \int_0^1 \int_{\mathbb{R}^d} |v_t|^2 \, d\mu_t \, dt, \tag{4.14}$$

where

$$\mathcal{A}(\mu_0, \mu_1) = \Big\{ (\mu, v) : \mu \in C([0,1], \mathcal{P}_2(\mathbb{R}^d)), \, v \in L^2([0,1] \times \mathbb{R}^d, \mu)^d,$$

$$\mu(0) = \mu_0, \mu(1) = \mu_1, \text{ and } \frac{\partial \mu_t}{\partial t} + \operatorname{div}(\mu_t v_t) = 0 \Big\}$$

is set space of admissible paths between μ_0 and μ_1.

4.3.2 Otto's calculus

The notion of calculus in these spaces is a little tricky, since $\mathcal{P}_p(\mathbb{R}^d)$ are not vector spaces. However, they are subsets of the linear space of measures. In fact, it can be though of as a manifold: we can define tangent through curves. In particular, $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and take $\varphi \in C_c^\infty(\mathbb{R}^d)$ a test function. Consider the curve of probabilities

$$\rho_t = (\mathrm{id} + t\nabla\varphi)_\sharp \rho_0.$$

It is not too difficult to see [2, Lecture 16] that ρ_t is a distributional solution of the continuity equation

$$\frac{\partial \rho_t}{\partial t} + \operatorname{div}(\rho_t v_t) = 0$$

where $v_t = \nabla\varphi \circ (\mathrm{id} + t\nabla\varphi)^{-1}$. Then formally the tangent space $T_{\rho_0}\mathcal{P}_2(\mathbb{R}^d)$ is made of

$$s := \frac{d}{dt}\Big|_{t=0} \rho_t = -\operatorname{div}(\rho_t \nabla\varphi).$$

This allows to formally construct the gradient of functions. Take the free energy of the diffusion term $\mathcal{U}$. In distributional sense the Wasserstein gradient is

$$\langle \nabla_{\mathfrak{W}_2} U[\rho_0], s \rangle := \frac{d}{dt}\Big|_{t=0} \mathcal{U}[\rho_t] = \int_{\mathbb{R}^d} U'(\rho_0) \frac{\partial}{\partial t}\Big|_{t=0} \rho_t$$

$$= -\int_{\mathbb{R}^d} U'(\rho_0) \operatorname{div}(\rho_0 \nabla\varphi)$$

$$= -\int_{\mathbb{R}^d} \varphi \operatorname{div}(\rho_0 \nabla U'(\rho_0)).$$

In general

$$\langle \nabla_{\mathfrak{W}_2} \mathcal{F}[\rho_0], s \rangle = -\int_\Omega \varphi \operatorname{div}\left(\rho \nabla \frac{\delta\mathcal{F}}{\delta\rho}[\rho_0] \right) \tag{4.15}$$

Hence, (ADE) in $\mathbb{R}^d$ is formally written can be formally written as the 2-Wasserstein gradient flow in the sense that

$$\frac{\partial\rho}{\partial t} = -\nabla_{\mathfrak{W}_2} \mathcal{F}[\rho_t]$$

of the free energy given by (FE).

Remark 4.12 If we work in a bounded domain, then we must study the space $\mathcal{P}_2(\overline{\Omega})$ and understand how the no-flux condition (BC) appears naturally in this context. One approach is to consider $\varphi \in C_c^\infty(\overline{\Omega})$ (meaning their support is bounded and does not reach the boundary). In order for $\mathrm{id} + t\nabla\varphi : \overline{\Omega} \to \overline{\Omega}$ when $t \in (-\varepsilon, \varepsilon)$ we must specify $\rho\nabla\varphi \cdot n = 0$. This is related to how the no-flux appears. We will continue this discussion below in Remark 4.13.

4.3.3 Rigorous gradient flow structure. The JKO scheme

The idea of writing gradient flows in terms of minimising movements is due to De Giorgi (see [101]). To draw a quick analogy think that we can to minimise a function $F : \mathbb{R}^d \to \mathbb{R}$. A continuous ODE going to local minimima is the $\mathbb{R}^d$ gradient flow

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\nabla F(x)$$

The implicit Euler scheme leads to the implicit gradient-descent method

$$\frac{x_{n+1} - x_n}{\tau} = -\nabla F(x_{n+1}).$$

We have avoided making explicit the dependence of x_n with respect to τ. We can rewrite this the minimisation of the so-called proximal function

$$x_{n+1} = \operatorname*{argmin}_{x \in \mathbb{R}^d} \left(\frac{|x - x_n|^2}{2} + \tau F(x) \right) \tag{4.16}$$

If F is smooth this minimisation problem has a unique solution for τ small. These are the so-called *minimising movements*. Then one can do a piece-wise constant interpolation $x^{(\tau)}(t) := x_n$ when $t \in [\tau n, \tau(n+1))$, or even linear interpolation.

The notion of minimising movement can be generalised to metric spaces and, in particular, we can construct in 2-Wasserstein space

$$\mu_{n+1} = \operatorname*{argmin}_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left(\frac{\mathfrak{W}_2(\mu, \mu_n)^2}{2} + \tau\mathcal{F}(\mu) \right). \tag{4.17}$$

Jordan, Kinderlehrer, and Otto proved in the seminal paper [130] that this procedure works for (HE-FP). The book [3] is devoted to proving how these minimising movements lead to (ADE) in $\mathbb{R}^d$ for fairly general $\mathcal{F}$. A good notion of solution for the limit of minimising movements is that of *curves of maximal slope*. Often, it is even possible to get back a suitable solution of a PDE.

Remark 4.13 In bounded domains, we continue the comment in Remark 4.12. In general, if one minimises over $\mathcal{P}_2(\overline{\Omega})$, then we expect to arrive (ADE) with the no-flux condition (BC*). We recommend [153, Chapter 8].

However, the general setting is tricky as presented by the following example.

Remark 4.14 Take $\Omega = (-1, 1)$ (i.e., $d = 1$), $U = W = 0$ and $V(x) = ax$. Then we recover the transport equation $\frac{\partial\rho}{\partial t} = a\frac{\partial\rho}{\partial x}$. The solutions of this equation are always of the form $\rho(t, x) = \rho_0(x + at)$. The boundary condition $a\rho = 0$ can be forced on one side, but never the other. From the perspective below one can always consider a Dirac delta on the right-hand side. This negates somehow $a\rho = 0$ on that side, but it is only possible solution. In fact, it is recovered from the vanishing viscosity formulation.

Notice that this examples does not satisfy (2.7). If we replaced V by a different smooth potential such that (2.7), then the characteristics will not reach the boundary in finite time. Hence, we will expect a smooth solution.

Remark 4.15 Otto's original paper [147] already covered p-Wassersten distances with $p \neq 2$. In particular, it showed that the *doubly nonlinear equation*

$$\frac{\partial u}{\partial t} = \Delta_p u^m$$

has a p-Wasserstein gradient-flow structure. Recently, a new article has appeared justifying the JKO scheme for these problems, see [42].

4.3.4 Convexity and McCann's condition

In the same way that (4.16) works better if $F : \mathbb{R}^d \to \mathbb{R}$ is convex, there is a suitable notion of convexity that works for (4.17). The correct extension is convexity along geodesics also called *displacement convexity*, i.e., if ρ_t is a geodesic from ρ_0 to ρ_1 then $t \mapsto \mathcal{F}[\rho_t]$ is a convex $[0, 1] \to \mathbb{R}$ function. It is usually called *displacement convexity* and was introduced by McCann in [145].[1]

Before we present this result, it is interesting to discuss the structure of geodesics in Wasserstein space (see, e.g., [2]). As the simplest case, assume that μ, ν are absolutely continuous probability distributions and compactly supported. Then, Monge's problem (4.11) for $p = 2$ admits an optimal transport map $T = \nabla \varphi$ such that $\nu = T_\sharp \mu$ (see, e.g., [2, Lecture 5]). Furthermore, the unique geodesic between μ and ν is given by

$$\rho_t = ((1 - t)\mathrm{id} + tT)_{\#}\mu.$$

In [145] the author proves that if V, W are strictly convex, $W(x) = W(-x)$ then

$$\int_{\mathbb{R}^d} V\rho, \quad \iint_{\mathbb{R}^d \times \mathbb{R}^d} W(x - y)\rho(x)\rho(y) \,\mathrm{d}x \,\mathrm{d}y$$

are displacement convex. For the diffusion term there is a more interesting condition, known as *McCann's condition*. It states that if we define

$$P'(s) := sU'(s), \quad P(0) := 0$$

then (2.11) is displacement convex if and only if

$$P'(s)s \geq (1 - 1/d)P(s) \geq 0, \quad \forall s \in (0, \infty). \tag{4.18}$$

This can also be written as $\lambda \mapsto \lambda^d U(\lambda^{-d})$ is convex. When $U = \frac{s^m}{m-1}$ this means $m \geq \frac{d-1}{d}$.

Remark 4.16 The case $m > \frac{d-1}{d}$ and $V(x) = \frac{|x|^2}{2}$ the functional is 0-convex and bounded below. But recalling Remark 3.2 for $m < \frac{d}{2+d}$ the (3.2) is not in $\mathcal{P}_2$, we cannot hope for asymptotic convergence in 2-Wasserstein sense.

Besides "regular" convexity, a powerful tool in the arsenal is the notion of λ-convexity. A function $f : (0, 1) \to \mathbb{R}$ is λ-convex if $f(t) - \lambda \frac{|t|^2}{2}$ is convex. Analogously as above

---

[1] It is interesting to mention that this work only mentions Wasserstein spaces in the title of one of the references, and rather speaks of a "a novel but natural interpolation between Borel probability measures"

we introduce the notion of displacement (or geodesically) λ-convex. In this setting for the gradient flow starting from two points we have

$$\mathfrak{W}_2(\mu_t, \widehat{\mu}_t) \leq e^{-\lambda t} \mathfrak{W}_2(\mu_0, \widehat{\mu}_0).$$

This implies uniqueness of the gradient flow and, if $\mathcal{F}$ is bounded below, of its minimisers.

A similar analysis of a free energy in $\mathbb{S}^1$ is done in [82]. The sense in which the equation is satisfied is delicate. For $U = V = 0$ and W "pointy", we send the reader to [56].

4.4 Global existence vs finite-time blow-up

As explained above, local-in-time existence is usually achieved through "standard" PDE methods or the use of gradient flow arguments, specially if starting from a nice initial datum ρ_0. Whether these solutions behave "nicely" for large t is a more difficult problem.

Transport for given potential Going back to the simplest example (CE) we can look at the examples $V = \frac{|x|^\alpha}{\alpha}$ with $\alpha > 0$. Then $v = -\nabla V = -|x|^{\alpha-2}x$ and $\rho_t = (X_t)_\sharp \rho_0$ with corresponding characteristic field is obtained by solving

$$\frac{\partial X_t}{\partial t} = -|X_t|^{\alpha-2} X_t,$$

with $X_t(0) = y$. We conclude that

$$X_t(y) = \begin{cases} ye^{-t} & \text{if } \alpha = 2, \\ (|y|^{2-\alpha} - (2-\alpha)t)^{\frac{1}{2-\alpha}} \frac{y}{|y|} & \text{if } \alpha \neq 2. \end{cases}$$

For $\alpha < 2$ the characteristics arrive at 0 at finite time and hence we get a Dirac delta at 0 at a certain finite time for any $\rho_0 \not\equiv 0$. This is not problematic in the distributional sense in this case. However, this behaviour happens also for other problems.

Keller–Segel problem. Some authors noticed by direct techniques that the Keller–Segel model has finite-time blow-up for initial data with large enough mass. [129] showed for (KSE) in a bounded domain with no-flux the existence of a critical mass M^* such that, if $\|\rho_0\|_{L^1} > M^*$, then $\rho(T^*)$ contains a Dirac delta. Later this result was obtained also in $\mathbb{R}^d$ in [126] using radial arguments, where the authors characterise the critical mass is $M^* = 8\pi$. This was later proved in more generality in [112]. Conversely, for $\|\rho_0\|_{L^1} < M^*$ global existence for (KSE) is known. In $d = 2$ it was shown in [28].

A nice proof of the formation of a Dirac comes through the analysis of the second-order moment. For (KSE) in $d = 2$ it was first noticed by [112] that, working with solutions of initial mass $\int_{\mathbb{R}^d} \rho_0 = M$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} |x|^2 \rho(t, x)\, \mathrm{d}x = 4M\left(1 - \frac{M}{8\pi}\right).$$

Hence, if $M > 8\pi$ necessarily the second-order moment becomes negative in finite time. This is incompatible with our non-negative solutions. There is complete concentration to a Dirac delta. This was later extended by [27] for (KSE) when $d > 2$ where the authors prove that

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} |x|^2 \rho(t, x)\, \mathrm{d}x = 2(d-2)\mathcal{F}[\rho(t, \cdot)] \leq 2(d-2)\mathcal{F}[\rho_0].$$

When $M > M^*$, there exist $\|\rho_0\|_{L^1} = M$ such that $\mathcal{F}[\rho_0] < 0$. We will perform the analysis of the free energy for this problem in Remark 6.15. We will discuss below in Sect. 5.1.3 the

critical case $M = 8\pi$ and some interesting *ad-hoc* constructions which are available in the literature.

5 Asymptotics

The aim of this section is to highlight some techniques coming from the PDE community that allow us to understand the behaviour as $t \to \infty$ (and sometimes $|x| \to \infty$) of "typical" solutions of the problem (namely those with "good" initial values).

