



# Nonlocal approximation of nonlinear diffusion equations

José Antonio Carrillo<sup>1</sup> · Antonio Esposito<sup>1</sup> · Jeremy Sheung-Him Wu<sup>2</sup>

Received: 8 October 2023 / Accepted: 6 March 2024 / Published online: 13 April 2024  
© The Author(s) 2024

## Abstract

We show that degenerate nonlinear diffusion equations can be asymptotically obtained as a limit from a class of nonlocal partial differential equations. The nonlocal equations are obtained as gradient flows of interaction-like energies approximating the internal energy. We construct weak solutions as the limit of a (sub)sequence of weak measure solutions by using the Jordan-Kinderlehrer-Otto scheme from the context of 2-Wasserstein gradient flows. Our strategy allows to cover the porous medium equation, for the general slow diffusion case, extending previous results in the literature. As a byproduct of our analysis, we provide a qualitative particle approximation.

**Mathematics Subject Classification** 35A15 · 35Q70 · 35D30

## 1 Introduction

Nonlinear diffusion equations are ubiquitous in several real world applications. They were introduced to analyse gas expansion in a porous medium, groundwater infiltration, and heat conduction in plasmas, to name a few applications in physics. These applications drove the first rigorous mathematical results by Zel'dovich and Kompaneets in [62] and Barenblatt in [2] regarding important particular weak solutions of nonlinear diffusion equations with homogeneous nonlinearity. The general filtration equation was then first developed in [38]. The use of these equations in oil recovery software is extensive nowadays. Another source of applications of this family of equations arises from population models in mathematical biology: ecological models [4, 6, 56] derived from probabilistic interpretations [34, 44],

---

Communicated by Andrea Mondino.

---

✉ Antonio Esposito  
antonio.esposito@maths.ox.ac.uk

José Antonio Carrillo  
carrillo@maths.ox.ac.uk

Jeremy Sheung-Him Wu  
jeremywu@math.ucla.edu

<sup>1</sup> Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK

<sup>2</sup> Mathematical Sciences Building, University of California, Los Angeles, CA 90095, USA

volume effect in Keller-Segel type models [8, 35, 48], volume exclusion in cell-cell adhesion models [11, 22], and many others.

Although a rigorous mathematical theory has been extensively provided over the years [17, 47, 58], there are particular aspects of renewed interest in view of novel applications as well as advances in mathematics. For instance, their derivation from interacting particles, with a distinction between deterministic and stochastic methods, has recently attracted attention for its implications in derivation of models in mathematical biology [22] and data science [27]. We take advantage of the gradient flow structure of nonlinear diffusions [47] to connect with nonlocal interaction equations. In fact, we rigorously derive particle approximations of nonlinear diffusions from these variational considerations by approximating their energy functional completing the approach started in [10].

For ease of presentation, let us focus on more standard diffusion equations. Let  $m \geq 1$  and consider the equation

$$\partial_t \rho = \Delta \rho^m,$$

which is better known as the heat equation for  $m = 1$ , or the porous medium equation (PME) in the case  $m > 1$ . A comprehensive study of the above PDE can be found, e.g., in the book of Vázquez, [58]. Owing to the advances in optimal transport theory, [54, 60, 61], starting from the seminal works of Jordan, Kinderlehrer, and Otto, [37, 47], such diffusion equations are known to be 2-Wasserstein gradient flows for a specific choice of the energy functional. More precisely, the previous equation can be written as

$$\begin{cases} \partial_t \rho + \nabla \cdot (\rho v) = 0, \\ v = -\nabla \frac{\delta \mathcal{H}_m}{\delta \rho}, \end{cases} \tag{1.1}$$

being  $\frac{\delta \mathcal{H}_m}{\delta \rho}$  the first variation of the energy functional

$$\mathcal{H}_m[\rho] = \begin{cases} \int_{\mathbb{R}^d} \rho(x) \log \rho(x) dx & m = 1 \\ \frac{1}{m-1} \int_{\mathbb{R}^d} \rho^m(x) dx & m > 1 \end{cases}. \tag{1.2}$$

In [37] the equation of interest was the linear Fokker–Planck equation, while Otto focused on the porous medium equation in [47]. Afterwards, a 2-Wasserstein gradient flow approach has been extended to other PDEs, in particular those modelling nonlocal interaction, [1, 14, 19, 20]. The latter equation is of the form (1.1) with  $v = -\nabla \frac{\delta \mathcal{W}}{\delta \rho}$  and

$$\mathcal{W}[\rho] = \frac{1}{2} \int_{\mathbb{R}^d} W * \rho(x) d\rho(x). \tag{1.3}$$

Recent works in the literature show a rigorous and fascinating connection between the two energies above for  $m > 1$  in (1.2) and the corresponding dynamics, by means of gradient flow techniques, c.f. [7, 10]. More precisely, exploiting the so-called *blob method* developed in [26], one can notice already at a formal level that an appropriate regularisation of  $\mathcal{H}_m$  transforms a diffusion equation (which is local) into an interaction PDE (which is nonlocal) by choosing a delocalising kernel. For simplicity, let  $m = 2$  and consider a standard family of non-negative radial mollifiers  $V_\varepsilon(x) = \varepsilon^{-d} V_1(x/\varepsilon)$  for  $\varepsilon > 0$  on  $\mathbb{R}^d$ . Using the commutativity of convolution with even functions such as  $V_\varepsilon$ , it is indeed not difficult to see

$$\begin{aligned} \mathcal{H}_2[V_\varepsilon * \rho] &= \int_{\mathbb{R}^d} (V_\varepsilon * \rho)^2(x) dx \\ &= \int_{\mathbb{R}^d} (V_\varepsilon * V_\varepsilon) * \rho(x) d\rho(x) = \int_{\mathbb{R}^d} W_\varepsilon * \rho(x) d\rho(x) = 2\mathcal{W}_\varepsilon[\rho], \end{aligned}$$

by setting  $W_\varepsilon := V_\varepsilon * V_\varepsilon$ . This observation sheds light on the aforementioned link between local and nonlocal PDEs. As a natural byproduct such a connection provides a rigorous particle approximation for a class of nonlinear diffusion equations. More precisely, this hinges on deterministic approaches for nonlocal interaction equations, since, *particles are solutions*, i.e. the following empirical measure  $\rho_t^N$  is a weak solution of (1.1) with  $v = -\nabla \frac{\delta \mathcal{W}}{\delta \rho}$  and  $\mathcal{W}$  as in (1.3)

$$\rho_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_i(t)},$$

where, for any  $i = 1, \dots, N$ ,  $X_i(t)$  solves the ODE

$$\dot{X}_i(t) = -\frac{1}{N} \sum_j \nabla W(X_i(t) - X_j(t)).$$

Further details on this aspect can be found, e.g., in [9, 14], and in [31, 32] in case of systems of nonlocal PDEs. This structure is advantageous for the computational approximation of continuous solutions to (1.1). The main issue when diffusion is present is that particles do not remain particles. Indeed, if the initial datum is a Dirac delta, we have an immediate smoothing effect, excluding measure solutions. However, numerical evidence of these deterministic particle methods [10] show that this can be achieved, as we shall see later on.

In this manuscript, we consider a general class of internal energy functionals  $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$  given by

$$\mathcal{F}[\rho] := \begin{cases} \int_{\mathbb{R}^d} F(\rho(x)) dx, & \rho \ll \text{Leb}(\mathbb{R}^d) \\ +\infty, & \text{otherwise} \end{cases},$$

where we identify the measure  $\rho$  with its density  $\rho(x)$  if it is absolutely continuous with respect to Lebesgue measure and  $\mathcal{P}_2(\mathbb{R}^d)$  denotes the set of probability measures with finite second order moment. We define the regularised internal energy functional  $\mathcal{F}^\varepsilon : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$  given by

$$\mathcal{F}^\varepsilon[\rho] := \int_{\mathbb{R}^d} F(V_\varepsilon * \rho(x)) dx,$$

which gives rise to a class of nonlocal PDEs

$$\partial_t \rho = \nabla \cdot (\rho \nabla V_\varepsilon * F'(V_\varepsilon * \rho)). \tag{NLE}$$

The functional  $\mathcal{F}$  includes  $\mathcal{H}_m$ , but it is not limited to it, c.f. Sect. 2. The reader is invited to verify

$$\frac{\delta \mathcal{F}^\varepsilon}{\delta \rho}(\rho) = V_\varepsilon * \left[ \frac{\delta \mathcal{F}}{\delta \rho}(V_\varepsilon * \rho) \right],$$

which motivates the consideration of (NLE) as the 2-Wasserstein gradient flow of  $\mathcal{F}^\varepsilon$ . Following the strategy proposed in [7], defining the pressure by  $P(x) := x F'(x) - F(x)$ , as in [1, 19, 42], we construct weak solutions of the nonlinear diffusion equation

$$\partial_t \rho = \Delta P(\rho) \tag{DE}$$

as a limit of a sequence of weak measure solutions of (NLE), in case  $F$  behaves like power laws of porous medium type, for  $m > 1$ .

The blob method for diffusion was first introduced in [10] for diffusion equations with the addition of local and nonlocal drifts. Let us mention that a similar approach was used on

the previous work [26] approximating nonlocal equations with singular kernels by smooth kernels. The authors in [10] consider a slightly different regularisation of the internal energy which is better for numerical purposes, see [10, Eq. (6)]. Despite this difference, the gradient flow perspective remains at the forefront of their and our present work. The corresponding nonlocal gradient flow is indeed different from (NLE), c.f. [10, Eq. (8)], but it coincides with ours in case  $m = 2$  for the energy  $\mathcal{H}_2$ . In [10],  $\Gamma$ -convergence of the regularised energy, as well as that of minimisers is proven for  $m \geq 1$ . The authors show that stability of gradient flows in the  $\varepsilon \rightarrow 0$  can be established for  $m \geq 2$  using the framework introduced by Sandier and Serfaty in [53, 55] and the concept of  $\lambda$ -gradient flows developed in [1]. This strategy requires to verify additional assumptions which are only known to hold in the case  $m = 2$  for an initial datum with finite second order moment and log-entropy, i.e.  $\mathcal{H}_1[\rho_0] < \infty$ . The result for  $m = 2$  was previously proven in [40], however on a bounded domain with periodic boundary conditions. The blob method in [10] is a deterministic particle method for linear and nonlinear diffusion on  $\mathbb{R}^d$ . Numerical simulations in [10, Section 6] suggest that the particle approximation remains valid even when  $\mathcal{H}_1[\rho_0] = \infty$ . Relaxing the condition  $\mathcal{H}_1[\rho_0] < \infty$  and rigorously proving a quantitative particle approximation is still an open problem, and left for future research.

In the case  $m = 2$ , in the same spirit of [10, 27], the authors in [7] construct weak solutions of the quadratic porous medium equation as a localising limit ( $\varepsilon \rightarrow 0$ ) of a sequence of weak measure solutions of the nonlocal interaction equation (NLE), for  $F(x) = x^2$ . The authors work directly at the level of the (nonlocal) equations by means of a time-discretisation scheme which allows to work with lack of convexity, as for instance in the case of cross-diffusion systems, or even PDEs with no purely gradient flow structure. As in [10], finite initial log-entropy is required, thus excluding particle approximation. However, simultaneously to [7], the authors in [27] focus on a weighted (quadratic) porous medium equation which is relevant, e.g. in sampling — the weight,  $\bar{\rho}$  in their notations, represents a target probability measure to be approximated from specific samples drawn from it. The blob method is indeed useful to develop a deterministic particle approximation for the weighted porous medium equation, and, as a byproduct, it provides a way to quantize a target  $\bar{\rho}$  in the long-time behaviour. We stress that also in this work it is essential to assume  $\mathcal{H}_1[\rho_0] < \infty$ , however using again  $\lambda$ -convexity of the regularised energy one can achieve a rigorous particle approximation as consequence of  $\lambda$ -stability (or contractivity) of Wasserstein gradient flows, as in [1]. This means one can achieve, so far, a qualitative result, as the initial datum needs to be approximated fast enough, c.f. [27, Theorem 1.4]. To the best of our knowledge, a quantitative result has not been achieved yet in more than one dimension. Still in one space dimension, the authors in [29] introduce a deterministic particle approximation for aggregation-diffusion equations, including the porous medium equation for the subquadratic ( $1 < m < 2$ ) and superquadratic ( $m > 2$ ) cases. This approach, however, is limited to one space dimension. All the previous three works do not make use of gradient flow techniques. Indeed, other attempts for a particle method have been proposed in the literature. Let us mention two simultaneous numerical methods for linear diffusion ( $m = 1$ ) [30, 52]. In one dimension and for nonlinear diffusions, there are other numerical methods based on the PDE satisfied by the transporting maps, see [21, 23, 36]. A nice survey of most of the available numerical methods for these families of equations can be found in [18].

Further related to particle methods, we mention the seminal paper by Oelschläger, [45], where a stochastic particle approximation is proven for classical and positive solutions of the quadratic porous medium equation in  $\mathbb{R}^d$ , and for weak solutions in one dimension, and the recent results in [24] for systems. In [34] very weak solutions of the viscous porous medium equation ( $m > 1$ ) are studied as a limit of a sequence of distributions of the solutions to

nonlinear stochastic differential equations generalising previous results [34, 43, 46]. In [49] strong  $L^1$ -solutions, c.f. [57], of the quadratic porous medium equation are derived from a stochastic mean field interacting particle system with the addition of a vanishing Brownian motion.

Our strategy is different from the aforementioned stochastic approaches as it is based on an optimal transport approach avoiding the addition of higher regularity induced by the (vanishing) viscosity method. We consider a time-discretisation of (NLE) à la Jordan-Kinderlehrer-Otto (JKO), c.f. [37]. This method provides uniform bounds on the approximating sequence in terms of the associated energy and second order moments. Although the sequence solving (NLE) is only a measure, we are able to prove strong  $L^m$ -compactness of a smoother sequence of solutions for the  $\varepsilon \rightarrow 0$  limit by using the so-called *flow interchange* technique, c.f. [41]. More precisely, one of our main contributions is to construct weak solutions of

$$\partial_t \rho = \Delta(\rho^m) \tag{PME}$$

as a subsequential  $\varepsilon \rightarrow 0$  limit of weak measure solutions to

$$\partial_t \rho^\varepsilon = \frac{m}{m-1} \nabla \cdot (\rho^\varepsilon \nabla V_\varepsilon * (V_\varepsilon * \rho^\varepsilon)^{m-1}), \tag{NLE-m}$$

for all  $m > 1$ . The same result is proven also for (NLE) and (DE). In particular, this extends [7] to the case  $m > 2$ , which is not trivial in view of the nonlinearities involved, and to a class of general nonlinear diffusion function. In [10], their gradient flow convergence result for  $m > 2$  was conditional on a uniform  $BV$  bound for  $\rho^\varepsilon$  while we make no such assumptions here. Furthermore, we are also able to treat the case  $1 < m < 2$  which is more challenging due to the lack of regularity at zero.

As a byproduct of our analysis we obtain an existence result for nonlocal diffusion equations related to a nonlocal internal energy functional. In particular, we are able to construct weak solutions to (NLE-m) via the JKO scheme for  $m > 1$ . While this may not be surprising, this is the first result in this direction to the best of our knowledge. We also provide a particle approximation for (NLE) in case  $F$  behaves like power laws, for  $m > 1$ . This result is purely qualitative, and quantitative estimates are not proven. Finally, we stress that the strategy we use to construct weak solutions does not require convexity of the internal energy, thus allowing to extend this method to non-convex energies, e.g. nonlinear cross-diffusion systems, see [7]. We leave the extension to systems for a future work as it deserves a deeper analysis.

The case  $m = 1$ , i.e. linear diffusion, is not completely covered in our theory, due to the lack of control on the compactness near the logarithmic singularity in the gradient flow approach. More precisely, our strategy does provide an approximating scheme, validated numerically in [10], but we are not able at this stage to identify the limit as solution of the heat equation. Indeed, the logarithmic singularity cannot be coped with for the case  $m = 1$  when the mollifier  $V_1$  is compactly supported. This is indeed one of the reasons we did not assume  $V_1$  is compactly supported in the case  $m \geq 2$ . Similar difficulties are found for the Landau equation [39] in plasma physics, for which efficient deterministic particle methods preserving all the properties of the Landau equation at the discrete level were introduced in [16] using the same strategy as in this work. Moreover such an approximated Landau equation has been analytically studied in [12, 13] showing the existence of solutions for the approximated problems where  $V_1(x) = e^{-|x|}$  with an appropriate mollification at the origin. The particular non-compactly supported kernel is crucial in the detailed estimates performed in [12]. Dealing with the logarithmic singularity in these problems is a challenging open problem.