5.1 Self-similarity

Some of the equations we are analysing have linear operators *homogeneous non-linearities*, and it is therefore interesting to exploit this structure. One of the main approaches comes from considering the mass-preserving change of variable

$$u(\tau, y) = \sigma(\tau)^d \rho\left(\tilde{t}(\tau), \sigma(\tau)y\right). \tag{5.1}$$

5.1.1 Self-similar scaling for (PME)

Let ρ be the solution of (PME). Applying the chain rule to (5.1)

$$\frac{\partial u}{\partial \tau} = \frac{\frac{d\sigma}{d\tau}}{\sigma} u \, \mathrm{div}_y \, y + \sigma^d \frac{d\sigma}{d\tau} \nabla_x \rho \cdot \nabla_y y + \sigma^d \frac{d\tilde{t}}{d\tau} \Delta_x \rho^m$$

since $\nabla_x x = (1, \ldots, 1) = \nabla_y y$ y $\mathrm{div}_x \, x = d = \mathrm{div}_y \, y$. Due to the chain rule

$$\frac{\partial u}{\partial y_i} = \sigma^{d+1} \frac{\partial \rho}{\partial x_i}, \quad \frac{\partial^2 u^m}{\partial y_i^2} = \sigma^{2+md} \frac{\partial^2 \rho^m}{\partial x_i^2}$$

Eventually, we simplify the equation for u to

$$\frac{\partial u}{\partial \tau} = \frac{\frac{d\sigma}{d\tau}}{\sigma} \mathrm{div}_y (uy) + \sigma^{-d(m-1)+2} \frac{d\tilde{t}}{d\tau} \Delta_x \rho^m$$

The only sensible choice is $\frac{d\sigma}{d\tau} = \sigma$ and $\frac{d\tilde{t}}{d\tau} = \sigma^{\frac{1}{\kappa}}$ where $\kappa = (d(m-1) + 2)^{-1}$, so we arrive to the simplified equation

$$\frac{\partial u}{\partial \tau} = \mathrm{div}_y (uy) + \Delta_y u^m.$$

Forward self-similar solutions Notice that adding the condition $\sigma(0) = 1$ we get $\sigma(\tau) = e^\tau$. We also have that $\frac{d\tilde{t}}{d\tau} = e^{\tau/\kappa}$. We have $t = \tilde{t}(\tau)$ so we set $\tilde{t}(0) = 0$. Eventually

$$t = \frac{e^{\frac{\tau}{\kappa}} - 1}{\kappa}$$

$$\kappa = (d(m-1) + 2)^{-1}$$

$$x = e^\tau y.$$

Notice that $\kappa > 0$ if and only if $m > m_c = \frac{d-2}{d}$. The Barenblatt solution (ZKB) corresponds to setting u as stationary. In fact, for $m > m_c$ we can write the family of self-similar solutions

$$\rho_t(x) := \frac{1}{\sigma_B(t+T)^d} F\left(\frac{x}{\sigma_B(t+T)}\right),$$

$$F_B(y) := \left(D - \frac{m-1}{2m}|y|^2\right)_+^{\frac{1}{m-1}}$$

$$\sigma_B(t) := (d|m - m_c|t)^{\frac{1}{d(m-m_c)}}.$$

Here we have chosen the presentation in [26, 83].

Backwards self-similar solutions When $\kappa < 0$ (i.e., $m > m_c = \frac{d-2}{d}$), we can build another type of solution. In [170, Section 5.2], Vázquez introduces the pseudo-Barenblatt solutions for the sub-critical regime $0 < m < m_c$ which can be written

$$\rho_t(x) := \frac{1}{\sigma_B(T-t)^d} F_B\left(\frac{x}{\sigma_B(T-t)}\right).$$

Here, T denotes the extinction time. They are suitable to deal with finite-time extinction [26].

5.1.2 Self-similar analysis in more general contexts

In the previous computation we can trade $\Delta_x \rho^m$ for any non-linear operator $\mathcal{L}$ with the scaling property

$$\mathcal{L}[u] = \sigma^q \mathcal{L}[\rho].$$

Self-similarity analysis for (CVE) was performed in [41, Theorem 3.1]. The result passes by an obstacle problem. For (NVE) the self-similar analysis is done in [157, Section 6.2]. For the fractional heat equation see [122].

Self-similar solutions are very particular of problem where all terms have the same scaling, and it is not reasonable to expect solutions of this nature to exist in general. Consider the problem

$$\frac{\partial \rho}{\partial t} = \mathcal{L}\rho.$$

When $\mathcal{L}$ is the sum of two terms with different scaling, it is not reasonable to expect self-similar structure. For (ADE) with $U = U_m$, $V = 0$, $W = W_k$ this is only possible in the so-called *fair-competition regime*

$$d(m-1) + k = 0.$$

The self-similar analysis in this setting was performed in [43].

5.1.3 The Keller–Segel problem: bubbles

For $d = 2$ a self-similar solution is not available, although [126] give some hints on the shape by matching asymptotics. Later, there were more advanced studies on stability [173–175]. In this direction, see also [95, 151].

In [125] the authors prove there are no self-similar solution of (KSE$_\chi$) when $d = 2$. However, for $d = 3$ they show that there exists a sequence of radial self-similar solutions $\rho^{(n)}$

$$\rho_t^{(n)}(x) = \frac{1}{T-t} F_n\left(\frac{x}{\sqrt{T-t}}\right), \quad \text{where } F_n(y) \sim \left(\frac{8\pi}{\chi} + a_n\right)\frac{1}{4\pi|y|^2} \text{ for } |y| \sim 0,$$

and $a_n \to 0$. This construction was later generalised for $d \geq 3$ by several authors [123, 156, 160]. Blow-up on the range $U = U_m$ for $m \geq 1$, $V = 0$ and $W = W_k$ with $k > 2 - d$ see [179].

It was later conjectured in [102] and proved in [100] that for $d = 2$ one can construct "approximate bubbles". In particular, an initial datum $\rho_0^\star$ with initial mass 8π so that the solutions of Keller-Segel problem (KSE) such that, for ρ_0 close to $\rho_0^\star$

$$\rho(t, x) \approx \frac{1}{\sigma(t)^2} F\left(\frac{x - \chi(t)}{\sigma(t)}\right), \quad F(y) = \frac{8}{(1 + |y|^2)^2}$$

as $t \to \infty$ where $\sigma(t) \approx c/\sqrt{\log t}$, $\chi(t) \to q$.

5.2 Relative entropy

If we are able to construct *a* global solution B_t of which we know some properties, we would like to see whether this is the "generic behaviour". Assume we are in a case where $\|\rho_t\|_{L^1}$ is preserved. We would say that B_t is attractive if

$$\|\rho_t - B_t\|_{L^1} \to 0, \quad \text{as } t \to \infty.$$

For L^p norms in general (which typically change over time), we would to see if the *relative error* tends to zero

$$\frac{\|\rho_t - B_t\|_{L^p}}{\|\rho_t\|_{L^p}} \to 0, \quad \text{as } t \to \infty. \tag{5.2}$$

These are usually called *intermediate asymptotics*, as opposed to the *long-time asymptotic* limit.

Naïve computations do not yield good results. Sometimes it is possible to do this computation directly via comparison arguments. But, in general, this is not possible. A useful tool is are the so-called *relative entropies*. We start by a simple example, (HE).

5.2.1 L^2 relative entropy for (HE)

An interesting alternative way of writing (HE-FP) as

$$\frac{\partial u}{\partial t} = \text{div}\left(G\nabla \frac{u}{G}\right),$$

where G is the Gaussian profile (3.1). Let us take $w = \frac{u}{G}$ we rewrite this problem as

$$\frac{\partial w}{\partial t} = G^{-1}\,\text{div}\,(G\nabla w) = \Delta w - x \cdot \nabla w.$$

This is the famous Ornstein–Uhlenbeck problem, which is also the dual (HE-FP). We can write the free-energy dissipation formula

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}|w - 1|^2 G\,\mathrm{d}x = -2\int_{\mathbb{R}^d}|\nabla w|^2 G\,\mathrm{d}x.$$

We can now take advantage of the Gaussian Poincaré inequality , see [93],

$$\int_{\mathbb{R}^d} w^2 G \, dx - \left(\int_{\mathbb{R}^d} w G \, dx \right)^2 \le \int_{\mathbb{R}^d} |\nabla w|^2 G.$$

Hence, assuming that $\int_{\mathbb{R}^d} wG = \int_{\mathbb{R}^d} u = 1$, we recover that

$$\int_{\mathbb{R}^d} |w_t - 1|^2 G \, dx \le e^{-2t} \int_{\mathbb{R}^d} |w_0 - 1|^2 G \, dx.$$

This approach can be generalised in many directions. We point the reader to [6] and the references therein.

5.2.2 L^1 Relative entropy argument for (HE)

For convenience, we will denote the solution of (HE-FP) by v. Notice that, with our choice of re-scaling $v_0 = \rho_0$. The relative entropy is defined

$$\mathcal{E}(v) = \int_{\mathbb{R}^d} v \log \frac{v}{G} = \int_{\mathbb{R}^d} v \left(\log v + \frac{x^2}{2} v \right) + C.$$

It is easy to check that

$$\frac{d}{dt} \mathcal{E}(v) = -\mathcal{I}(v),$$

where $\mathcal{I}$ is the so-called *Fisher information*

$$\mathcal{I}(v) = \int_{\mathbb{R}^d} v \left| \frac{\nabla v}{v} + x \right|^2 = \int_{\mathbb{R}^d} v \left| \nabla \log \frac{v}{G} \right|^2.$$

The Gaussian logarithmic Sobolev inequality (see [113, 124, 168])

$$\frac{1}{2} \int_{\mathbb{R}^d} (|f|^2 \log |f|^2) G \, dx \le \int_{\mathbb{R}^d} |\nabla f|^2 G, \quad \text{if} \int_{\mathbb{R}^d} |f|^2 G = 1. \qquad (5.3)$$

This inequality ensures the relationship $\mathcal{E} \le \frac{1}{2}\mathcal{I}$, and hence we recover an ordinary differential inequation that yields $\mathcal{E}(v_t) \le \mathcal{E}(\rho_0)e^{-2t}$. Lastly, we will take advantage of the Czsizar–Kullback inequality taking $G = K_1$

$$\|f - G\|_{L^1}^2 \le 2\mathcal{E}(f).$$

Eventually, we deduce that if ρ is a solution (HE)

$$\|\rho_t - K_t\|_{L^1} \le \sqrt{2\mathcal{E}(\rho_0)} t^{-\frac{1}{2}}.$$

Remark 5.1 It is also worth point out that (5.3) is equivalent to the Euclidean log-Sobolev inequality, which in scale invariant form (see [113, 178]) given by

$$\frac{d}{2} \log \left(\frac{2}{\pi de} \int_{\mathbb{R}^d} |\nabla f|^2 \right) \ge \int_{\mathbb{R}^d} |f|^2 \log |f|^2, \quad \text{if} \int_{\mathbb{R}^d} |f|^2 = 1. \qquad (5.4)$$

Notice that is usually applied to $f = \sqrt{\rho}$, and hence the condition is simply that $\rho \in L^1(\mathbb{R}^d) \cap \mathcal{P}(\mathbb{R}^d)$.

5.2.3 L^1 relative entropy for (PME)

In [83, 103] the authors extend the relative-entropy study to $m > 1$. Now we denote by v the solution of (PME-FP). Suppose there exists $\hat{v}$ of suitable mass and define the relative entropy and the Fisher information

$$H(v) = \int_{\mathbb{R}^d} \left(\frac{2}{m-1} v^m + |x|^2 v \right) \qquad I(v) = \int_{\mathbb{R}^d} v \left| x + \frac{m}{m-1} \nabla v^{m-1} \right|^2 .$$

The suitable relative entropy for $m > 1$ is $\mathcal{H}(u||\hat{u}) = \mathcal{H}(u) - \mathcal{H}(\hat{u})$. This was later extended to more general families than (PME-FP) in [72]. This relative-entropy arguments rely heavy on Gagliardo–Nirenberg–Sobolev inequalities. In fact, there is a deep connection between the smoothing properties of (PME) and these types of inequalities (see, e.g., [103]). In [26, 29] the authors develop the correct relative entropies and Fisher information functions to cover the cases $m < 1$ with the correct re-scaling of the Barenblatt/pseudo-Barenblatt explained in Sect. 5.1.1.

Remark 5.2 This kind of relative entropy arguments for diffusive problems (which do not have a natural stationary state) rely on the self-similar solution. It is possible to prove intermediate asymptotics without homogeneity or self-similarity techniques. For diffusive problems of the form $\frac{\partial \rho}{\partial t} = \Delta \Phi(\rho)$, a very nice analysis using Wasserstein spaces can be found in [59].

5.2.4 Relative entropy argument for (ADE)

The entropy arguments for (PME) are stable enough that allow some families of perturbations into the range (ADE). For $W = 0$ and confining potentials see [72]. The reader may find information on the strictly displacement-convex free-energy functionals [76, 77]. For $(\text{KSE}_{m,\chi})$ with $m \in [1, 2 - \frac{2}{d}]$ (and, in fact, a larger family of W radially decreasing), intermediate asymptotics in the sense (5.2) are obtained in [14]. In [61] the authors take advantage of relative entropy arguments to prove that in the case $U(\rho) = \rho \log \rho$, $V = 0$ and W smooth and bounded (5.2) with $p = 1$ and $B_t = K_t$ (i.e., the heat kernel). See [46] for previous results in this direction.

5.3 Study of the mass variable of radial solutions

One of the significant limitations to prove the above results rigorously is the lack of regularity of ρ. Some regularity can be regained by passing to the so-called *mass variable*.

Let $d = 1$. Assume $x \in \mathbb{R}$ and

$$M(t, x) = \int_{-\infty}^{x} \rho_t(y) \, \mathrm{d}y. \tag{5.5}$$

Integrating (ADE) we show that M satisfies a PDE

$$\frac{\partial M}{\partial t} = \Phi' \left(\frac{\partial M}{\partial x} \right) \frac{\partial^2 M}{\partial x^2} + \frac{\partial M}{\partial x} \frac{\partial V}{\partial x} + \frac{\partial M}{\partial x} \frac{\partial}{\partial x} \left(W * \frac{\partial M}{\partial x} \right)$$

If $\Phi(s) = |s|^{m-1}s$ (PME for ρ), this is the p-Laplacian for M with $p = m + 1$. If $x \in \mathbb{R}^d$ with $d > 1$ it is better to define the mass in volumetric coordinates

$$M(t, v) = \int_{A_v} \rho(t, x) \, \mathrm{d}x, \quad \text{where } A_v = B(0, r) \text{ such that } |A_v| = v.$$

Then, if ρ is radially, we recover for a weight κ

$$\frac{\partial M}{\partial t} = \kappa(v)^2 \Phi' \left(\frac{\partial M}{\partial v} \right) \frac{\partial^2 M}{\partial v^2} + \kappa(v)^2 \frac{\partial M}{\partial v} \frac{\partial V}{\partial v} + \kappa(v)^2 \frac{\partial M}{\partial v} \frac{\partial}{\partial v} \left(W *_{\mathbb{R}^d} \frac{\partial M}{\partial v} \right)$$

This is a Hamilton–Jacobi type equation, and is well suited for the theory of viscosity solutions. In [60] the authors discuss the problem when $W = 0$. They show that one can take advantage of stability properties of viscosity solutions to characterise the steady state. Since $\kappa(0) = 0$ and $\kappa(v) > 0$ for $v > 0$, this equation can be used to show that for smooth V and W and radially symmetric ρ, singularities can only form at $x = 0$. They also show that a Dirac may form, but only in the limit $t \to \infty$.

Many authors have taken advantage of this kind of idea in their arguments. For example, we point to [123, 125, 134, 156].

There is also an interesting connection between the mass variable and the Wasserstein distance for ρ. The simplest case states that if $\rho_1, \rho_2 \in \mathcal{P}(\mathbb{R})$, and M_i are their primitives, then

$$\mathfrak{W}_1(\rho_1, \rho_2) = \|M_1 - M_2\|_{L^1(\mathbb{R})}.$$

A similar formula holds if $\rho_i \in \mathcal{P}(\mathbb{R}^d)$, and are radially symmetric.

The inverse of the mass function Furthermore, if we consider the generalised inverse

$$M^{-1}(s) = \inf\{v \in [-\infty, \infty) : M(v) \geq s\}$$

and $\rho_1, \rho_2 \in \mathcal{P}(\mathbb{R})$, and M_i are their primitives, then

$$\mathfrak{W}_p(\rho_1, \rho_2) = \|M_1^{-1} - M_2^{-1}\|_{L^p(0,1)}.$$

Interestingly, if $\rho(t, x)$ solves a conservation problem, we can also deduce and equation for $u(t, s) = M^{-1}(t, s)$. This equation is usually degenerate, but the coefficients do not depend on s (see, e.g., [68]).

5.4 On the attractiveness of attractors

Relative entropy arguments allow us to show that the steady state (or energy minimiser) in L^1 is an attractor for a large range of equations: from (PME-FP) to all the examples discussed in Sect. 5.2.4.

A more difficult question is to understand the case where the energy minimiser may contain a singular part, e.g.,

$$\widehat{\mu} = \widehat{\rho} + (1 - \|\widehat{\rho}\|_{L^1})\delta_0, \tag{5.6}$$

and $\|\widehat{\rho}\|_{L^1} < 1$. We will see an example in the aggregation dominated range (see Corollary 6.13 below). It is difficult to say whether a measure of this kind can be called a *steady state*, since it cannot be plugged into even the distributional formulation. If the energy is λ-displacemente convex, there is no doubt that the minimiser is a global attractor.

In [60] the authors prove that (ADE*) $m \in (0, 1)$, V smooth, and $K = 0$, the solution converges to the free energy minimiser, even when this contains a singular part. In [57] the authors use the mass to prove that for (ADE*) $m \in (0, 1)$, in large class of K.

In the recent preprint [23], covering (ADE) when $k = 2 - 2s \in (2 - d, 0)$ and $m = \frac{2d}{d+2s}$ (part of the *fair-competition regime* described in Sect. 6.3.3) shows that there is an explicit steady state (given in (6.3) below), and solutions with $\|\rho_0\|_{L^m} > \|\widehat{\rho}\|_{L^m}$ have $\|\rho_t\|_{L^m}$ blow-up in finite time (and hence even we cannot study them in the distributional sense).

6 Minimisation of the free energy

As mentioned in Sect. 4.3.4, displacement λ-convexity with $\lambda > 0$ guarantees exponential contraction with rate

$$\mathfrak{W}_2(\mu_t, \widehat{\mu}_t) \leq e^{-\lambda t} \mathfrak{W}_2(\mu_0, \widehat{\mu}_0)$$

In particular, if $\widehat{\mu}$ is a global minimiser of $\mathcal{F}$ (which is unique by convexity), then the JKO schemes is stationary and so is the gradient flow. This means that

$$\mathfrak{W}_2(\mu_t, \widehat{\mu}) \leq e^{-\lambda t} \mathfrak{W}_2(\mu_0, \widehat{\mu})$$

Therefore, it is a global attractor with exponential rate. First, we will try to characterise the local minimisers using variational arguments. This will lead us to Euler–Lagrange conditions, and generally corresponds to 2-Wasserstein local minimisers. On each connect component of the support of a steady state, the Euler–Lagrange condition comes also from the free-energy dissipation (4.9), and are simply $\frac{\delta F}{\delta \rho} = C$.

However, (ADE) can have steady states that do not satisfy the Euler–Lagrange conditions, specially when the support has several components. A famous example that is known to the community is the case $U = U_m$ with $m > 1$, $W = 0$, and a potential V with two wells, e.g., $V(x) = |x - x_0|^2 + |x - x_1|^2$. Then we can build formal steady states

$$\hat{\rho}(x) = \left((U_m')^{-1}(h_1 - V(x)) \right)_+ + \left((U_m')^{-1}(h_2 - V(x)) \right)_+ \quad \text{even with } h_1 \neq h_2,$$

(6.1)

provided the support of these two terms are disjoint. These are not 2-Wasserstein local minimisers, but they are, actually, ∞-Wasserstein local minimisers. Furthermore, these steady states attract some initial data. The study of ∞-Wasserstein local minimisation became quite popular, and we cite several references below.

6.1 Local minimisers: Euler–Lagrange conditions

The aim of this section is show that if $\hat{\rho} \in L^1(\mathbb{R}^d) \cap \mathcal{P}(\mathbb{R}^d)$ is a local minimiser of $\mathcal{F}$ over $L^1(\mathbb{R}^d) \cap \mathcal{P}(\mathbb{R}^d)$ then we expect it to satisfy the following: there exists $h \in \mathbb{R}$ such that

$$\frac{\delta \mathcal{F}}{\delta \rho}[\hat{\rho}](x) = h \quad \text{a.e. in supp } \widehat{\rho} \tag{EL$_1$}$$

$$\frac{\delta \mathcal{F}}{\delta \rho}[\hat{\rho}](x) \geq h \quad \text{a.e. in } \mathbb{R}^d \tag{EL$_2$}$$

We recall that our main interest is the free energy $\mathcal{F}$ given by (FE*) for which the first variation $\frac{\delta \mathcal{F}}{\delta \rho}$ is given by (2.9).

6.1.1 First variation over the space of absolutely continuous probability measures

We follow the argument of [10, 27, 43, 45, 60, 66, 67, 84] where the reader may find rigorous proofs in different ranges. Our aim is to take variations of the forms

$$\widehat{\rho_\varepsilon} = \widehat{\rho} + \varepsilon\varphi$$

where, if we pick test functions φ suitably, we still have $\widehat{\rho_\varepsilon} \in L^1 \cap \mathcal{P}$ so

$$\mathcal{F}(\widehat{\rho}) \leq \mathcal{F}(\widehat{\rho_\varepsilon}).$$

Variations on the support of $\hat{\rho}$ Consider a test function $\psi \in C_c^\infty(\mathbb{R}^d)$ and introduce

$$\varphi(x) = \left(\psi(x) - \int_{\mathbb{R}^d} \psi(y)\hat{\rho}(y)\,\mathrm{d}y\right)\hat{\rho}(x)$$

Then for $\varepsilon < \frac{1}{2}\|\psi\|_{L^\infty}^{-1}$ we have $\hat{\rho}_\varepsilon \in L^1 \cap \mathcal{P}_2$ and, using that $\widehat{\rho}$ is a minimiser and the definition of first variation, we get

$$\int_{\mathbb{R}^d} \frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}]\varphi = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left(\mathcal{F}(\rho_\varepsilon) - \mathcal{F}(\rho)\right) \geq 0.$$

Expanding this integral and using $\|\hat{\rho}\|_{L^1} = 1$ we get

$$0 \leq \int_{\mathbb{R}^d} \frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](x)\left(\psi(x) - \int_{\mathbb{R}^d}\psi(y)\hat{\rho}(y)\,\mathrm{d}y\right)\hat{\rho}(x)\,\mathrm{d}x$$

$$= \int_{\mathbb{R}^d} \frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](x)\hat{\rho}(x)\psi(x)\,\mathrm{d}x - \left(\int_{\mathbb{R}^d}\frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](x)\hat{\rho}(x)\,\mathrm{d}x\right)\left(\int_{\mathbb{R}^d}\psi(y)\hat{\rho}(y)\,\mathrm{d}y\right)$$

Lastly, trading x and y on the last two integrals we recover that for all $\psi \in C_c^\infty(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d}\left(\frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](x) - h(\hat{\rho})\right)\psi(x)\hat{\rho}(x)\,\mathrm{d}x \geq 0, \qquad \text{where } h(\hat{\rho}) = \int_{\mathbb{R}^d}\frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](y)\hat{\rho}(y)\,\mathrm{d}y.$$

Since this also holds for $-\psi$ the equality holds above, and hence

$$\left(\frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](x) - h(\hat{\rho})\right)\widehat{\rho}(x) = 0, \qquad \text{for a.e. } x \in \mathbb{R}^d$$

Hence we recover (EL$_1$). This gives us complete information on the boundary of the support. But it could jump uncontrollably from 0 to the other profile.

Variations outside the support of $\hat{\rho}$ Now we take the variations

$$\varphi(x) = \psi(x) - \hat{\rho}(x)\int_{\mathbb{R}^d}\psi(y)\,\mathrm{d}y.$$

If $\psi \geq 0$ and $\varepsilon < \|\psi\|_{L^1}^{-1}$ then $\hat{\rho}_\varepsilon \geq 0$. Notice that we cannot hope for $\psi < 0$ outside the support of $\hat{\rho}$. With a similar argument as before we get

$$\int_{\mathbb{R}^d}\psi\left(\frac{\delta\mathcal{F}}{\delta\rho}[\hat{\rho}](x) - h(\hat{\rho})\right) \geq 0.$$

Since this holds true for all ψ is the conditions above, we have proven (EL$_2$).

Remark 6.1 For (ADE) we have that

$$h(\hat{\rho}) = \int_{\mathbb{R}^d} (U'(\hat{\rho})\hat{\rho} + V\hat{\rho} + (W * \hat{\rho})\hat{\rho})$$

With PME diffusion

$$h(\hat{\rho}) = \mathcal{F}(\hat{\rho}) + (m-1) \int \rho^m + \frac{1}{2} \int_{\mathbb{R}^d} (W * \hat{\rho})\hat{\rho}$$

6.1.2 Solving the Euler–Lagrange equation when $U \not\equiv 0$

When $U \not\equiv 0$ then U' is non-decreasing, and we get from (EL$_2$)

$$\hat{\rho} \geq (U')^{-1}(h - V - W * \hat{\rho}) \tag{6.2}$$

we also know that $\hat{\rho} \geq 0$, and (6.2) equation holds with equality when $\hat{\rho} > 0$. If the right-hand of (6.2) is positive, so is $\hat{\rho}$ and equality holds. When the right-hand side of (6.2) is non-positive, then $\hat{\rho} = 0$. Therefore, we have that

$$\hat{\rho} = \left((U')^{-1}\left(h - V - W * \hat{\rho}\right) \right)_+, \qquad \text{for some constant } h \in \mathbb{R}, \tag{EL$_3$}$$

In particular, observe that when $U = U_m$, $V = \frac{|x|^2}{2}$ and $W = 0$ (i.e., (PME-FP)) this is the famous Barenblatt solution (ZKB). More generally, the same approach holds true for the case $W = 0$. In this case, (EL$_3$) is a one-parameter family $\hat{\rho}_h$. Furthermore, since U' is non-decreasing we have that $\hat{\rho}_h$ is point-wise non-decreasing with h. Hence, in most cases h can be recovered from the fixed value $\int_{\mathbb{R}^d} \hat{\rho}_h$.

Remark 6.2 For the fast diffusion case $U'(0) = \infty$. Hence, if $|h| < \infty$ and $W * \rho \in L^\infty_{loc}$ then $\hat{\rho} > 0$.

Notice that when $\hat{\rho} > 0$ then the solutions of (EL$_3$) are precisely stationary solutions in the sense of (ADE-S). When the support of $\hat{\rho}$ minimisers are still formally stationary, however the regularity on the free boundary (the boundary of the support) is more challenging.

The cases $U = 0$ are also interesting. Take, for example (CVE) and re-scale to a Fokker–Planck problem. Then $\frac{|x|^2}{2} + (-\Delta)^{-s}\hat{\rho} \geq h$. Let $v := (-\Delta)^{-s}\rho$. Hence, $(-\Delta)^s v = \hat{\rho} \geq 0$. We also get, due to (EL$_1$), that when $\hat{\rho} > 0$ then equality holds in the last equation. Hence, we can write

$$\min\left\{ (-\Delta)^s v(x),\, v(x) + \frac{|x|^2}{2} - h \right\} = 0, \qquad \text{for all } x \in \mathbb{R}^d.$$

This is the well-known fractional obstacle problem. In this particular application it is discussed in [41]. Regularity of solutions was proved in [158].

Lastly, there is the case of the Newtonian/Riesz potential $W = -cW_k$ for $k = 2 - 2s \in [2 - d, 0)$. Then $u := W * \rho = (-\Delta)^{-s}\rho$. Hence, we can re-write (EL$_3$) as

$$(-\Delta)^s u = ((U')^{-1}(u + h - V))_+$$

The case of U_m with $m > 1$, $V = 0$, and $s \in (0, 1)$ was studied in [89].

The particular case $k \in (2 - d, 0)$ and $m = \frac{2d}{2d+k}$ has the correct scaling (sometimes called *conformal*). The explicit self-similar solution can be found using the result in [92] and the references therein (see also [23]) to be

$$\hat{\rho}(x) = B \left(\frac{\lambda}{\lambda^2 + |x - x_0|^2} \right)^{\frac{d+2s}{2}}. \tag{6.3}$$

Remark 6.3 This also shows that in $\mathbb{R}^d$, when U' is strictly increasing, $U'(0) = -\infty$, and V, W is bounded, there can exist no stationary states. We compute the lower bound

$$\hat{\rho}_h(x) \geq (U')^{-1}(h - \|V\|_\infty - \|W\|_\infty \|\rho\|_\infty) > 0.$$

This means that, $\hat{\rho}_t \notin L^1(\mathbb{R}^d)$ is not for any $h \in \mathbb{R}$.