### Structure of the manuscript

Section 2 sets the assumptions, notations, and definitions we use in this paper. At the end of Sect. 2, we state the precise results obtained once the appropriate notions of solutions are introduced. Section 3 focuses on the construction of weak solutions  $\rho^\varepsilon$  to (NLE) (c.f. Theorem 2.1) based on the JKO scheme [37]. Section 4 discusses the strong compactness criteria used to construct a limit  $\rho$  (which is the candidate weak solution to (DE)) from the sequence  $\rho^\varepsilon$ . Section 5 verifies that the limit  $\rho$  is a weak solution to (DE) (c.f. Theorem 2.2) by passing to the limit  $\varepsilon \rightarrow 0$  from (NLE). In Sect. 6, we sketch the ideas behind the proofs of Theorem 2.3, which gives conditions for uniqueness of solutions to (NLE), and Corollary 2.1, which provides a particle approximation to (DE). Finally, Sect. 7 collects various technical results which, possibly with minor adaptations, already exist in the literature.

## 2 Preliminaries and results

The mollifying sequence is generated by  $V_\varepsilon(x) = \varepsilon^{-d} V_1(x/\varepsilon)$  for  $\varepsilon > 0$ . We assume that the generating function  $V_1$  satisfies

$$(V) \quad V_1 \in C_b(\mathbb{R}^d; [0, +\infty)) \cap C^1(\mathbb{R}^d), \|V_1\|_{L^1} = 1, V_1(x) = V_1(-x), \int_{\mathbb{R}^d} |x|^2 V_1(x) dx < +\infty, |\nabla V_1| \in L^1(\mathbb{R}^d), \text{ and } |\nabla V_1(x)| \leq C(1 + |x|).$$

Depending on the results we prove, we assume the function  $F : [0, +\infty) \rightarrow (-\infty, +\infty]$  satisfies some combination of the following assumptions:

(F1)  $F$  is a proper, convex, and lower semicontinuous function such that

$$F(0) = 0, \quad \liminf_{s \uparrow +\infty} \frac{F(s)}{s} = +\infty, \quad \liminf_{s \downarrow 0} \frac{F(s)}{s^\alpha} > -\infty, \quad \text{for some } \alpha > \frac{d}{d+2}.$$

(F2)  $F \in C^1([0, +\infty))$ .

(F3)  $F \in C([0, +\infty)) \cap C^2((0, +\infty))$ .

(F<sub>m</sub>) There exist  $c_1, c_2 > 0$  and  $m \geq 1$  such that  $c_1 x^{m-2} \leq F''(x) \leq c_2 x^{m-2}$  for all  $x > 0$ .

**Remark 2.1** (Comments on the assumptions) (F1) is lifted directly from [1, Example 9.3.6] so that  $\mathcal{F}$  enjoys certain properties; it is well-defined and the associated JKO scheme is well-posed c.f. [37].

For the reader’s convenience, we observe the condition  $\liminf_{s \downarrow 0} \frac{F(s)}{s^\alpha} > -\infty$  for some  $\alpha > \frac{d}{d+2}$  ensures (c.f. [1, Remark 9.3.7] and Lemma A.1) that  $F^-(\rho) \in L^1(\mathbb{R}^d)$  whenever  $\rho \in \mathcal{P}_2(\mathbb{R}^d)$  is absolutely continuous with respect to Lebesgue measure. In particular, on any sublevel subset  $\{\rho \in \mathcal{P}_2(\mathbb{R}^d) \mid m_2(\rho) \leq C\}$ , the functional  $\mathcal{F}$  is uniformly bounded below.

The superlinear growth (c.f. [1, Remark 9.3.8]) and convexity ensure that  $\mathcal{F}$  is lower semicontinuous in  $\mathcal{P}_1(\mathbb{R}^d)$ .

Assumption (F3) mainly refers to energies lacking regularity at the origin as in the case  $F(x) = x \log x$ . We stress that (F2) is used to construct solutions to (NLE) in Sect. 5, however it is not used to derive the compactness estimates in Sect. 4. Conversely, (F3) is used for the compactness estimates in Sect. 4 but is not assumed to construct solutions to (NLE). The motivating examples which satisfy all of (F1), (F2), (F3), and (F<sub>m</sub>) are power laws  $F(x) = \frac{1}{m-1} x^m$  for  $m > 1$ .

A further discussion can be found after the statements of Theorem 2.1 and Theorem 2.2.

Throughout the manuscript we will denote by  $\mathcal{P}(\mathbb{R}^d)$  the set of probability measures on  $\mathbb{R}^d$ , for  $d \in \mathbb{N}$ , and by  $\mathcal{P}_p(\mathbb{R}^d) := \{\rho \in \mathcal{P}(\mathbb{R}^d) : m_p(\rho) < +\infty\}$ , being  $m_p(\rho) := \int_{\mathbb{R}^d} |x|^p d\rho(x)$  the  $p^{\text{th}}$ -order moment of  $\rho$ , for  $1 \leq p < \infty$ . We shall use  $\mathcal{P}_p^a(\mathbb{R}^d)$  for elements in  $\mathcal{P}_p(\mathbb{R}^d)$  which are absolutely continuous with respect to the Lebesgue measure. For  $p = 2$ , the 2-Wasserstein distance between  $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$  is

$$d_W^2(\mu_1, \mu_2) := \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \left\{ \int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma(x, y) \right\}, \tag{2.1}$$

where  $\Gamma(\mu_1, \mu_2)$  is the class of all transport plans between  $\mu_1$  and  $\mu_2$ , that is the class of measures  $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$  such that, denoting by  $\pi_i$  the projection operator on the  $i$ -th component of the product space, the marginality condition

$$(\pi_i)_\# \gamma = \mu_i \quad \text{for } i = 1, 2$$

is satisfied. In the expression above, marginals are the push-forward of  $\gamma$  through  $\pi_i$ . For a measure  $\rho \in \mathcal{P}(\mathbb{R}^d)$  and a Borel map  $T : \mathbb{R}^d \rightarrow \mathbb{R}^n$ ,  $n \in \mathbb{N}$ , the push-forward of  $\rho$  through  $T$  is defined by

$$\int_{\mathbb{R}^n} f(y) dT\#\rho(y) = \int_{\mathbb{R}^d} f(T(x)) d\rho(x) \quad \text{for all Borel functions } f \text{ on } \mathbb{R}^n.$$

Setting  $\Gamma_0(\mu_1, \mu_2)$  as the class of optimal plans, i.e. minimizers of (2.1), the 2-Wasserstein distance can be written as

$$d_W^2(\mu_1, \mu_2) = \int_{\mathbb{R}^{2d}} |x - y|^2 d\gamma(x, y), \quad \gamma \in \Gamma_0(\mu_1, \mu_2).$$

We denote the 1-Wasserstein distance with  $d_1$  and it is defined by

$$d_1(\mu_1, \mu_2) := \min_{\gamma \in \Gamma(\mu_1, \mu_2)} \left\{ \int_{\mathbb{R}^{2d}} |x - y| d\gamma(x, y) \right\}. \tag{2.2}$$

We refer the reader to [1, 54, 61] for further details on optimal transport theory and Wasserstein spaces.

Below we recall the concepts of solutions used throughout the manuscript, distinguishing between measure and weak solutions.