Remark 6.4 (*Bifurcation analysis*) Several authors have paid attention to the structure of local minimisers in terms of bifurcations. These analyses can be done for (KE) in terms of Fourier coefficients. In [64] the authors study bifurcation of (McKVE) (with periodic boundary conditions, or equivalently in $\mathbb{T}^d$, with $W = \kappa \bar{W}$, and $V = 0$) by applying Crandall-Rabinowitz theory. They see local minimisers as solutions of the equation

$$(\rho, \kappa) \mapsto \rho - \frac{e^{-\beta \kappa \bar{W} * \rho}}{\int e^{-\beta \kappa (\bar{W} * \rho)(y)} \, \mathrm{d}y},$$

over the space of even functions. This was later extended in [65] to (ADE) (again in the torus but with $U = U_m$ and $V = 0$) by taking the suitable extension of the energy.

Remark 6.5 When $U = 0$ the equation $V + W * \rho = h$ has been studied using different approaches, but we include these references in Sect. 6.3.2 below.

6.2 Extension of the free energy to measure space

Our free energy $\mathcal{F}$ is defined for $\rho \in \mathcal{P}_2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. This set is not dense in $\mathcal{P}_2(\mathbb{R}^d)$ with the total variation metric since for $\|\rho - \delta_0\|_{TV} = 2$. However, it is dense in the narrow topology. We can therefore try to extend $\mathcal{F}$. The natural weak lower-semicontinuous extension is given by

$$\widetilde{\mathcal{F}}(\mu) = \inf_{\rho_n \rightharpoonup \mu} \liminf_n \mathcal{F}(\rho_n).$$

Clearly, this is not so easy to compute directly.

Let us look first at the diffusion term in the homogeneous range U_m. As an extreme example, we have to make sense of δ_0^m. An easy approach is to take $\rho_1 \in \mathcal{P}_2(\mathbb{R}^d) \cap C_c^\infty(\mathbb{R}^d)$, define $\rho_\varepsilon = \varepsilon^{-d} \rho_1(\varepsilon^{-1}x)$. Then $\rho_\varepsilon \rightharpoonup \delta_0$. Integrating we recover

$$\mathcal{U}_m(\rho_\varepsilon) = \varepsilon^{d(1-m)} \mathcal{U}_m(\rho_1).$$

Therefore, the extension has to be

$$\widetilde{\mathcal{U}_m}(\delta_0) = \begin{cases} +\infty & \text{if } m > 1, \\ 0 & \text{if } m \in (0, 1). \end{cases}$$

The rigorous construction of $\widetilde{\mathcal{U}}$ in general setting can be found in [107]. In fact

$$\widetilde{\mathcal{U}}(\mu) = \begin{cases} +\infty & \text{if } \lim_{s \to \infty} \frac{U(s)}{s} = \infty \text{ and } \mu_{\text{sing}} \neq 0, \\ \mathcal{U}(\mu_{ac}) & \text{if } \lim_{s \to \infty} \frac{U(s)}{s} = 0, \end{cases}$$

where $\mu = \mu_{\text{ac}} + \mu_{\text{sing}}$ the absolutely continuous and singular parts of the measure. If V, W are well-behaved (e.g., bounded by $C(1 + |x|^2)$) then

$$\widetilde{\mathcal{F}}(\alpha\delta_0 + \rho) = \widetilde{\mathcal{U}}(\rho) + \int V \, \mathrm{d}\mu + \frac{1}{2} \iint (W * \mu) \, \mathrm{d}\mu.$$

For example, if U is sublinear and V, W are well-behaved

$$\widetilde{\mathcal{F}}(\alpha\delta_0 + \rho) = \mathcal{F}(\rho) + \alpha V(0) + 2\alpha \int_{\mathbb{R}^d} W(x)\rho(x) \, \mathrm{d}x + \alpha^2 W(0).$$

The first variation in this setting is a little more involved, but it follows the same general scheme.

However, if W is singular at 0 the landscape can be richer, for example as mentioned in Remark 6.15 for (KSE_χ), the correct behaviour extension is more difficult to construct. In fact, we point the reader to [106] where the authors construct example of W with infinitely many radially decreasing steady states.

6.3 Existence of minimisers

6.3.1 General comments

For example, it is easy to show that when $m > 1$, and V, W are λ-convex for $\lambda > 0$ and bounded below, then the free-energy functional admits a unique minimiser with the properties above using the direct method of calculus of variations. First, the functional is bounded below since $U_m \geq 0$ for $m > 1$. We take a minimising sequence ρ_n, and we use Prokhorov's theorem to prove it has a narrow limit $\hat{\mu}$. Using weak lower semi-continuity we show that the limit is indeed a minimiser.

The situation in bounded domain Ω is easier, since all sequences of measures are tight (i.e., mass cannot escape to infinity) and the Lebesgue spaces are embedded in each other. We focus therefore on the more complicated case $\mathbb{R}^d$.

To find energy minimisers in $\mathbb{R}^d$, one can try to construct minimising profiles by scaling. Given a profile ρ_1, we can rescale it like

$$\rho_\lambda(x) = \lambda^d \rho_1(\lambda x).$$

It might happen that *full diffusion* (i.e., $\lambda \to 0$) is energy beneficial. Indeed, we have the scaling

$$\mathcal{U}_m(\rho_\lambda) = \begin{cases} \lambda^{d(m-1)}\mathcal{U}_m(\rho_1) & \text{if } m \neq 1 \\ \mathcal{U}_m(\rho_1) + \log\lambda & \text{if } m = 1, \end{cases}$$

Notice that, when $m > 1$ then $d(m-1)$ and $\mathcal{U}_m(\rho_1) > 0$, whereas if $m < 1$ they are both negative. Furthermore, we observe that, when $m > 1$, then the energy is bounded below; and for $m \in (0, 1]$ it is not.

It might also happen that *full concentration* (i.e., $\lambda \to \infty$) is best for the energy, for example for U, $W = 0$ and $V = |x|^2$. Then we expect the minimiser to be a δ_0.

6.3.2 The case $U = 0$

There is a long literature looking for minimisers when $U = 0$. In this setting, some authors have studied the $\mathfrak{W}_p$ local minimisers with $p \in (1, \infty]$. In [10] the authors realised that, when

$V = 0$, the minimisers can be supported over sets of different dimensions, depending on the attractive-repulsive nature of W. The stability of these minimisers was later studied in [11]. In [12] the authors discuss the compact support of minimisers under general settings. For general conditions on existence of global minimisers see [159]. The regularity of compactly supported ∞-Wasserstein when $V = 0$ was studied in [53]. Recently, analysis of the Fourier transform has been applied to determine further structure of minimisers [79, 81].

There is the particularly interesting case of $V = 0$ and the power-type attractive-repulsive potential

$$W(x) = \frac{|x|^\gamma}{\gamma} - \frac{|x|^\alpha}{\alpha}. \tag{6.4}$$

In this direction, there have a number of works. The authors of [85] show existence of minimisers coming as a limit of empirical measures (recall (2.4)) when $\gamma > \alpha$. In dimension 1 one can construct solutions of $W * \rho = E$ by inverting Fredholm operators (see [70]). Nevertheless, the measures

$$\mu_{\mathfrak{m}} = \mathfrak{m}\delta_0 + (1 - \mathfrak{m})\delta_1$$

are of special relevance. In [131] the authors show in $d = 1$ if $\alpha \geq 2$ and γ is large enough, then $\mu_{\frac{1}{2}}$ is unique $\mathfrak{W}_p$ minimiser for $p \in [1, \infty)$. The $\mathfrak{W}_\infty$ minimisation when $d = 1$ is richer:

- If $\gamma > \alpha > 2$, for every $\mathfrak{m} \in (0, 1)$, $\mu_{\mathfrak{m}}$ is a $\mathfrak{W}_\infty$-strict local minimiser.
- If $\gamma > 3$, $\alpha = 2$, for every $\mathfrak{m} \in (\frac{1}{\gamma-1}, \frac{\gamma-2}{\gamma-1})$, $\mu_{\mathfrak{m}}$ is a $\mathfrak{W}_\infty$-strict local minimiser.
- If $\gamma > 3$, $\alpha = 2$,for every $\mathfrak{m} \in (0, \frac{1}{\gamma-1}] \cup [\frac{\gamma-2}{\gamma-1}, 1)$, $\mu_{\mathfrak{m}}$ is a $\mathfrak{W}_\infty$-saddle point.
- If $3 > \gamma > \alpha = 2$, for every $\mathfrak{m} \in (0, 1)$, $\mu_{\mathfrak{m}}$ is a $\mathfrak{W}_\infty$-saddle point.
- If $\gamma = 3$, $\alpha = 2$, for every $m \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$, $\mu_{\mathfrak{m}}$ is a $\mathfrak{W}_\infty$-saddle point.
- If $\gamma = 3$, $\alpha = 2$, $\mu_{\frac{1}{2}}$ is a $\mathfrak{W}_\infty$-strict local minimiser.

When $d > 1$ the suitable extension of $\rho_{\mathfrak{m}}$ is a "Dirac delta" supported on the surface of a ball, see [98, 99]. The explicit shape of the $\mathfrak{W}_\infty$-minimiser when $\gamma \in (2, 3)$ and $\alpha = 2$ is given in [119, 120]. There is also significant interest in the case $\gamma = 2, \alpha \in (-2, 0)$, see [80] and the references therein.

There are several related problems with different choices of V, W which have been studied over the last few years due to the modelling, we point the reader to [74, 75, 144], and the references therein.

6.3.3 The power cases $U = U_m$, $V = 0$, and $W = \chi W_k$

The family of cases $U_m(\rho) = \frac{\rho^m}{m-1}$, $V = 0$, and $W = \chi W_k$ has been studied in extensive detail. We analyse the scaling of the second term of the energy to recover

$$\mathcal{W}_k(\rho_\lambda) = \begin{cases} \lambda^{-k}\mathcal{W}(\rho_1) & \text{if } k \neq 0 \\ \mathcal{W}_k(\rho_1) - \log \lambda & \text{if } k = 0, \end{cases}$$

The scaling of the energy is given by

$$\mathcal{F}_{m,\chi,k}(\rho_\lambda) = \lambda^{-k}\left(\lambda^{d(m-1)+k}\mathcal{U}_m(\rho_1) + \chi\mathcal{W}(\rho_1)\right)$$

The sign inside the parenthesis is crucial. As pointed out in [43, 67] the two terms are in balance when $d(m - 1) = k$, and this leads to the critical value

$$m_c = \frac{d - k}{d} \tag{6.5}$$

called the *fair-competition regime*. For $m > m_c$ we have the so-called *diffusion-dominated regime* and for $m \in (0, m_c)$ the *aggregation-dominated regime*. Notice that $k > 0$ implies $m_c < 1$ (fast-diffusion) whereas $k < 0$ implies $m_c > 1$ (slow-diffusion).

Remark 6.6 The case scaling argument can be repeated for $\mathcal{F} = \mathcal{U}_m + \chi \mathcal{V}_\lambda$, and we recover also the critical value $m_c = \frac{d-\lambda}{d}$. The critical value of (PME-FP) corresponds precisely to $\lambda = 2$.

Remark 6.7 Notice that (KSE$_{m,\chi}$) corresponds precisely to $\lambda = 2 - d$, so we recover $m_c = \frac{2d-2}{d}$. The case $U = U_m$ with $m \geq 2$, $V = 0$ and W radially decreasing was studied in [13], where the main tool is radially decreasing rearrangement of ρ (which will be presented in Sect. 6.4).

Remark 6.8 (*Hardy–Littlewood–Sobolev*) To show that the free energy functionals are bounded below, we recall the Hardy–Littlewood–Sobolev inequality $k \in (-d, 0)$ and $\rho \in \mathcal{P}(\mathbb{R}^d)$

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\lambda \rho(x)\rho(y)\, dx\, dy \leq C_{\text{HLS}}(k, d) \int_{\mathbb{R}^d} \rho(x)^{m_c}\, dx, \qquad (6.6)$$

where we recall the definition of m_c given in (6.5). A proof can be found in [43, Theorem 3.1] The logarithmic version is that if $\log(1 + |\cdot|^2)\rho \in L^1$ and $\rho \in \mathcal{P}(\mathbb{R}^d)$ then

$$-\iint_{\mathbb{R}^d \times \mathbb{R}^d} \log|x - y|\rho(x)\rho(y)\, dx\, dy \leq \tfrac{1}{d} \mathcal{U}_1(\rho) + C_0. \qquad (6.7)$$

The fair-competition regime was discussed in [43], the diffusion-dominated regime in [67], and the aggregation-dominated regime in [54].

Since there is fair competition, the rescaling done for (PME) suitably rescales the equation. There is, as usual, a new term $\text{div}(x\rho)$ coming from the time derivative. This new equation is itself the 2-Wasserstein gradient flow of the rescaled free energy

$$\mathcal{F}_{\text{resc}} := \mathcal{F} + \mathcal{V}_2,$$

where we recall the definition (3.7). The main results are as follows

Theorem 6.9 (Fair-competition case [43]) *Let $m = m_c$ and $k \in (-d, 0)$. Then*

1. *Stationary states satisfy $\mathcal{F}(\hat{\rho}) = 0$.*
2. *We have that*

$$\mathcal{F}(\rho) \geq \frac{1 - \chi C_{\text{HLS}}(k, d)}{d(m - 1)} \|\rho\|_{L^m}^m.$$

   Due to the previous item, we define $\chi_c = 1/C_{\text{HLS}}(k, d)$.
3. *If $\chi = \chi_c$, then there exists a global minimiser.*
4. *If $\chi < \chi_c$, then there exists a global minimiser for $\mathcal{F}_{\text{resc}}$, but not for $\mathcal{F}$.*
5. *If $\chi > \chi_c$, then $\mathcal{F}$ and $\mathcal{F}_{\text{resc}}$ are not bounded below.*

Theorem 6.10 (Diffusion-dominated regime [67]) *Let $m > m_c$, $d \geq 1$, $\chi > 0$, and $k \in (-d, 0)$. Then there exists a global minimiser $\hat{\rho} \in L^1_+(\mathbb{R}^d) \cap L^m(\mathbb{R}^d)$, and it is radially symmetric and compactly supported.*

The work [66] also contains contributions towards existence of minimisers in the diffusion-dominated range. However, the key contribution is the radial symmetry as described in Sect. 6.4. More or less in parallel, work was done on the aggregation-dominated regime.