**Definition 2.1** (Weak measure solution to (NLE)) Suppose  $F$  satisfies **(F1)** and **(F2)**. An absolutely continuous curve  $\rho^\varepsilon : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ , mapping  $t \in [0, T] \mapsto \rho_t^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$ , is a weak measure solution to **(NLE)** if, for every  $\varphi \in C_c^1(\mathbb{R}^d)$  and any  $t \in [0, T]$ , it holds

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) - \int_{\mathbb{R}^d} \varphi(x) d\rho_0(x) \\ &= - \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot [\nabla V_\varepsilon * F'(V_\varepsilon * \rho_s^\varepsilon)](x) d\rho_s^\varepsilon(x) ds. \end{aligned} \tag{2.3}$$

**Definition 2.2** (Weak measure solution to (NLE-m)) An absolutely continuous curve  $\rho^\varepsilon : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ , mapping  $t \in [0, T] \mapsto \rho_t^\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$ , is a weak measure solution to **(NLE-m)** for  $m > 1$  if, for every  $\varphi \in C_c^1(\mathbb{R}^d)$  and any  $t \in [0, T]$ , it holds

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) - \int_{\mathbb{R}^d} \varphi(x) d\rho_0(x) \\ &= - \frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot [\nabla V_\varepsilon * (V_\varepsilon * \rho_s^\varepsilon)^{m-1}](x) d\rho_s^\varepsilon(x) ds. \end{aligned} \tag{2.4}$$

**Remark 2.2** By considering fixed  $\varepsilon > 0$  and the corresponding scaling for  $V_1$  satisfying **(V)**, the driving velocity field satisfies

$$\begin{aligned}
 & \int_0^T \int_{\mathbb{R}^d} |[\nabla V_\varepsilon * (V_\varepsilon * \rho_t^\varepsilon)^{m-1}](x)| d\rho_t^\varepsilon(x) dt \\
 & \leq \int_0^T \iint_{\mathbb{R}^{2d}} |\nabla V_\varepsilon(x - y)|(V_\varepsilon * \rho_t^\varepsilon)^{m-1}(y) dy d\rho_t^\varepsilon(x) dt \\
 & = \int_0^T \int_{\mathbb{R}^d} (|\nabla V_\varepsilon| * \rho_t^\varepsilon)(y)(V_\varepsilon * \rho_t^\varepsilon)^{m-1}(y) dy dt \\
 & \leq \int_0^T \|V_\varepsilon * \rho_t^\varepsilon\|_{L^\infty}^{m-1} \left( \int_{\mathbb{R}^d} (|\nabla V_\varepsilon| * \rho_t^\varepsilon)(y) dy \right) dt \tag{2.5} \\
 & \leq \varepsilon^{-md} \|V_1\|_{L^\infty} \int_0^T \| |\nabla V_\varepsilon| * \rho_t^\varepsilon \|_{L^1} dt \\
 & \leq \varepsilon^{-md} \|V_1\|_{L^\infty} T \|\nabla V_\varepsilon\|_{L^1} \\
 & = \frac{T}{\varepsilon^{md+1}} \|V_1\|_{L^\infty} \|\nabla V_1\|_{L^1} < \infty.
 \end{aligned}$$

[1, Lemma 8.2.1] provides the existence of a continuous representative for distributional solutions of continuity equations with velocity fields in  $L^1([0, T]; L^1(\rho_t))$ . This justifies Definition 2.2 in the sense that the right-hand side of (2.4) is well-defined. Note that a similar computation holds true for the velocity field in **(NLE)** by applying Lemma A.3, thus justifying Definition 2.1 in the sense that the right-hand side of (2.3) is well-defined.

**Definition 2.3** (Weak solution to **(PME)**) A weak solution to the Cauchy problem for  $m > 1$

$$\begin{cases} \partial_t \rho = \Delta \rho^m \\ \rho(0, \cdot) = \rho_0 \end{cases} \tag{PME}$$

on the time interval  $[0, T]$  with initial datum  $\rho_0 \in \mathcal{P}_2^a(\mathbb{R}^d) \cap L^m(\mathbb{R}^d)$  is an absolutely continuous curve  $\rho \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$  satisfying the following properties:

- (1) for almost every  $t \in [0, T]$  the measure  $\rho(t)$  has a density with respect to the Lebesgue measure, still denoted by  $\rho(t)$ , such that  $\rho \in L^\infty([0, T]; L^m(\mathbb{R}^d))$  and  $\nabla \rho^{\frac{m}{2}} \in L^2([0, T]; L^2(\mathbb{R}^d))$ ;
- (2) for any  $\varphi \in C_c^1(\mathbb{R}^d)$  and all  $t \in [0, T]$  it holds

$$\int_{\mathbb{R}^d} \varphi(x) \rho(t, x) dx = \int_{\mathbb{R}^d} \varphi(x) \rho_0(x) dx - \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot \nabla \rho(s, x)^m dx ds;$$

- (3)  $\rho^{-\frac{1}{2}} |\nabla \rho^m| \in L^1([0, T]; L^2(\mathbb{R}^d))$ .

**Remark 2.3** For the sake of clarity we point out the weak solution we obtain initially satisfies, for any test function  $\varphi \in C_c^1(\mathbb{R}^d)$ ,

$$\int_{\mathbb{R}^d} \varphi(x) \rho(t, x) dx = \int_{\mathbb{R}^d} \varphi(x) \rho_0(x) dx - 2 \int_0^t \int_{\mathbb{R}^d} \rho(s, x)^{\frac{m}{2}} \nabla \varphi(x) \cdot \nabla \rho(s, x)^{\frac{m}{2}} dx ds.$$

The chain rule in Sobolev spaces gives sense to  $\nabla \rho^m$  in  $L^1(\mathbb{R}^d)$ , hence the more standard concept of weak solution for porous medium equation. A further application of the chain rule identifies  $\nabla \rho^m = \frac{m}{m-1} \rho \nabla \rho^{m-1}$ , for  $m \geq 2$ ; the same result, however, does not hold in the



case  $1 < m < 2$ . Further details are provided in the proof of Theorem 2.2 in Sect. 5. Finally, the last condition in Definition 2.3 is a consequence of uniqueness of very weak solutions, cf. [28], and the theory in [1].

Equally, the same concept is extended to general diffusion equations.

**Definition 2.4** (Weak solution to (DE)) Let  $F$  satisfy (F1), (F2), (F3), and (F<sub>m</sub>) for some  $m > 1$ . A weak solution to the Cauchy problem

$$\begin{cases} \partial_t \rho = \Delta P(\rho) \\ \rho(0, \cdot) = \rho_0 \end{cases} \tag{DE}$$

on the time interval  $[0, T]$  with initial datum  $\rho_0 \in \mathcal{P}_2^a(\mathbb{R}^d)$  such that  $\mathcal{F}[\rho_0] < \infty$  is an absolutely continuous curve  $\rho \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$  satisfying the following properties:

- (1) for almost every  $t \in [0, T]$  the measure  $\rho(t)$  has a density with respect to the Lebesgue measure, still denoted by  $\rho(t)$ , such that  $\rho \in L^\infty([0, T]; L^m(\mathbb{R}^d))$  and  $\nabla \rho^{\frac{m}{2}} \in L^2([0, T]; L^2(\mathbb{R}^d))$ ;
- (2) for any  $\varphi \in C_c^1(\mathbb{R}^d)$  and all  $t \in [0, T]$  it holds

$$\int_{\mathbb{R}^d} \varphi(x) \rho(t, x) dx = \int_{\mathbb{R}^d} \varphi(x) \rho_0(x) dx - \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot \nabla P(\rho(s, x)) dx ds;$$

- (3)  $\rho^{-\frac{1}{2}} |\nabla P(\rho)| \in L^1([0, T]; L^2(\mathbb{R}^d))$ .

With the previous definitions, we are ready to state the results of this manuscript.

**Theorem 2.1** (Existence for (NLE)) Fix  $\varepsilon > 0$  and let  $V_1$ , the generator of the mollifying sequence  $V_\varepsilon(x) = \varepsilon^{-d} V_1(x/\varepsilon)$ , satisfy (V). Let  $F$  satisfy (F1) and (F2) and suppose  $\mathcal{F}^\varepsilon[\rho_0] < +\infty$ . Then, there exist weak measure solutions  $\rho^\varepsilon$  to (NLE) such that  $\rho^\varepsilon(0) = \rho_0$ .

**Theorem 2.2** ( $\lim_{\varepsilon \rightarrow 0}(\text{NLE}) = (\text{DE})$ ) Let  $F$  satisfy (F1), (F2), (F3), and (F<sub>m</sub>) for some  $m > 1$ . Suppose  $\rho_0 \in \mathcal{P}_2^a(\mathbb{R}^d)$  such that  $\mathcal{F}[\rho_0] < \infty$  and  $V_1$  satisfies (V). In the case  $1 < m < 2$ , assume further that  $\text{supp} V_1 \subset B_R$  for some  $R > 0$ . Let  $\rho^\varepsilon$  be a sequence of weak measure solutions to (NLE) from Theorem 2.1 with initial condition  $\rho^\varepsilon(0) = \rho_0$ . Then, the sequence  $\rho^\varepsilon$  converges narrowly to the unique weak solution  $\rho$  of (DE) as  $\varepsilon \downarrow 0$ .