Theorem 6.11 (Aggregation-dominated regime [54]) *Assume that* $W \in L^1_{loc}(\mathbb{R}^d)$ *and* $W(x) = W(-x)$.

1. *(Linear diffusion) Assume, furthermore, that* $W \in L^\infty(\mathbb{R}^d \setminus B_\delta)$. *Then, for any* $\varepsilon > 0$ *the energy* $\varepsilon \mathcal{U}_1 + \mathcal{W}$ *does not admit any* $\mathfrak{W}_p$-*local minimisers. Furthermore, if W is Lipschitz continuous, then it does not admit any critical points.*
2. *(Fast diffusion) If* $m < \frac{d}{d+k}$ *then the free energy* $\varepsilon U_m + \mathcal{W}_k$ *is not bounded below in* $\mathcal{P}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$.
3. *(Slow diffusion) If* $m > 1$ *and* $k \in ((1-m)d, 0)$ *then the energy*

$$\mathcal{F}(\rho) = \begin{cases} \mathcal{U}_m(\rho) + \mathcal{W}_k(\rho) & \text{if } \rho \in L^m(\mathbb{R}^d) \\ +\infty & \text{if } \rho \in \mathcal{P}(\mathbb{R}^d) \setminus L^m(\mathbb{R}^d) \end{cases}$$

has a global minimiser.

The authors also provide a condition for sharp existence of global minimisers. In [55] the authors introduce a reversed HLS inequality to deal with the aggregation-dominated regime.

Theorem 6.12 (Reversed HLS [55])

1. *If* $k > 0$ *and* $m \in (\frac{d}{d+k}, 1)$ *then there exist a constant* $C \in \mathbb{R}$ *such that*

$$\frac{\iint_{\mathbb{R}^d \times \mathbb{R}^d} f(x)|x-y|^k f(y) \, dx \, dy}{\left(\int_{\mathbb{R}^d} f(x) \, dx \right)^\alpha \left(\int_{\mathbb{R}^d} f(x)^m \, dx \right)^{\frac{2-\alpha}{m}}} \geq C, \tag{6.8}$$

where

$$\alpha = \frac{2d - m(2d+k)}{d(1-m)}.$$

2. *The equality is in* (6.8) *achieved if* $d = 1, 2$ *or* $d = 3$ *and* $m \geq \min\{\frac{d-2}{d}, \frac{2d}{2d+k}\}$.
3. *If* $m \in (\frac{d}{d+k}, 1)$ *and either* $\lambda \in [2, 4]$ *or,* $\lambda \geq 1$ *and* $m \geq \frac{d-1}{d}$ *then the minimiser of* (6.8) *is unique (up to translation, dilation, or multiplication).*

Part of the results in [55] appeared first as a preprint [51] written in terms of (ADE) instead of functional inequalities. We present the results as stated there as a corollary of the published paper.

Corollary 6.13 *Let* $m \in (\frac{d}{d+k}, 1)$ *and* $k > 0$. *Then*

1. *There exists a unique* $\widehat{\mu}$ *that minimises the free energy* $\widetilde{\mathcal{U}}_m(\widehat{\mu}) + \chi \mathcal{W}(\widehat{\mu})$, *and they coincide with minimisers of* (6.8).
2. *It is of the form* (5.6) *for some* $\widehat{\rho} \in L^1$.
3. *For* $m > \frac{d-1}{d}$ *and* $k \geq 1$ *the functional is geodesically convex in* $\mathfrak{W}_2$, *and hence, the minimiser is unique.*

The authors also present the suitable Euler–Lagrange equation for $\widehat{\mu}$. The computation is more involved that (6.1.1) that follows the same philosophy.

The case $\lambda = 4$ was studied in full detail in [52]. Here the authors characterise precisely the dimensions d and index m for has a Dirac delta, i.e., it is (5.6) with $\|\widehat{\rho}\|_{L^1} < 1$.

Remark 6.14 A particular amount of work has been devoted to the case $V = 0$ whereas $V \neq 0$ raised much lower interest. The extension of the results is more or less direct, see [57, 60].

Remark 6.15 Keller–Segel corresponds precisely to the logarithmic cases $d = 2, m = 1$, and $W = \frac{-1}{2\pi} W_0$ which yield (assuming $\|\rho\|_{L^1} = 1$)

$$\mathcal{F}[\rho_\lambda] = \mathcal{F}(\rho_1) + \left(2 - \frac{\chi}{4\pi}\right) \log \lambda.$$

This leads to the critical value of $\chi^* = 8\pi$ or, equivalently recalling the change of variable in Sect. 3.4, critical mass $M^* = 8\pi$.

6.4 Radial symmetry

There are several approaches to prove that, when V, W are radially symmetric then the minimisers must be radially symmetric. One approach is to use the associated stationary problem, and Alexandrov reflexions.

A different approach is to prove that we can find radially symmetric minimising sequences. For this, we can take advantage of the rearrangement theory (see [166] and the references therein). First, let us "re-arrange" a set. If $A \subset \mathbb{R}^d$ we define the rearrangement of A as the ball of radius 0

$$A^\star := B(0, R) \qquad \text{such that } |A^\star| = |A|.$$

We define the *radially decreasing re-arrangement* of a non-negative function f as the function $f^\star$ such that

$$\{x : f^\star(x) \geq t\} = \{x : f(x) \geq t\}^\star.$$

We can write f in terms of its level sets via the *layer-cake representation*

$$f(x) = \int_0^{f(x)} dt = \int_0^\infty 1_{[0, f(x)]}(t) \, dt = \int_0^\infty 1_{L_t^+(f)}(x) \, dt,$$

where 1_A is the indicator function of A, and we have introduced the super-level set notation

$$L_t^+(f) := \{x : f(x) \geq t\}.$$

Hence, we can simply define

$$f^\star(x) := \int_0^\infty 1_{L_t^+(f)^\star}(x) \, dt.$$

It is not difficult to check that if f, g, h are non-negative functions vanishing at infinity we have the conservation of all L^p norms, which can be more generally written as

$$\int_{\mathbb{R}^d} U(f^\star(x)) \, dx = \int_{\mathbb{R}^d} U(f(x)) \, dx, \qquad \text{for all } U \text{ convex.}$$

Furthermore, we have Hardy–Littlewood's inequality

$$\int_{\mathbb{R}^d} f^\star(x) g^\star(x) \, dx \geq \int_{\mathbb{R}^d} f(x) g(x) \, dx,$$

and Riesz's inequality

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} f^{\star}(x) g^{\star}(y) h^{\star}(x - y) \, \mathrm{d}x \, \mathrm{d}y \geq \iint_{\mathbb{R}^d \times \mathbb{R}^d} f(x) g(y) h(x - y) \, \mathrm{d}x \, \mathrm{d}y.$$

By reverting these inequalities, if V, W are non-negative and radially increasing we have that

$$\mathcal{F}[\rho^{\star}] \leq \mathcal{F}[\rho].$$

Hence, the expected behaviour is that if $\hat{\rho}$ is a minimiser, then so is its radially decreasing rearrangement. Furthermore, any minimising sequence ρ_k can be replaced by a radially-decreasing minimising sequence. However, rigorously proving this intuition is rather technical. A fairly general theorem can be seen in [66], where the authors take advantage of the continuous Steiner rearrangement.

7 Numerical methods

7.1 Finite volumes

Finite volume schemes are a family of schemes for conservation equations which incorporate the idea of transport of mass. They are easiest to introduce $d = 1$. Take a spatial mesh $x_i = i \Delta x$ we aim to approximate numerically the averages

$$\rho_i \approx \frac{1}{x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \rho(x) \, \mathrm{d}x.$$

The conservation equation (2.2) yields

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \rho(t, x) \, \mathrm{d}x = -\left(F(x_{i+\frac{1}{2}}) - F(x_{i-\frac{1}{2}}) \right).$$

Discretising in time we arrive at the general form

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + \frac{F_{i+\frac{1}{2}}^{n+1} - F_{i-\frac{1}{2}}^{n+1}}{\Delta x} = 0.$$

The no-flux condition is approximated by $F_{-N-\frac{1}{2}} = F_{N+\frac{1}{2}} = 0$, so $i = -K, \ldots, K$. For transport problems in which $\mathbf{F} = \rho \mathbf{v}$, the flux is taken up-wind to preserve positivity of ρ_i^n, namely

$$F_{i+\frac{1}{2}} = \rho_i (v_{i+\frac{1}{2}})_+ + \rho_{i+1} (v_{i+\frac{1}{2}})_-, \tag{7.1}$$

where $u_+ = \max\{u, 0\}$, $u_- = \min\{u, 0\}$ so that $v = v_+ + v_-$. For (ADE) the velocity field $v_{i+\frac{1}{2}}$ is an approximation of the velocity field

$$v_{i+\frac{1}{2}} \approx -\nabla (U'(\rho) + V + W * \rho)(x_{i+\frac{1}{2}}).$$

In [7] the authors introduce a scheme with energy decay for (ADE). When $W = 0$, if U is concave then so is $E(x, \rho) = U(\rho) + V(x)\rho$ and hence we can proof a version of (4.8)

with an inequality

$$\frac{1}{\Delta t}\left(\sum_i E(x_i, \rho_i^{n+1}) - \sum_i E(x_i, \rho_i^n) \right)$$

$$\leq \sum_i \frac{\partial E}{\partial \rho}(x_i, \rho_i^{n+1}) \frac{\rho_i^{n+1} - \rho_i^n}{\Delta t}$$

$$\leq -\sum_i \frac{\partial E}{\partial \rho}(x_i, \rho_i^{n+1}) \frac{F_{i+\frac{1}{2}}^{n+1} - F_{i-\frac{1}{2}}^{n+1}}{\Delta x}$$

$$= \sum_i \frac{\frac{\partial E}{\partial \rho}(x_{i+1}, \rho_{i+1}^{n+1}) - \frac{\partial E}{\partial \rho}(x_i, \rho_i^{n+1})}{\Delta x} F_{i+\frac{1}{2}}^{n+1}.$$

Due to the convexity tricked used above, it is natural to take the implicit method

$$v_{i+\frac{1}{2}}^{n+1} = -\frac{\frac{\partial E}{\partial \rho}(x_{i+1}, \rho_{i+1}^{n+1}) - \frac{\partial E}{\partial \rho}(x_i, \rho_i^{n+1})}{\Delta x}$$

$$F_{i+\frac{1}{2}}^{n+1} = \rho_i^{n+1}(v_{i+\frac{1}{2}}^{n+1})_+ + \rho_{i+1}^{n+1}(v_{i+\frac{1}{2}}^{n+1})_-.$$

Then we have the energy decay

$$\frac{1}{\Delta t}\left(\sum_i E(x_i, \rho_i^{n+1}) - \sum_i E(x_i, \rho_i^n) \right) \leq -\sum_i \left(\rho_i^{n+1}(v_{i+\frac{1}{2}}^{n+1})_+^2 + \rho_{i+1}^{n+1}(v_{i-\frac{1}{2}}^{n+1})_-^2 \right) \leq 0$$

When W is introduced similar estimates are possible, although the bilinear terms need delicate handling. In [7] the authors prove that a successful scheme is

$$\xi_i^{n+1} = U'(\rho_i^{n+1}) + V(x_i) + \Delta x \sum_{j=1}^n W(x_i - x_j) \frac{\rho_j^{n+1} + \rho_j^n}{2} \tag{7.2a}$$

$$v_{i+\frac{1}{2}}^{n+1} = -\frac{\xi_{i+1}^{n+1} - \xi_i^{n+1}}{\Delta x}, \tag{7.2b}$$

$$F_{i+\frac{1}{2}}^{n+1} = \rho_i^{n+1}(v_{i+\frac{1}{2}}^{n+1})_+ - \rho_{i+1}^{n+1}(v_{i+\frac{1}{2}}^{n+1})_-, \tag{7.2c}$$

with the corresponding boundary no-flux conditions. This scheme has energy decay for any W even. The authors of [7] also prove that under certain assumptions on W the decay of the energy holds for other discretisations without the midpoint $(\rho_j^{n+1} + \rho_j^n)/2$. And a proof of convergence through compactness for the cases $m > 2$ was later introduced in [9] convergence. This scheme can be adapted to (ADE*) by

$$\xi_i^{n+1} = U'(\rho_i^{n+1}) + V(x_i) + \Delta x \sum_{j=1}^n K(x_i, x_j) \frac{\rho_j^{n+1} + \rho_j^n}{2}. \tag{7.2a'}$$

A very similar scheme for (AE) is studied in [105], where the authors prove convergence in Wasserstein sense even for pointy potentials. This scheme is later extended for linear diffusion in [137].

A higher order method for the case $U, V = 0$ and W Lipschitz was constructed in [58]. The authors prove convergence in 1-Wasserstein norm.

Remark 7.1 Some finite-volume methods have been studied actually correspond to the gradient of a discrete energy in a discrete version of the Wasserstein metric, see [152].

Remark 7.2 This kind of method have been generalised to $\partial_t \rho = \text{div}(m(\rho)\nabla \frac{\delta F}{\delta \rho}))$ when m is concave (see [8]).

7.2 Methods based on discretising the Wasserstein distance

There are several possibilities of numerical computing the Wasserstein gradient flow. One could combine a method numerically computing the Wasserstein distance (see, e.g., [150]) with a numerical optimisation in the JKO scheme. A very interesting paper in a similar direction is [50], where the authors combine of the JKO scheme with the Benamou–Brenier formula (recall (4.14) above) to construct a discrete gradient-flow method, which they call *primal-dual method*.

A different approach is to use a modified energy for which the gradient flow is easier to compute. This leads to the particle/blob methods we will discuss in the next section.

7.3 Particle/Blob methods

When $U = 0$ and V, W are Lipschitz, (ADE) admits exact solutions given by empirical distributions

$$\mu_t^{(K)} = \sum_{i=1}^K w_i \delta_{X_t^{(i)}}.$$

Hence, the PDE can be reduced to a finite set of ODEs, that can be solved by any method for ODEs. If we are given $\rho_0 \in L^1$, then for any K we approximate it by an empirical measure $\mu_0^{(K)}$.