**Remark 2.4** In the case  $F$  is a power law given by  $F(x) = \frac{1}{m-1} |x|^m$  for some  $m > 1$ , all of (F1), (F2), (F3), and (F<sub>m</sub>) are fulfilled; Theorem 2.2 holds for (PME).

At first glance, our compactness estimates only show that a subsequence of  $\rho^\varepsilon$  converges narrowly to  $\rho$ . However, we appeal to [5, 28, 58] which imply that weak solutions to (DE) and (PME) are unique. Hence, the entire sequence converges.

In Theorem 2.1, the construction of weak measure solutions  $\rho^\varepsilon$  to (NLE) leverages the JKO scheme [37]. Just at the level of the JKO scheme, only (F1) is required (c.f. Proposition 3.1) for which all of the regularised Rényi entropies  $\mathcal{H}_m^\varepsilon[\rho] = \mathcal{H}_m[V_\varepsilon * \rho]$  for any  $m \geq 1$  are admissible. In fact, assumption (F2) enters only when verifying  $\rho^\varepsilon$  is a weak measure solution of (NLE) (c.f. Sect. 3). This excludes  $F(x) = x \log x$ , but all the power laws for  $m > 1$  are permitted in this consistency result. Moreover, the assumption that (F<sub>m</sub>) holds for some  $m > 1$  in Theorem 2.2 is only used to verify that the limit  $\rho$  is a weak solution to (DE). On the other hand, the construction of the limit  $\rho$  from the sequence  $\rho^\varepsilon$  allows to relax assumption (F<sub>m</sub>) to any  $m \geq 1$  provided the initial condition  $\rho^\varepsilon(0) = \rho_0$  belongs in  $L^m \cap L \log L$  (c.f. Sect. 4), thus including all of the regularised Rényi entropies  $\mathcal{H}_m^\varepsilon$ . To

summarise in the specific case of  $\mathcal{F}^\varepsilon = \mathcal{H}_m^\varepsilon$  as the regularised energy, the construction of curves  $\rho^\varepsilon$  and  $\rho$  without consideration of the respective equations (NLE-m) and (PME) can be done for any  $m \geq 1$ . However, our technique requires  $m > 1$  to verify that  $\rho^\varepsilon$  is a weak measure solution of (NLE-m). Moreover, when  $1 < m < 2$ , we insist that the generator,  $V_1$  of the mollifying sequence, satisfies (V) and has compact support (in the case  $m \geq 2$  only (V) is required). It is certainly interesting to investigate how we can close this gap to  $m = 1$  and we leave this direction for future research.

In Theorem 2.2 we prove that the solutions  $\rho^\varepsilon$  to (NLE) coming from the construction in Theorem 2.1 converge to  $\rho$ , the unique weak solution of (DE). It is natural to ask whether other solutions  $\tilde{\rho}^\varepsilon$  to (NLE) (not necessarily those constructed via the JKO scheme c.f. Sect. 3) also converge to  $\rho$ . Actually, under additional assumptions on the nonlinearity  $F$  and the mollifier  $V$ , the sequence  $\rho^\varepsilon$  is unique.

**Theorem 2.3** (Uniqueness of solutions to (NLE)) *Let  $F$  satisfy (F1), (F2), (F3), and (F<sub>m</sub>) for some  $m > 1$ . Assume  $V_1$  satisfies (V),  $V_1 \in C^2(\mathbb{R}^d)$ , and  $D^2V_1 \in L^\infty(\mathbb{R}^d)$ . Then, the weak measure solution  $\rho^\varepsilon$  in Theorem 2.1 is unique among absolutely continuous curves  $\rho : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$  satisfying (NLE) in the sense of Definition 2.1.*

The following concluding result is completely analogous to Theorem 1.2 of [27].

**Corollary 2.1** (Particle approximation to (DE)) *Let  $F$  satisfy (F1), (F2), (F3), and (F<sub>m</sub>) for some  $m > 1$ . Assume  $V_1$  satisfies (V),  $V_1 \in C^2(\mathbb{R}^d)$ , and  $D^2V_1 \in L^\infty(\mathbb{R}^d)$ . In the case  $1 < m < 2$ , assume moreover that  $\text{supp}V_1 \subset B_R$  for some  $R > 0$ . For any  $t \in [0, T]$ ,  $N \in \mathbb{N}$ , the empirical measure  $\rho_\varepsilon^N(t) = \frac{1}{N} \sum_{j=1}^N \delta_{x_\varepsilon^j(t)}$  is a weak solution to (NLE) provided the particles satisfy the following ODE system*

$$\dot{x}_\varepsilon^i(t) = -\nabla \int_{\mathbb{R}^d} V_\varepsilon(x_\varepsilon^i(t) - y) F' \left( \frac{1}{N} \sum_{j=1}^N V_\varepsilon(y - x_\varepsilon^j(t)) \right) dy \quad \forall i = 1, \dots, N.$$

Suppose that (up to a subsequence) as  $\varepsilon \rightarrow 0$  there exist  $N = N(\varepsilon) \rightarrow +\infty$  such that

$$e^{-\lambda_F^\varepsilon t} d_W(\rho_\varepsilon^N(0), \rho(0)) \rightarrow 0, \quad \text{for } \lambda_F^\varepsilon \approx -\varepsilon^{-2-d(m-1)}, \quad t \in [0, T],$$

with  $\rho_0 \in \mathcal{P}_2^a(\mathbb{R}^d)$  such that  $\mathcal{F}[\rho_0] < \infty$  and  $T > 0$ . Then  $\rho_\varepsilon^N(t)$  converges narrowly to a weak solution of (DE),  $\rho(t)$ , for any  $t \in [0, T]$ .

In view of Corollary 2.1 and [27], if the initial distribution of particles  $x_\varepsilon^i$  is cleverly chosen (so that  $d_W(\rho_\varepsilon^N(0), \rho(0)) = O(1/N)$ ), then one can take  $N = o\left(e^{-1/\varepsilon^{2+d(m-1)}}\right)$  to fulfill the hypothesis on the initial condition. However, it was also suggested in [26, 27] by numerical evidence that a much smaller number of particles  $N \sim \varepsilon^{-1.01}$  for  $m = 2$  in one dimension still yields good accuracy. Bridging this gap between theory and practice is left for future investigation.

### 3 Results on the nonlocal equation

In this section we focus on (NLE). We show existence of weak measure solutions by means of the JKO scheme [37] which is needed to derive uniform bounds for the nonlocal-to-local limit proven in Sect. 5. Although this is not the main purpose of the paper, and it may be unsurprising, this is indeed an existence result for weak measure solutions to a class of nonlocal PDEs, including nonlocal interactions but not limited to this case. To the best of

our knowledge this is the first general result in this context — the structure of  $\mathcal{F}^\varepsilon$  does not fit in the classical framework of functionals considered in [1]. Note that we do not require the functional to satisfy convexity, for instance as in [1, 14].

We consider initial data  $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$  such that  $\sup_{\varepsilon>0} \mathcal{F}^\varepsilon[\rho_0] < +\infty$ . In the case of nonlinear diffusion equations,  $F(x) = \frac{1}{m-1}|x|^m$  with  $m > 1$  we denote the corresponding energy functionals by

$$\mathcal{J}_m^\varepsilon[\rho] := \frac{1}{m-1} \int_{\mathbb{R}^d} |V_\varepsilon * \rho(x)|^m dx.$$

**Remark 3.1** In the case of power laws  $F(x) = \frac{1}{m-1}|x|^m$  for  $m > 1$ , the condition  $\sup_{\varepsilon>0} \mathcal{J}_m^\varepsilon[\rho_0] < +\infty$  is guaranteed when  $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d) \cap L^m(\mathbb{R}^d)$ . More precisely, Young’s convolution inequality gives

$$\begin{aligned} \mathcal{J}_m^\varepsilon[\rho_0] &= \frac{1}{m-1} \int_{\mathbb{R}^d} |V_\varepsilon * \rho_0(x)|^m dx = \frac{1}{m-1} \|V_\varepsilon * \rho_0\|_{L^m(\mathbb{R}^d)}^m \\ &\leq \frac{1}{m-1} \|V_\varepsilon\|_{L^1}^m \|\rho_0\|_{L^m}^m = \frac{1}{m-1} \|V_1\|_{L^1}^m \|\rho_0\|_{L^m}^m < \infty. \end{aligned}$$

We now proceed with the JKO scheme associated to  $\mathcal{F}^\varepsilon$ . First, we define a sequence recursively as follows:

- fix a time step  $\tau \in (0, 1)$  such that  $\rho_{\tau,\varepsilon}^0 := \rho_0$ ;
- for  $n \in \mathbb{N}$  and given  $\rho_{\tau,\varepsilon}^n \in \mathcal{P}_2(\mathbb{R}^d)$ , choose

$$\rho_{\tau,\varepsilon}^{n+1} \in \operatorname{argmin}_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \frac{d_W^2(\rho_{\tau,\varepsilon}^n, \rho)}{2\tau} + \mathcal{F}^\varepsilon[\rho] \right\}. \tag{3.1}$$

The above sequence is well-defined for  $\tau$  sufficiently small independently of  $\varepsilon$  (given explicitly in Lemma A.2).

Let  $T > 0$  be fixed, and define a piecewise constant interpolation as follows: take  $N := \lceil \frac{T}{\tau} \rceil$  the largest integer less than or equal to  $\frac{T}{\tau}$  and set

$$\rho_\tau^\varepsilon(t) = \rho_{\tau,\varepsilon}^n \quad t \in ((n-1)\tau, n\tau], \quad n = 0, 1, \dots, N,$$

being  $\rho_{\tau,\varepsilon}^n$  defined in (3.1). As usually proven, we derive energy and moments bounds sufficient to show narrow compactness.

**Proposition 3.1** (Narrow compactness, energy, & moments bound) *Let  $0 < \varepsilon_0 < \infty$  be fixed and suppose  $F$  satisfied (F1). There exists an absolutely continuous curve  $\tilde{\rho}^\varepsilon : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$  such that the piecewise constant interpolation  $\rho_\tau^\varepsilon$  admits a subsequence  $\rho_{\tau_k}^\varepsilon$  narrowly converging to  $\tilde{\rho}^\varepsilon$  uniformly in  $t \in [0, T]$  and  $0 < \varepsilon \leq \varepsilon_0$  as  $k \rightarrow +\infty$ . Moreover, for any  $t \in [0, T]$ , the following uniform bounds in  $\tau$  and  $0 < \varepsilon \leq \varepsilon_0$  hold*

$$\mathcal{F}^\varepsilon[\tilde{\rho}^\varepsilon(t)] \leq \sup_{\varepsilon>0} \mathcal{F}^\varepsilon[\rho_0], \quad m_2(\tilde{\rho}^\varepsilon) \leq C \left( T, m_2(\rho_0), \varepsilon_0^2 m_2(V_1), \sup_{\varepsilon>0} \mathcal{F}^\varepsilon[\rho_0] \right),$$

where  $C(T, m_2(\rho_0), \varepsilon_0^2 m_2(V_1), \sup_{\varepsilon>0} \mathcal{F}^\varepsilon[\rho_0]) > 0$  is a uniform constant depending only on the quantities in the brackets.

The following proof is based on [1, 37].

**Proof** From the definition of the sequence  $\{\rho_{\tau,\varepsilon}^n\}_{n=0,\dots,N}$  it holds

$$\frac{d_W^2(\rho_{\tau,\varepsilon}^n, \rho_{\tau,\varepsilon}^{n+1})}{2\tau} + \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}] \leq \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^n], \quad \forall n = 0, \dots, N - 1. \tag{3.3}$$

which implies  $\mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}] \leq \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^n]$ , and, in particular, the following bound for the regularised internal energy

$$\sup_{0 \leq n \leq N, N\tau \leq T} \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^n] \leq \mathcal{F}^\varepsilon[\rho_0], \tag{3.4}$$

where the supremum is over all  $n = 0, \dots, N$  and  $\tau \in (0, 1)$  such that  $N\tau \leq T$  with  $N := \lceil \frac{T}{\tau} \rceil$ . By summing up over  $k$  in inequality (3.3), we obtain

$$\sum_{k=m}^n \frac{d_W^2(\rho_{\tau,\varepsilon}^k, \rho_{\tau,\varepsilon}^{k+1})}{2\tau} \leq \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^m] - \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}], \quad \forall 0 \leq m \leq n \leq N - 1. \tag{3.5}$$

**Bounded second moment:** We claim the existence of some uniform constant  $C > 0$  (depending on the quantities discussed in the statement of this result) such that

$$\sup_{0 \leq n \leq N, N\tau \leq T} m_2(\rho_{\tau,\varepsilon}^n) \leq C. \tag{3.6}$$

By Remark A.1, for any fixed  $n = 0, \dots, N - 1$ , we begin with

$$m_2(\rho_{\tau,\varepsilon}^{n+1}) \leq 2d_W^2(\rho_0, \rho_{\tau,\varepsilon}^{n+1}) + 2m_2(\rho_0).$$

We use the triangle inequality and Cauchy-Schwarz to estimate the  $d_W^2$  term

$$m_2(\rho_{\tau,\varepsilon}^{n+1}) \leq 2(n + 1) \sum_{k=0}^n d_W^2(\rho_{\tau,\varepsilon}^k, \rho_{\tau,\varepsilon}^{k+1}) + 2m_2(\rho_0).$$

We replace the summation with (3.5) and use  $n + 1 \leq N$  to obtain

$$m_2(\rho_{\tau,\varepsilon}^{n+1}) \leq 4T(\mathcal{F}^\varepsilon[\rho_0] - \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}]) + 2m_2(\rho_0).$$

We insert the lower bound for  $\mathcal{F}^\varepsilon$  from (A.1) so that we have

$$m_2(\rho_{\tau,\varepsilon}^{n+1}) \leq 4T(\mathcal{F}^\varepsilon[\rho_0] + c_1 + c_2 C_{d,\alpha}(1 + \varepsilon^2 m_2(V_1) + m_2(\rho_{\tau,\varepsilon}^{n+1}))^\alpha) + 2m_2(\rho_0).$$

Keeping in mind that we can assume  $\alpha < 1$  without loss of generality, this final inequality implies the bound (3.6). This can be seen by analysing sequences  $x_n \geq 0$  satisfying  $x_n \leq C_1 + C_2 x_n^\alpha$ .

**Bounded squared 2-Wasserstein distance:** We claim the existence of some uniform constant  $c > 0$  such that

$$\sum_{k=m}^n d_W^2(\rho_{\tau,\varepsilon}^k, \rho_{\tau,\varepsilon}^{k+1}) \leq c\tau, \quad \forall 0 \leq m \leq n \leq N - 1. \tag{3.7}$$

We insert the upper bound of  $\mathcal{F}^\varepsilon$  (3.4) and the lower bound of  $\mathcal{F}^\varepsilon$  (A.1) into (3.5) to obtain

$$\begin{aligned} \sum_{k=m}^n d_W^2(\rho_{\tau,\varepsilon}^k, \rho_{\tau,\varepsilon}^{k+1}) &\leq 2\tau(\mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^m] - \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}]) \\ &\leq 2\tau(\mathcal{F}^\varepsilon[\rho_0] + c_1 + c_2 C_{d,\alpha}(1 + \varepsilon^2 m_2(V_1) + m_2(\rho_{\tau,\varepsilon}^{n+1}))^\alpha). \end{aligned}$$

By the uniform second moment estimate (3.6), the inequality (3.7) is verified.