When $U \neq 0$ this is no longer the case, because the PDE is "regularising". Furthermore, the free of empirical measures is not defined for $U = U_m, m \geq 1$. However, there are two approaches to approximate the free energy so that the $\mathfrak{W}_2$-gradient flow admits again these kinds of solutions.

7.3.1 The particle method

A first approach was introduced in [78] and perfected in [69]. Fix an initial mass

$$\mu_0^{(K)} = \sum_{i=1}^K w_i \delta_{X_0^{(i)}}.$$

Fix these weights of the particles and K the number of particles. Take the set of empirical distributions of K elements and fixed weights

$$\mathcal{A}_{K,w}(\mathbb{R}^d) = \left\{ \mu = \sum_{i=1}^K w_i \delta_{x_i} \,:\, \text{for some } x_i \in \mathbb{R}^d \right\}.$$

The aim is to approximate the free energy $\mathcal{F}$ so that

$$\mathcal{F}_K \left[\sum_{i=1}^{K} w_i \delta_{x_i} \right] := \mathcal{F} \left[\sum_{i=1}^{K} w_i \frac{1_{B(x_i, r_i)}}{|B(x_i, r_i)|} \right]$$

$$= \sum_i |B(x_i, r_i)|^d U \left(\frac{w_i}{|B(x_i, r_i)|} \right) + \sum_{i=1}^{K} w_i V(x_i) \qquad (7.3)$$

$$+ \frac{1}{2} \sum_{i,j=1}^{K} w_i w_j W(x_i - x_j),$$

for suitably selected radii. Then the scheme is simply to apply a suitable adaptation of the JKO scheme (4.17) with time-step τ, that is

$$\mu_{n+1} := \underset{\mu \in \mathcal{A}_{K,w}}{\text{argmin}} \left(\frac{\mathfrak{W}_2(\mu, \mu_n)^2}{2} + \tau \mathcal{F}_K[\mu] \right).$$

7.3.2 The blob method

The blob approach aims to modify the energy so that solutions by particles are admissible. To preserve the gradient-flow structure we modify the free energy by changing the term $\int U(\rho)$. A popular idea introduced by [86] is

$$\mathcal{U}_\eta[\rho] = \int_{\mathbb{R}^d} F \circ (\eta * \rho) \, d\rho, \quad \text{where } F(s) = \frac{U(s)}{s}$$

where η is a smooth probably distribution typically close to a Dirac delta. Then, the equation becomes

$$\partial_t \rho = \text{div} \left(\rho \nabla \frac{\delta U_\eta}{\delta \rho} \right) \qquad \text{where } \frac{\delta U_\eta}{\delta \rho} = \eta * \left((F' \circ (\eta * \rho)) \rho \right) + F \circ (\eta * \rho).$$

This problem admits particle-type solutions.

Alternative choices of approximation of the free energy are available. See, e.g., [36, 88].

7.4 Finite elements

Finite element schemes rely on the weak formulation and look for solution in Sobolev spaces. They are extremely useful and accurate for problems which have an L^2-gradient flow structure (e.g., the p-Laplacian problem instead of (PME)). Nevertheless, authors have found success in this approach, e.g., for (KSE)-type problems [114–116] and later for more general (ADE) in [35]. For more details on finite elements for parabolic problems see [167].

8 Other related problems

8.1 Second-order problems

Our family of problems is of first order, since in the particle description we are specifying the velocities of the particles. There is a family of models which introduce the type of fields

but work in acceleration formulation

$$
\begin{cases}
\frac{dx_t^{(i)}}{dt} = v_t^{(i)} \\
\frac{dv_t^{(i)}}{dt} = -\nabla V(x_t^{(i)}) - \frac{1}{N} \sum_j \nabla W(x_t^{(i)} - x_t^{(j)}) + \sqrt{2\sigma} \, d\mathcal{B}_t^{(i)}
\end{cases}
$$

with independent Brownian motions $\mathcal{B}_t^{(i)}$. When $\sigma \neq 0$ we recover a diffusion term. This leads to the Vlasov–Poisson–Fokker–Planck equations. We point the reader to [49] and the references therein.

8.2 Cahn–Hilliard problem

Instead of a known potential, some authors have considered the Cahn–Hilliard type problem

$$
\frac{\partial \rho}{\partial t} = \operatorname{div}(\rho \nabla(-\Delta \rho)) - \chi \Delta \rho^m,
$$

where the diffusion is of fourth order, and the aggregation is modelled by a second order operator. This problem also has a Wasserstein gradient-flow structure. We point the reader, e.g., to [87] and the references therein.

8.3 The (fractional) thin-film problem

A family of related problem is the so-called thin-film problem and its fractional variant

$$
\frac{\partial \rho}{\partial t} = \operatorname{div}\left(u\nabla((-\Delta)^s u + V(x))\right).
$$

These are 2-Wasserstein gradient flows, where we replace $U(\rho)$ by $[u]_{H^s}^2$ and V is usually 0 or $|x|^2/2$. Usually $s \in [0, 1]$ where $s = 0$ corresponds in fact to U_2. We point the reader to [155] and the references therein.

8.4 Cross-diffusion systems

In our model, all particles are of the same "type". Hence, there is only need for one density. If one considers models several "types" of particles, we will arrive at a system of PDEs. A similar approach to ours can be done in the so-called *cross-diffusion systems*. They correspond to the suitable Wasserstein flow of several-variable energies like

$$
\mathcal{F}[\rho_1, \rho_2] = \int_{\mathbb{R}^d} U(\rho_1, \rho_2) + \frac{1}{2} \int_{\mathbb{R}^d} \rho_1 W_1 * \rho_1 + \frac{1}{2} \int_{\mathbb{R}^d} \rho_2 W_2 \rho_2 + \int_{\mathbb{R}^d} \rho_1 K * \rho_2.
$$

See, e.g., [34, 71, 91, 109] and the references therein.

8.5 Non-linear mobility

Several authors have studied the family of problems with so-called *non-linear mobility* m given by

$$
\frac{\partial \rho}{\partial t} = \operatorname{div}\left(m(\rho)\nabla \frac{\delta \mathcal{F}}{\delta \rho}\right). \tag{8.1}
$$

This family of problems correspond to a Wasserstein-type gradient flow, with a modified distance. The correct way to modify the Wasserstein distance is through an adapted Benamou–Brenier formula (recall (4.14)). The details can be found in [73, 111, 141]. It turns out this is only a valid distance if m is concave.

8.5.1 The Caffarelli–Vázquez problem with non-linear mobility

Some authors developed a theory for the (CVE) with non-linear mobility of the form $m(\rho) = \rho^{m-1}$ with $m > 0$. The motivation for came as a fractional generalisation of (PME) with non-local pressure. A theory of existence and properties of weak solutions was developed in [161, 162]. Recently, it was shown that in $d = 1$ one can analyse a suitable equation for the mass, for which there is a monotone and convergent scheme [104]. In fact, this numerical method for the mass uses an up-winding given for ρ as (7.1). This group in Trondheim also prove convergence in ρ variable in the so-called bounded Lipschitz distance (also called Fortet–Mourier distance), where we point the reader to [177, page 97]. This distance is defined as an extension of the duality formula (4.13) as

$$d_{bL}(\mu, \nu) := \sup \left\{ \int \psi \, d\mu - \int \psi \, d\nu : \text{for } \psi \text{ such that } \|\psi\|_{L^\infty} + \|\nabla\psi\|_{L^\infty} \leq 1 \right\},$$

where we point out the addition of the bound of ψ in L^∞.

8.5.2 Newtonian vortex with non-linear mobility

Some authors became interested in the case of non-linear mobility version of (NVE) given by

$$\begin{cases} \frac{\partial \rho}{\partial t} = \operatorname{div}(\rho^\alpha \nabla v) \\ -\Delta v = \rho, \end{cases} \tag{8.2}$$

The cases $\alpha \in (0, 1)$ (where m is concave) are studied in [62], where the authors show that for $\alpha < 1$ there exist nice self-similar solutions. The cases $\alpha > 1$ [63], where the authors find self-similar solutions do not have finite mass, and in fact the attracting solution is a family of characteristic functions. The technique in [62, 63] is to find solutions by generalised characteristics of the mass equation.

8.5.3 Saturation

The problem (ADE) with $m(\rho) = \rho\psi(\rho)$ with $\psi(1) = 0$ has recently gained the attention of some experts. This addition ensures the that we can work with solutions such that $0 \leq \rho \leq 1$ (in particular no formation of δ), and hence some authors use the term *saturation* to describe this choice.

The papers [110, 117] explore the case $U = 0$. The case $U \neq 0$ is, so far, only understood numerically. A finite-volume numerical scheme was constructed for this case [8]. Their numerical results suggest that steady-states will be of the form

$$\hat\rho = \min\{1, (U')^{-1}(h - V - W * \hat\rho)\}_+$$

for suitable h such that the total mass is preserved. However, there is currently no rigorous proof that this is the correct choice.

References

1. Amann, H.: Dynamic theory of quasilinear parabolic systems—III. Global existence. Math. Zeitschrift **205**(1), 331 (1990)
2. Ambrosio, L., Brué, E., Semola, D.: Lectures on Optimal Transport. Springer International Publishing, Berlin (2021)
3. Ambrosio, L., Gigli, N., Savare, G.: Gradient Flows. Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, pp 1–27 (2005)
4. Ambrosio, L., Mainini, E., Serfaty, S.: Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices. Annales de l'I.H.P.Analyse non linéaire **28**(2), 217–246 (2011)
5. Ambrosio, L., Serfaty, S.: A gradient flow approach to an evolution problem arising in superconductivity. Commun. Pure Appl. Math. **61**(11), 1495–1539 (2008)
6. Arnold, A., Markowich, P., Toscani, G., Unterreiter, A.: On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker–Planck type equations. Commun. Partial Differ. Equ. **26**(1–2), 43–100 (2001)
7. Bailo, R., Carrillo, J.A., Hu, J.: Fully discrete positivity-preserving and energy-dissipating schemes for aggregation-diffusion equations with a gradient-flow structure. Commun. Math. Sci. **18**(5), 1259–1303 (2020)
8. Bailo, R., Carrillo, J.A., Hu, J.: Bound-preserving finite-volume schemes for systems of continuity equations with saturation. arXiv:2110.08186
9. Bailo, R., Carrillo, J.A., Murakawa, H., Schmidtchen, M.: Convergence of a fully discrete and energy-dissipating finite-volume scheme for aggregation-diffusion equations. Math. Models Methods Appl. Sci. **30**(13), 2487–2522 (2020)
10. Balagué, D., Carrillo, J.A., Laurent, T., Raoul, G.: Dimensionality of local minimizers of the interaction energy. Arch. Ration. Mech. Anal. **209**(3), 1055–1088 (2013)
11. Balagué, D., Carrillo, J.A., Laurent, T., Raoul, G.: Nonlocal interactions by repulsive-attractive potentials: radial ins/stability. Phys. D Nonlinear Phenom. **260**, 5–25 (2013)
12. Balagué, D., Carrillo, J.A., Yao, Y.: Confinement for repulsive-attractive kernels. Discrete Contin. Dyn. Syst. B **19**(5), 1227–1248 (2014)
13. Bedrossian, J.: Global minimizers for free energies of subcritical aggregation equations with degenerate diffusion. Appl. Math. Lett. **24**(11), 1927–1932 (2011)
14. Bedrossian, J.: Intermediate asymptotics for critical and supercritical aggregation equations and Patlak–Keller–Segel models. Commun. Math. Sci. **9**(4), 1143–1161 (2011)
15. Bedrossian, J., Rodríguez, N., Bertozzi, A.L.: Local and global well-posedness for aggregation equations and Patlak–Keller–Segel models with degenerate diffusion. Nonlinearity **24**(6), 1683–1714 (2011)