Compactness: Now, let us consider  $0 < s < t$  such that  $s \in ((m - 1)\tau, m\tau]$  and  $t \in ((n - 1)\tau, n\tau]$  (which implies  $|n - m| < \frac{|t-s|}{\tau} + 1$ ); by Cauchy-Schwarz inequality and (3.7), we obtain

$$d_W(\rho_\tau^\varepsilon(s), \rho_\tau^\varepsilon(t)) \leq \sum_{k=m}^{n-1} d_W(\rho_{\tau,\varepsilon}^k, \rho_{\tau,\varepsilon}^{k+1}) \leq \left( \sum_{k=m}^{n-1} d_W^2(\rho_{\tau,\varepsilon}^k, \rho_{\tau,\varepsilon}^{k+1}) \right)^{\frac{1}{2}} |n - m|^{\frac{1}{2}} \tag{3.8}$$

$$\leq c \left( \sqrt{|t - s|} + \sqrt{\tau} \right),$$

where  $c$  is a positive constant. Thus  $\rho_\tau^\varepsilon$  is  $\frac{1}{2}$ -Hölder equicontinuous, up to a negligible error of order  $\sqrt{\tau}$ . By using a refined version of Ascoli-Arzelà’s theorem, [1, Proposition 3.3.1], we obtain that  $\rho_\tau^\varepsilon$  admits a subsequence  $\rho_{\tau_k}^\varepsilon$  narrowly converging to a limit  $\tilde{\rho}^\varepsilon$  as  $k \rightarrow +\infty$  uniformly on  $[0, T]$ . Since  $|\cdot|^2$  is lower semicontinuous and bounded from below, we actually have for any  $t \in [0, T]$

$$\liminf_{k \rightarrow +\infty} \int_{\mathbb{R}^d} |x|^2 d\rho_{\tau_k}^\varepsilon(x) \geq \int_{\mathbb{R}^d} |x|^2 d\tilde{\rho}^\varepsilon(x).$$

Moreover,  $\mathcal{F}^\varepsilon$  is lower semicontinuous and bounded from below since  $V_\varepsilon * \rho$  is bounded. Then an application of Fatou’s lemma implies

$$\liminf_{k \rightarrow +\infty} \mathcal{F}^\varepsilon[\rho_{\tau_k}^\varepsilon] \geq \mathcal{F}^\varepsilon[\tilde{\rho}^\varepsilon],$$

whence the thesis follows by applying the above inequalities to (3.4) and (3.6). □

Next, we show that  $\tilde{\rho}^\varepsilon$  provided by Proposition 3.1 is indeed a solution to (NLE), thus proving Theorem 2.1. Since we make use of (F2), the theorem below does not include linear diffusion corresponding to  $F(x) = x \log x$ .

**Proof of Theorem 2.1** Let us consider two consecutive elements of the sequence  $\{\rho_{\tau,\varepsilon}^n\}_{n \in \mathbb{N}}$  defined from the JKO step (3.1), i.e.  $\rho_{\tau,\varepsilon}^n$  and  $\rho_{\tau,\varepsilon}^{n+1}$ . We perturb  $\rho_{\tau,\varepsilon}^{n+1}$  by using the map  $P^\sigma = \text{id} + \sigma \zeta$ , for some  $\zeta \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$  and  $\sigma > 0$ , that is we consider the perturbation

$$\rho^\sigma := P_\#^\sigma \rho_{\tau,\varepsilon}^{n+1}. \tag{3.9}$$

Being  $\rho_{\tau,\varepsilon}^{n+1}$  a minimiser of (3.1), we have

$$\frac{1}{2\tau} \left[ \frac{d_W^2(\rho_{\tau,\varepsilon}^n, \rho^\sigma) - d_W^2(\rho_{\tau,\varepsilon}^n, \rho_{\tau,\varepsilon}^{n+1})}{\sigma} \right] + \frac{\mathcal{F}^\varepsilon[\rho^\sigma] - \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}]}{\sigma} \geq 0. \tag{3.10}$$

We now let  $\sigma \rightarrow 0$  in (3.10) analysing the two terms involved separately.

The energy functional terms in (3.10): In this part of the proof, we aim to show

$$\frac{\mathcal{F}^\varepsilon[\rho^\sigma] - \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}]}{\sigma} \rightarrow \int_{\mathbb{R}^d} \zeta(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})](x) d\rho_{\tau,\varepsilon}^{n+1}(x), \quad \sigma \rightarrow 0. \tag{3.11}$$

We apply the mean-value form of the Taylor expansion to  $F$

$$\begin{aligned}
 \frac{1}{\sigma} \int_{\mathbb{R}^d} (F(V_\varepsilon * \rho^\sigma(x)) - F(V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})) dx &= \frac{1}{\sigma} \int_{\mathbb{R}^d} (V_\varepsilon * \rho^\sigma(x) \\
 &\quad - V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1}(x)) \underbrace{\int_0^1 F'(tV_\varepsilon * \rho^\sigma(x) + (1-t)V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1}(x)) dt}_{=: M_\varepsilon^\sigma(x)} dx \\
 &= \frac{1}{\sigma} \int_{\mathbb{R}^d} (V_\varepsilon * M_\varepsilon^\sigma(x)) d[\rho^\sigma - \rho_{\tau,\varepsilon}^{n+1}](x) \\
 &= \int_{\mathbb{R}^d} \frac{(V_\varepsilon * M_\varepsilon^\sigma)(P^\sigma(x)) - (V_\varepsilon * M_\varepsilon^\sigma)(x)}{\sigma} d\rho_{\tau,\varepsilon}^{n+1}(x) \\
 &= \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \right\} d\rho_{\tau,\varepsilon}^{n+1}(x).
 \end{aligned} \tag{3.12}$$

In the last few lines, we used the definition of  $\rho^\sigma$  from (3.9) and expanded the convolution. The limit (3.11) is achieved by applying Egorov’s theorem. First, we prove convergence up to sets of  $\rho_{\tau,\varepsilon}^{n+1}$ -measure zero for the term in curly brackets, which is a sequence of functions of  $x$ , i.e.:

$$\begin{aligned}
 &\int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \\
 &\rightarrow \zeta(x) \cdot \int_{\mathbb{R}^d} \nabla V_\varepsilon(x - y) F'(V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1}(y)) dy, \quad \sigma \rightarrow 0, \rho_{\tau,\varepsilon}^{n+1}\text{-almost every } x \in \mathbb{R}^d.
 \end{aligned} \tag{3.13}$$

This is exactly  $\zeta(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})](x)$  which appears as the integrand in (3.11). Assuming this is true for now, by Egorov’s theorem, for every  $\eta > 0$ , there exists a measurable set  $S_\eta \subset \mathbb{R}^d$  such that  $\rho_{\tau,\varepsilon}^{n+1}(S_\eta) < \eta$  and the convergence (3.13) is uniform on  $\mathbb{R}^d \setminus S_\eta$ . Continuing from the last line of (3.12), we have

$$\begin{aligned}
 &\frac{\mathcal{F}^\varepsilon[\rho^\sigma] - \mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}]}{\sigma} \\
 &= \int_{S_\eta} \left\{ \int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \right\} d\rho_{\tau,\varepsilon}^{n+1}(x) \\
 &\quad + \int_{\mathbb{R}^d \setminus S_\eta} \left\{ \int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \right\} d\rho_{\tau,\varepsilon}^{n+1}(x).
 \end{aligned} \tag{3.14}$$

The integral over  $\mathbb{R}^d \setminus S_\eta$  passes well in the limit  $\sigma \rightarrow 0$  owing to (3.13) and Egorov’s theorem, so (3.11) is achieved once we show that the integral over  $S_\eta$  is small. We apply the mean-value form of Taylor’s theorem for  $V_\varepsilon$  and Lemma A.3 (with  $\rho = t\rho^\sigma + (1-t)\rho_{\tau,\varepsilon}^{n+1}$

and  $C = 1$ ) to estimate  $M_\varepsilon^\sigma$  and obtain

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \right| \\ & \leq \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} |\zeta(x)| \int_{\mathbb{R}^d} \int_0^1 |\nabla V_\varepsilon(x + s\sigma\zeta(x) - y)| ds dy \\ & = \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} |\zeta(x)| \int_0^1 \int_{\mathbb{R}^d} |\nabla V_\varepsilon(x + s\sigma\zeta(x) - y)| dy ds \\ & = \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} |\zeta(x)| \int_0^1 \int_{\mathbb{R}^d} |\nabla V_\varepsilon(z)| dz ds \\ & = \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} \|\nabla V_\varepsilon\|_{L^1} |\zeta(x)|. \end{aligned}$$

In the second to last line, we have used Fubini and the linear change of variables  $z = x + s\sigma\zeta(x) - y$  for fixed  $x$ . Therefore, the integral over  $S_\eta$  from (3.14) can be estimated by

$$\begin{aligned} & \left| \int_{S_\eta} \left\{ \int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \right\} d\rho_{\tau,\varepsilon}^{n+1}(x) \right| \\ & \leq \|\zeta\|_{L^\infty} \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} \|\nabla V_\varepsilon\|_{L^1} \eta, \end{aligned}$$

which is negligible by taking  $\eta \rightarrow 0$ .