16. Benedetto, D., Caglioti, E., Carrillo, J.A., Pulvirenti, M.: A non-Maxwellian steady distribution for one-dimensional granular media. J. Stat. Phys. **91**(5/6), 979–990 (1998)
17. Benilan, P., Crandall, M.: The continuous dependence on Φ of solutions of $u_t - \Delta\phi(u) = 0$. Ind. Univ. Math. J. **30**(2), 161–177 (1981)
18. Bertozzi, A.L., Brandman, J.: Finite-time blow-up of L∞-weak solutions of an aggregation equation. Commun. Math. Sci. **8**(1), 45–65 (2010)
19. Bertozzi, A.L., Carrillo, J.A., Laurent, T.: Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. Nonlinearity **22**(3), 683–710 (2009)
20. Bertozzi, A.L., Garnett, J.B., Laurent, T.: Characterization of radially symmetric finite time blowup in multidimensional aggregation equations. SIAM J. Math. Anal. **44**(2), 651–681 (2012)
21. Bertozzi, A.L., Laurent, T., Léger, F.: Aggregation via Newtonian potential and aggregation patches. Math. Models Methods Appl. Sci. **22** (2012)
22. Bertozzi, A.L., Laurent, T., Rosado, J.: Lp theory for the multidimensional aggregation equation. Commun. Pure Appl. Math. **64**(1), 45–83 (2011)
23. Bian, S.: The Aggregation-Diffusion Equation with Energy Critical Exponent. arXiv:2302.09490
24. Biler, P.: Existence and asymptotics of solutions for a parabolic-elliptic system with nonlinear no-flux boundary conditions. Nonlinear Anal. Theory Methods Appl. **19**(12), 1121–1136 (1992)
25. Biler, P., Imbert, C., Karch, G.: The nonlocal porous medium equation: Barenblatt profiles and other weak solutions. Arch. Ration. Mech. Anal. **215**(2), 497–529 (2015)
26. Blanchet, A., Bonforte, M., Dolbeault, J., Grillo, G., Vázquez, J.L.: Asymptotics of the fast diffusion equation via entropy estimates. Arch. Ration. Mech. Anal. **191**(2), 347–385 (2009)
27. Blanchet, A., Carrillo, J.A., Laurençot, P.: Critical mass for a Patlak–Keller–Segel model with degenerate diffusion in higher dimensions. Calc. Var. Partial Differ. Equ. **35**(2), 133–168 (2009)
28. Blanchet, A., Dolbeault, J., Perthame, B.: Two-dimensional Keller–Segel model: optimal critical mass and qualitative properties of the solutions. Electron. J. Differ. Equ. **44**, 32 (2006)
29. Bonforte, M., Dolbeault, J., Grillo, G., Vázquez, J.L.: Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities. Proc. Natl. Acad. Sci. **107**(38), 16459–16464 (2010)
30. Bonforte, M., Figalli, A.: Total variation flow and sign fast diffusion in one dimension. J. Differ. Equ. **252**(8), 4455–4480 (2012)
31. Brändle, C., Vázquez, J.L.: Viscosity solutions for quasilinear degenerate parabolic equations of porous medium type. Ind. Univ. Math. J. **54**(3), 817–860 (2005)
32. Brezis, H., Friedman, A.: Nonlinear parabolic equations involving measures as initial data. J. Math. Pures Appl. **62**, 73–97 (1983)
33. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, New York (2010)
34. Burger, M., Carrillo, J., Pietschmann, J.-F., Schmidtchen, M.: Segregation effects and gap formation in cross-diffusion models. Interfaces Free Bound. **22**(2), 175–203 (2020)
35. Burger, M., Carrillo, J.A., Wolfram, M.-T.: A mixed finite element method for nonlinear diffusion equations. Kinet. Relat. Models **3**(1), 59–83 (2010)
36. Burger, M., Esposito, A.: Porous medium equation and cross-diffusion systems as limit of nonlocal interaction. Nonlinear Anal. **235**, 113347 (2023)
37. Caffarelli, L.A., Soria, F., Vázquez, J.L.: Regularity of solutions of the fractional porous medium flow. J. Eur. Math. Soc. (JEMS) **15**(5), 1701–1746 (2013)
38. Caffarelli, L.A., Vazquez, J.L.: Nonlinear porous medium flow with fractional potential pressure. Arch. Ration. Mech. Anal. **202**(2), 537–565 (2011)
39. Caffarelli, L.A., Vázquez, J.L.: Regularity of solutions of the fractional porous medium flow with exponent 1/2. Algebra i Analiz **27**(3), 125–156 (2015)
40. Caffarelli, L.A., Vázquez, J.L.: Viscosity solutions for the porous medium equation. In: Proc. Symp. Pure Math., pp. 13–26 (1999)
41. Caffarelli, L.A., Vázquez, J.L.: Asymptotic behaviour of a porous medium equation with fractional diffusion. Discrete Contin. Dyn. Syst. **29**(4), 1393–1404 (2011)
42. Caillet, T., Santambrogio, F.: Doubly Nonlinear Diffusive PDEs: New Existence Results via Generalized Wasserstein Gradient Flows. (2024). arXiv:2402.02882
43. Calvez, V., Carrillo, J.A., Hoffmann, F.: Equilibria of homogeneous functionals in the fair-competition regime. Nonlinear Anal. Adv. React.-Cross-Diffus. Syst. **159**, 85–128 (2017)
44. Calvez, V., Carrillo, J.A.: Volume effects in the Keller–Segel model: energy estimates preventing blow-up. J. Math. Pures Appl. **86**(2), 155–175 (2006)
45. Cañizo, J.A., Carrillo, J.A., Patacchini, F.S.: Existence of compactly supported global minimisers for the interaction energy. Arch. Ration. Mech. Anal. **217**(3), 1197–1217 (2015)

46. Cañizo, J.A., Carrillo, J.A., Schonbek, M.E.: Decay rates for a class of diffusive-dominated interaction equations. J. Math. Anal. Appl. **389**(1), 541–557 (2012)
47. Carrillo, J.: Entropy solutions for nonlinear degenerate problems. Arch. Ration. Mech. Anal. **147**(4), 269–361 (1999)
48. Carrillo, J.A., Choi, Y.-P., Ha, S.-Y., Kang, M.-J., Kim, Y.: Contractivity of transport distances for the kinetic Kuramoto equation. J. Stat. Phys. **156**(2), 395–415 (2014)
49. Carrillo, J.A., Choi, Y.-P., Salem, S.: Propagation of chaos for the Vlasov–Poisson–Fokker–Planck equation with a polynomial cut-off. Commun. Contemp. Math. **21**(04), 1850039 (2019)
50. Carrillo, J.A., Craig, K., Wang, L., Wei, C.: Primal dual methods for wasserstein gradient flows. Found. Comput. Math. **22**(2), 389–443 (2022)
51. Carrillo, J.A., Delgadino, M.G.: Free Energies and the Reversed HLS Inequality (2018)
52. Carrillo, J.A., Delgadino, M.G., Frank, R.L., Lewin, M.: Fast Diffusion Leads to Partial Mass Concentration in Keller–Segel Type Stationary Solutions (2020). arXiv:2012.08586
53. Carrillo, J.A., Delgadino, M.G., Mellet, A.: Regularity of local minimizers of the interaction energy via obstacle problems. Commun. Math. Phys. **343**(3), 747–781 (2016)
54. Carrillo, J.A., Delgadino, M.G., Patacchini, F.S.: Existence of ground states for aggregation-diffusion equations. Anal. Appl. **17**(3), 393–423 (2019)
55. Carrillo, J.A., Delgadino, M.G., Dolbeault, J., Frank, R.L., Hoffmann, F.: Reverse Hardy–Littlewood–Sobolev inequalities. J. Math. Pures et Appl. **132**, 133–165 (2019)
56. Carrillo, J.A., DiFrancesco, M., Figalli, A., Laurent, T., Slepčev, D.: Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. Duke Math. J. **156**(2), 229–271 (2011)
57. Carrillo, J.A., Fernández-Jiménez, A., Gómez-Castro, D.: Partial Mass Concentration for Fast-Diffusions with Non-Local Aggregation Terms. arXiv:2304.04582
58. Carrillo, J.A., Fjordholm, U.S., Solem, S.: A second-order numerical method for the aggregation equations. Math. Comput. **90**(327), 103–39 (2021)
59. Carrillo, J.A., Francesco, M.D., Toscani, G.: Intermediate asymptotics beyond homogeneity and self-similarity: long time behavior for $u_t = \Delta\phi(u)$. Arch. Ration. Mech. Anal. **180**(1), 127–149 (2006)
60. Carrillo, J.A., Gómez-Castro, D., Vázquez, J.L.: Infinite-time concentration in aggregation-diffusion equations with a given potential. J. Math. Pures Appl. **157**, 346–398 (2022)
61. Carrillo, J.A., Gómez-Castro, D., Yao, Y., Zeng, C.: Asymptotic simplification of aggregation-diffusion equations towards the heat kernel. Arch. Ration. Mech. Anal. **247**, 11 (2023)
62. Carrillo, J.A., Gómez-Castro, D., Vázquez, J.L.: A fast regularisation of a Newtonian vortex equation. In: Annales de l'Institut Henri Poincaré C, Analyse non linéaire (2022)
63. Carrillo, J.A., Gómez-Castro, D., Vázquez, J.L.: Vortex formation for a non-local interaction model with Newtonian repulsion and superlinear mobility. Adv. Nonlinear Anal. **11**(1), 937–967 (2022)
64. Carrillo, J.A., Gvalani, R.S., Pavliotis, G.A., Schlichting, A.: Long-time behaviour and phase transitions for the Mckean–Vlasov equation on the torus. Arch. Ration. Mech. Anal. **235**(1), 635–690 (2020)
65. Carrillo, J.A., Gvalani, R.S.: Phase transitions for nonlinear nonlocal aggregation-diffusion equations. Commun. Math. Phys. **382**(1), 485–545 (2021)
66. Carrillo, J.A., Hittmeir, S., Volzone, B., Yao, Y.: Nonlinear aggregation-diffusion equations: radial symmetry and long time asymptotics. Invent. Math. **218**(3), 889–977 (2019)
67. Carrillo, J.A., Hoffmann, F., Mainini, E., Volzone, B.: Ground states in the diffusion-dominated regime. Calc. Var. Partial Differ. Equ. **57**(5), 1–28 (2018)
68. Carrillo, J.A., Hopf, K., Rodrigo, J.L.: On the singularity formation and relaxation to equilibrium in 1D Fokker-Planck model with superlinear drift. Adv. Math. (N. Y). **360**, 106883 (2020)
69. Cañillo, J.A., Huang, Y., Patacchini, F.S., Wolansky, G.: Numerical study of a particle method for gradient flows. Kinet. Relat. Models **10**(3), 613–641 (2017)
70. Carrillo, J.A., Huang, Y.: Explicit equilibrium solutions for the aggregation equation with power-law potentials. Kinet. Relat. Models **10**(1), 171–192 (2017)
71. Carrillo, J.A., Huang, Y., Schmidtchen, M.: Zoology of a nonlocal cross-diffusion model for two species. SIAM J. Appl. Math. **78**(2), 1078–1104 (2018)
72. Carrillo, J.A., Jüngel, A., Markowich, P.A., Toscani, G., Unterreiter, A.: Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. Mon. Math. **133**(1), 1–82 (2001)
73. Carrillo, J.A., Lisini, S., Savaré, G., Slepčev, D.: Nonlinear mobility continuity equations and generalized displacement convexity. J. Funct. Anal. **258**(4), 1273–1309 (2010)
74. Carrillo, J.A., Mateu, J., Mora, M.G., Rondi, L., Scardia, L., Verdera, J.: The ellipse law: Kirchhoff meets dislocations. Commun. Math. Phys. **373**(2), 507–524 (2020)
75. Carrillo, J.A., Mateu, J., Mora, M.G., Rondi, L., Scardia, L., Verdera, J.: The equilibrium measure for an anisotropic nonlocal energy. Calc. Var. Partial Differ. Equ. **60**(3), 109 (2021)

76. Carrillo, J.A., McCann, R.J., Villani, C.: Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. Rev. Mat. Iberoam. **19**(3), 971–1018 (2003)

77. Carrillo, J.A., McCann, R.J., Villani, C.: Contractions in the 2-Wasserstein length space and thermalization of granular media. Arch. Ration. Mech. Anal. **179**(2), 217–263 (2006)

78. Carrillo, J.A., Patacchini, F.S., Sternberg, P., Wolansky, G.: Convergence of a particle method for diffusive gradient flows in one dimension. SIAM J. Math. Anal. **48**(6), 3708–3741 (2016)

79. Carrillo, J.A., Shu, R.: From radial symmetry to fractal behavior of aggregation equilibria for repulsive-attractive potentials. Calc. Var. Partial Differ. Equ. **62**(1), 28 (2022)

80. Carrillo, J.A., Shu, R.: Global minimizers of a large class of anisotropic attractive-repulsive interaction energies in 2D. Commun. Pure Appl. Math. **77**(2), 1353–1404 (2024)

81. Carrillo, J.A., Shu, R.: Minimizers of 3D anisotropic interaction energies. arXiv:2206.14054

82. Carrillo, J.A., Slepčev, D.: Example of a displacement convex functional of first order. Calc. Var. Partial Differ. Equ. **36**(4), 547–564 (2009)

83. Carrillo, J.A., Toscani, G.: Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. Ind. Univ. Math. J. **49**(1), 113–142 (2000)

84. Carrillo, J.A., Castorina, D., Volzone, B.: Ground states for diffusion dominated free energies with logarithmic interaction. SIAM J. Math. Anal. **47**(1), 1–25 (2015)

85. Carrillo, J.A., Chipot, M., Huang, Y.: On global minimizers of repulsive-attractive power-law interaction energies. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **372**(2028), 20130399 (2014)

86. Carrillo, J.A., Craig, K., Patacchini, F.S.: A blob method for diffusion. Calc. Var. Partial Differ. Equ. **58**(2), 53 (2019)

87. Carrillo, J.A., Esposito, A., Falcó, C., Fernández-Jiménez, A.: Competing Effects in Fourth-Order Aggregation-Diffusion Equations. arXiv:2307.14706

88. Carrillo, J.A., Esposito, A., Wu, J.S.-H.: Nonlocal Approximation of Nonlinear Diffusion Equations. arXiv:2302.08248

89. Chan, H., González, M.D.M., Huang, Y., Mainini, E., Volzone, B.: Uniqueness of entire ground states for the fractional plasma problem. Calc. Var. Partial Differ. Equ. **59**(6), 195 (2020)

90. Chapman, S.J., Rubinstein, J., Schatzman, M.: A mean-field model of superconducting vortices. Eur. J. Appl. Math. **7**(2), 97–111 (1996)

91. Chen, L., Daus, E.S., Jüngel, A.: Rigorous mean-field limit and cross-diffusion. Z. Angew. Math. Phys. **70**(4), 122 (2019)

92. Chen, W., Li, C., Ou, B.: Classification of solutions for an integral equation. Commun. Pure Appl. Math. **59**(3), 330–343 (2006)

93. Chernoff, H.: A note on an inequality involving the normal distribution. Ann. Probab. **9**(3), 533–535 (1981)

94. Chizat, L., Colombo, M., Fernández-Real, X., Figalli, A.: Infinite-Width Limit of Deep Linear Neural Networks (2022) arXiv:2211.16980

95. Collot, C., Ghoul, T.-E., Masmoudi, N., Nguyen, V.T.: Refined description and stability for singular solutions of the 2D Keller–Segel system. Commun. Pure Appl. Math. **75**(7), 1419–1516 (2022)

96. Crandall, M.G., Liggett, T.M.: Generation of semi-groups of nonlinear transformations on general Banach spaces. Am. J. Math. **93**(2), 265 (1971)

97. Crandall, M.G., Ishii, H., Lions, P.-L.: User's guide to viscosity solutions of second order partial differential equation. Bull. Am. Math. Soc. **27**(1), 1–67 (1992)

98. Davies, C., Lim, T., McCann, R.J.: Classifying minimum energy states for interacting particles: regular simplices. Commun. Math. Phys. **399**(2), 577–598 (2023)

99. Davies, C., Lim, T., McCann, R.J.: Classifying Minimum Energy States for Interacting Particles: Spherical Shells. arXiv:2107.11718

100. Davila, J., del Pino, M., Dolbeault, J., Musso, M., Wei, J.: Existence and Stability of Infinite Time Blow-up in the Keller–Segel System. arXiv:1911.12417

101. De Giorgi, E.: New problems on minimizing movements. In: Boundary Value Problems for PDE and Applications, vol. 29, pp. 81–98 (1993) **(Cited By 138)**

102. del Pino, M.: Bubbling Blow-Up in Critical Parabolic Problems. In: Carrillo, J.A., del Pino, M., Figalli, A., Mingione, G., Vázquez, J.L., Bonforte, M., Grillo, G. (eds.) Nonlocal and Nonlinear Diffusions and Interactions: New Methods and Directions, vol. 2186, pp. 73–116. Springer International Publishing, Cham (2017)

103. del Pino, M., Dolbeault, J.: Best constants for Gagliardo–Nirenberg inequalities and applications to nonlinear diffusions. J. Math. Pures Appl. **81**(9), 847–875 (2002)

104. del Teso, F., Jakobsen, E.R.: A Convergent Finite Difference-Quadrature Scheme for the Porous Medium Equation with Nonlocal Pressure. arXiv:2303.05168

105. Delarue, F., Lagoutière, F., Vauchelet, N.: Convergence analysis of upwind type schemes for the aggregation equation with pointy potential. Ann. Henri Lebesgue **3**, 217–260 (2020)
106. Delgadino, M.G., Yan, X., Yao, Y.: Uniqueness and nonuniqueness of steady states of aggregation-diffusion equations. Commun. Pure Appl. Math. **75**(1), 3–59 (2022)
107. Demengel, F., Temam, R.: Convex functions of a measure and applications. Ind. Univ. Math. J. **33**(5), 673–709 (1986)
108. DiBenedetto, E.: Degenerate Parabolic Equations. Universitext. Springer, New York (1993)
109. Di Francesco, M., Esposito, A., Fagioli, S.: Nonlinear degenerate cross-diffusion systems with nonlocal interaction. Nonlinear Anal. Theory Methods Appl. **169**, 94–117 (2018)
110. Di Francesco, M., Fagioli, S., Radici, E.: Deterministic particle approximation for nonlocal transport equations with nonlinear mobility. J. Differ. Equ. **266**(5), 2830–2868 (2019)
111. Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. Calc. Var. Partial Differ. Equ. **34**(2), 193–231 (2009)
112. Dolbeault, J., Perthame, B.: Optimal critical mass in the two dimensional Keller–Segel model. C.R. Math. **339**(9), 611–616 (2004)
113. Dolbeault, J., Toscani, G.: Stability results for logarithmic Sobolev and Gagliardo–Nirenberg inequalities. Int. Math. Res. Not. **2016**(2), rnv131 (2015)
114. Epshteyn, Y.: Discontinuous Galerkin methods for the chemotaxis and haptotaxis models. J. Comput. Appl. Math. **224**(1), 168–181 (2009)
115. Epshteyn, Y., Izmirlioglu, A.: Fully discrete analysis of a discontinuous finite element method for the Keller–Segel chemotaxis model. J. Sci. Comput. **40**(1–3), 211–256 (2009)
116. Epshteyn, Y., Kurganov, A.: New interior penalty discontinuous Galerkin methods for the Keller–Segel chemotaxis model. SIAM J. Numer. Anal. **47**(1), 386–408 (2009)
117. Fagioli, S., Tse, O.: On gradient flow and entropy solutions for nonlocal transport equations with nonlinear mobility. Nonlinear Anal. **221**, 112904 (2022)
118. Fernández-Real, X., Figalli, A.: The Continuous Formulation of Shallow Neural Networks as Wasserstein-Type Gradient Flows, Analysis at Large: Dedicated to the Life and Work of Jean Bourgain. Springer, pp. 29–57 (2022)
119. Frank, R.L.: Minimizers for a one-dimensional interaction energy. Nonlinear Anal. **216**, 112691 (2022)
120. Frank, R.L., Matzke, R.W.: Minimizers for an Aggregation Model with Attractive-Repulsive Interaction. arXiv:2307.13769
121. Friedman, A., Kamin, S.: The asymptotic behavior of gas in an N-dimensional porous medium. Trans. Am. Math. Soc. **262**(2), 551–563 (1980)
122. Gentil, I., Imbert, C.: The Lévy–Fokker–Planck equation: Φ-entropies and convergence to equilibrium. Asymptot. Anal. **59**(3–4), 125–138 (2008)
123. Giga, Y., Mizoguchi, N., Senba, T.: Asymptotic behavior of type I blowup solutions to a parabolic-elliptic system of drift-diffusion type. Arch. Ration. Mech. Anal. **201**(2), 549–573 (2011)
124. Gross, L.: Logarithmic Sobolev. Am. J. Math. **97**(4), 1061–1083 (1975)
125. Herrero, M., Medina, E., Velázquez, J.: Self-similar blow-up for a reaction-diffusion system. J. Comput. Appl. Math. **97**(1–2), 99–119 (1998)
126. Herrero, M.A., Velázquez, J.J.L.: Chemotactic collapse for the Keller–Segel model. J. Math. Biol. **35**(2), 177–194 (1996)
127. Huang, C., Svobodny, T.: Evolution of mixed-state regions in type-II superconductors. SIAM J. Math. Anal. **29**(4), 1002–1021 (1998)
128. Jabin, P.-E., Wang, Z.: Mean field limit for stochastic particle systems. In: Active Particles, Modeling and Simulation in Science, Engineering and Technology, ed. by N. Bellomo, P. Degond, and E. Tadmor, vol. 1, pp. 379–402. Springer International Publishing, Cham (2017)
129. Jäger, W., Luckhaus, S.: On explosions of solutions to a system of partial differential equations modelling chemotaxis. Trans. Am. Math. Soc. **329**(2), 819–824 (1992)
130. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker–Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998)
131. Kang, K., Kim, H.K., Lim, T., Seo, G.: Uniqueness and characterization of local minimizers for the interaction energy with mildly repulsive potentials. Calc. Var. Partial Differ. Equ. **60**(1), 15 (2021)
132. Karlsen, K.H., Risebro, N.H.: On the uniqueness and stability of entropy solutions of nonlinear degenerate parabolic equations with rough coefficients. Discrete Contin. Dyn. Syst. A **9**(5), 1081–1104 (2003)
133. Keller, E.F., Segel, L.A.: Initiation of slime mold aggregation viewed as an instability. J. Theor. Biol. **26**(3), 399–415 (1970)
134. Kim, I., Yao, Y.: The Patlak–Keller–Segel model and its variations: properties of solutions via maximum principle. SIAM J. Math. Anal. **44**(2), 568–602 (2012)

135. Kuramoto, Y.: Self-Entrainment of a Population of Coupled Non-Linear Oscillators. In: International Symposium on Mathematical Problems in Theoretical Physics (Kyoto Univ., Kyoto, 1975), Lecture Notes in Phys., vol. 39, pp. 420–422. Springer, Berlin (1975)
136. Ladyzhenskaia, O.A., Solonnikov, V.A., Ural'tseva, N.N.: Linear and quasi-linear equations of parabolic type, vol. 23. American Mathematical Soc. (1968)
137. Lagoutière, F., Santambrogio, F., Tien, S.T.: Vanishing Viscosity Limit for Aggregation-Diffusion Equations
138. Laurent, T.: Local and global existence for an aggregation equation. Commun. Partial Differ. Equ. **32**(12), 1941–1964 (2007)
139. Lin, F., Zhang, P.: On the hydrodynamic limit of Ginzburg–Landau vortices. Discrete Contin. Dyn. Syst. **6**(1), 121–142 (1999)
140. Lisini, S., Mainini, E., Segatti, A.: A gradient flow approach to the porous medium equation with fractional pressure. Arch. Ration. Mech. Anal. **227**(2), 567–606 (2018)
141. Lisini, S., Marigonda, A.: On a class of modified Wasserstein distances induced by concave mobility functions defined on bounded intervals. Manuscr. Math. **133**(1–2), 197–224 (2010)
142. Luckhaus, S., Sugiyama, Y.: Asymptotic profile with the optimal convergence rate for a parabolic equation of chemotaxis in super-critical cases. Ind. Univ. Math. J. **56**(3), 1279–1297 (2007)
143. Masmoudi, N., Zhang, P.: Global solutions to vortex density equations arising from sup-conductivity. Ann. l'IHP Analyse non linéaire **22**(4), 441–458 (2005)
144. Mateu, J., Mora, M.G., Rondi, L., Scardia, L., Verdera, J.: Explicit minimisers for anisotropic Coulomb energies in 3D. Adv. Math. **434**, 109333 (2023)
145. McCann, R.J.: A convexity principle for interacting gases. Adv. Math. **128**(1), 153–179 (1997)
146. Nanjundiah, V.: Chemotaxis, signal relaying and aggregation morphology. J. Theor. Biol. **42**(1), 63–105 (1973)
147. Otto, F.: Double degenerate diffusion equations as steepest descent, pp. 1–43 (1996) **(unpublished)**
148. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Commun. Partial Differ. Equ. **26**(1–2), 101–174 (2001)
149. Patlak, C.S.: Random walk with persistence and external bias. Bull. Math. Biophys. **15**(3), 311–338 (1953)
150. Peyré, G., Cuturi, M.: Computational Optimal Transport (2020). arXiv:1803.00567
151. Raphaël, P., Schweyer, R.: On the stability of critical chemotactic aggregation. Math. Ann. **359**, 267–377 (2014)
152. Reda, F.A., Maury, B.: 15. Interpretation of finite volume discretization schemes for the Fokker–Planck equation as gradient flows for the discrete wasserstein distance. In: Bergounioux, M., Oudet, É., Rumpf, M., Carlier, G., Champion, T., Santambrogio, F. (eds.) Topological Optimization and Optimal Transport, pp. 400–416. De Gruyter (2017)
153. Santambrogio, F.: Optimal Transport for Applied Mathematicians. Progress in Nonlinear Differential Equations and Their Applications, vol. 87, pp. xxvii+353. Birkhäuser/Springer, Cham (2015)
154. Schätzle, R., Styles, V.: Analysis of a mean field model of superconducting vortices. Eur. J. Appl. Math. **10**(4), 319–352 (1999)
155. Segatti, A., Vázquez, J.L.: On a fractional thin film equation. Adv. Nonlinear Anal. **9**(1), 1516–1558 (2020)
156. Senba, T.: Blowup behavior of radial solutions to Jaeger–Luckhaus system in high dimensional domains. Funkcialaj Ekvacioj **48**(2), 247–271 (2005)
157. Serfaty, S., Vázquez, J.L.: A mean field equation as limit of nonlinear diffusions with fractional Laplacian operators. Calc. Var. Partial Differ. Equ. **49**(3–4), 1091–1120 (2014)
158. Silvestre, L.: Regularity of the obstacle problem for a fractional power of the Laplace operator. Commun. Pure Appl. Math. **60**(1), 67–112 (2007)
159. Simione, R., Slepčev, D., Topaloglu, I.: Existence of ground states of nonlocal-interaction energies. J. Stat. Phys. **159**(4), 972–986 (2015)
160. Souplet, P., Winkler, M.: Blow-up profiles for the parabolic-elliptic Keller–Segel system in dimensions $n \geq 3$. Commun. Math. Phys. **367**(2), 665–681 (2019)
161. Stan, D., del Teso, F., Vázquez, J.L.: Finite and infinite speed of propagation for porous medium equations with nonlocal pressure. J. Differ. Equ. **260**(2), 1154–1199 (2016)
162. Stan, D., del Teso, F., Vázquez, J.L.: Existence of weak solutions for a general porous medium equation with nonlocal pressure. Arch. Ration. Mech. Anal. **233**(1), 451–496 (2019)
163. Sugiyama, Y.: Application of the Best Constant of the Sobolev Inequality to Degenerate Keller–Segel Models. Advances in Differential Equations, vol. 12, no 2 (2007)
164. Sugiyama, Y.: Time Global Existence and Asymptotic Behavior of Solutions to Degenerate Quasi-Linear Parabolic Systems of Chemotaxis. Differential and Integral Equations, vol. 20, no 2 (2007)

165. Sugiyama, Y., Kunii, H.: Global existence and decay properties for a degenerate Keller–Segel model with a power factor in drift term. J. Differ. Equ. **227**(1), 333–364 (2006)
166. Talenti, G.: The art of rearranging. Milan J. Math. **84**(1), 105–157 (2016)
167. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems. 2nd ed. Springer Series in Computational Mathematics, vol. 25. Springer, Berlin (2006)
168. Toscani, G.: Entropy production and the rate of convergence to equilibrium for the Fokker–Planck equation. Q. Appl. Math. **57**(3), 521–541 (1999)
169. Urbano, J.M.: The Method of Intrinsic Scaling. In: Morel, J., Takens, F., Teissier, B. (eds.) Lecture Notes in Mathematics, vol. 1930. Springer, Berlin (2008)
170. Vázquez, J.L.: Smoothing and Decay Estimates for Nonlinear Diffusion Equations. Oxford University Press, Oxford (2006)
171. Vázquez, J.L.: The Porous Medium Equation. Oxford University Press, Oxford (2006)
172. Vázquez, J.L.: Asymptotic Behaviour Methods for the Heat Equation. Convergence to the Gaussian (2017). arXiv:1706.10034
173. Velázquez, J.J.: Stability of some mechanisms of chemotactic aggregation. SIAM J. Appl. Math. **62**(5), 1581–1633 (2002)
174. Velázquez, J.J.: Point dynamics in a singular limit of the Keller–Segel model 1: motion of the concentration regions. SIAM J. Appl. Math. **64**(4), 1198–1223 (2004)
175. Velázquez, J.J.: Point dynamics in a singular limit of the Keller–Segel model 2: formation of the concentration regions. SIAM J. Appl. Math. **64**(4), 1224–1248 (2004)
176. Villani, C.: Mathematics of granular materials. J. Stat. Phys. **124**(2–4), 781–822 (2006)
177. Villani, C.: Optimal Transport. Old and New. Grundlehren Der Mathematischen Wissenschaften, vol. 338, p. 973. Springer, Berlin (2009)
178. Weissler, F.B.: Logarithmic Sobolev inequalities for the heat-diffusion semigroup. Trans. Am. Math. Soc. **237**, 255 (1978)
179. Yao, Y., Bertozzi, A.L.: Blow-up dynamics for the aggregation equation with degenerate diffusion. Phys. D **260**, 77–89 (2013)
180. Yin, Z.: On the global existence of solutions to quasilinear parabolic equations with homogeneous Neumann boundary conditions. Glas. Math. J. **47**(2), 237–248 (2005)