Proving (3.13): Throughout this step, we fix  $x \in \mathbb{R}^d$ . We again use the mean-value form of Taylor's theorem to rewrite the difference quotient appearing in (3.13)

$$\begin{aligned} & \int_{\mathbb{R}^d} \left( \frac{V_\varepsilon(P^\sigma(x) - y) - V_\varepsilon(x - y)}{\sigma} \right) M_\varepsilon^\sigma(y) dy \\ & = \zeta(x) \cdot \int_{\mathbb{R}^d} \left( \int_0^1 \nabla V_\varepsilon(x + s\sigma\zeta(x) - y) ds \right) M_\varepsilon^\sigma(y) dy. \end{aligned}$$

We majorise the integrand with the sequence

$$\left| \int_0^1 \nabla V_\varepsilon(x + s\sigma\zeta(x) - y) ds M_\varepsilon^\sigma(y) \right| \leq \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} \int_0^1 |\nabla V_\varepsilon(x + s\sigma\zeta(x) - y)| ds.$$

We seek to apply Theorem A.1 on  $X = \mathbb{R}^d$  with

$$\begin{aligned} f^\sigma(y) & := \int_0^1 \nabla V_\varepsilon(x + s\sigma\zeta(x) - y) ds M_\varepsilon^\sigma(y), \\ g^\sigma(y) & := \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} \int_0^1 |\nabla V_\varepsilon(x + s\sigma\zeta(x) - y)| ds. \end{aligned}$$

We have already shown the majorisation  $|f^\sigma(y)| \leq g^\sigma(y)$  and the convergence

$$\begin{aligned} f^\sigma(y) & \rightarrow f(y) := \nabla V_\varepsilon(x - y) F'(V_\varepsilon * \rho(y)), \\ g^\sigma(y) & \rightarrow g(y) := \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} |\nabla V_\varepsilon(x - y)|, \end{aligned}$$

for almost every  $y \in \mathbb{R}^d$  can be proven using the usual Dominated Convergence Theorem. In particular, the growth estimate  $|\nabla V_1(z)| \leq C(1 + |z|)$  treats the integration  $\int_0^1 ds$ . On the other hand, for  $M_\varepsilon^\sigma(y)$ , the composition  $F'(tV_\varepsilon * \rho^\sigma(y) + (1 - t)V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1}(y))$  is bounded

uniformly in  $\sigma$  by Lemma A.3. We verify the last assumption of Theorem A.1 using Fubini and the change of variables  $z = -s\sigma\zeta(x) + y$ .

$$\begin{aligned} \int_{\mathbb{R}^d} g^\sigma(y)dy &= \|F'\|_{L^\infty(\{0, \|V_\varepsilon\|_{L^\infty})} \int_{\mathbb{R}^d} \int_0^1 |\nabla V_\varepsilon(x + s\sigma\zeta(x) - y)|dsdy \\ &= \|F'\|_{L^\infty(\{0, \|V_\varepsilon\|_{L^\infty})} \int_0^1 \int_{\mathbb{R}^d} |\nabla V_\varepsilon(x + s\sigma\zeta(x) - y)|dy ds \\ &= \|F'\|_{L^\infty(\{0, \|V_\varepsilon\|_{L^\infty})} \int_0^1 \int_{\mathbb{R}^d} |\nabla V_\varepsilon(x - z)|dz ds \\ &= \int_{\mathbb{R}^d} g(y)dy. \end{aligned}$$

Therefore, we can apply Theorem A.1 and (3.13) is established.

The 2-Wasserstein terms in (3.10): the treatment here is standard and we reproduce the proof in [7, Theorem 3.1] for completeness. Consider an optimal transport plan  $\gamma_{\tau,\varepsilon}^{n+1} \in \Gamma_o(\rho_{\tau,\varepsilon}^n, \rho_{\tau,\varepsilon}^{n+1})$  between  $\rho_{\tau,\varepsilon}^n$  and  $\rho_{\tau,\varepsilon}^{n+1}$ . By definition of  $d_W$ , we have

$$\begin{aligned} \frac{1}{2\tau} \left[ \frac{d_W^2(\rho_{\tau,\varepsilon}^n, \rho^\sigma) - d_W^2(\rho_{\tau,\varepsilon}^n, \rho_{\tau,\varepsilon}^{n+1})}{\sigma} \right] &\leq \frac{1}{2\tau\sigma} \iint_{\mathbb{R}^{2d}} (|x - P^\sigma(y)|^2 - |x - y|^2) d\gamma_{\tau,\varepsilon}^n(x, y) \\ &= \frac{1}{2\tau\sigma} \iint_{\mathbb{R}^{2d}} (|x - y - \sigma\zeta(y)|^2 - |x - y|^2) d\gamma_{\tau,\varepsilon}^n(x, y) \\ &= -\frac{1}{\tau} \iint_{\mathbb{R}^{2d}} (x - y) \cdot \zeta(y) d\gamma_{\tau,\varepsilon}^n(x, y) + o(\sigma), \end{aligned}$$

where in the last equality we applied a first order Taylor expansion. By sending  $\sigma$  to 0 and recalling (3.10), it holds

$$\frac{1}{\tau} \iint_{\mathbb{R}^{2d}} (x - y) \cdot \zeta(y) d\gamma_{\tau,\varepsilon}^n(x, y) \leq \int_{\mathbb{R}^d} \zeta(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})](x) d\rho_{\tau,\varepsilon}^{n+1}(x).$$

Repeating the same computation for  $\sigma \leq 0$ , we actually obtain an equality, that is, for  $\zeta = \nabla\varphi$

$$\frac{1}{\tau} \iint_{\mathbb{R}^{2d}} (x - y) \cdot \nabla\varphi(y) d\gamma_{\tau,\varepsilon}^n(x, y) = \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})](x) d\rho_{\tau,\varepsilon}^{n+1}(x). \tag{3.15}$$

Note that the Hölder estimate (3.8) and  $(x - y) \cdot \nabla\varphi(y) = \varphi(x) - \varphi(y) + o(|x - y|^2)$  imply

$$\frac{1}{\tau} \iint_{\mathbb{R}^{2d}} (x - y) \cdot \nabla\varphi(y) d\gamma_{\tau,\varepsilon}^n(x, y) = \frac{1}{\tau} \int_{\mathbb{R}^d} \varphi(x) d(\rho_{\tau,\varepsilon}^n - \rho_{\tau,\varepsilon}^{n+1})(x) + O(\tau).$$

Now, let  $0 \leq s < t$  be fixed, with

$$h = \left\lceil \frac{s}{\tau} \right\rceil + 1 \quad \text{and} \quad k = \left\lceil \frac{t}{\tau} \right\rceil.$$

Taking into account the last equality, by summing in (3.15) over  $j$  from  $h$  to  $k$ , we obtain

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(x) d\rho_{\tau,\varepsilon}^{k+1} - \int_{\mathbb{R}^d} \varphi(x) d\rho_{\tau,\varepsilon}^h + O(\tau^2) \\ = -\tau \sum_{j=h}^k \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_{\tau,\varepsilon}^{j+1})](x) d\rho_{\tau,\varepsilon}^{j+1}(x), \end{aligned}$$



which is equivalent to

$$\begin{aligned} & \int_{\mathbb{R}^d} \varphi(x) d\rho_\tau^\varepsilon(t)(x) - \int_{\mathbb{R}^d} \varphi(x) d\rho_\tau^\varepsilon(s)(x) + O(\tau^2) \\ &= - \int_s^t \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon(r))](x) d\rho_\tau^\varepsilon(r)(x) dr. \end{aligned} \tag{3.16}$$

It remains to pass the limit  $\tau \downarrow 0$  up to a subsequence for  $\rho_\tau^\varepsilon \rightharpoonup \tilde{\rho}^\varepsilon$  as in Proposition 3.1. More specifically, the result there states that  $\rho_\tau^\varepsilon$  narrowly converges uniformly in  $t \in [0, T]$  to  $\tilde{\rho}^\varepsilon$  as (a subsequence of)  $\tau \downarrow 0$ . Clearly, the left-hand side of (3.16) passes easily in the limit  $\tau \downarrow 0$  so we only focus on the right-hand side. Let us take the following statement for granted: for fixed  $\varepsilon > 0$  and almost every  $r \in [0, T]$ , we have

$$\begin{aligned} & \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon(r))](x) d\rho_\tau^\varepsilon(r)(x) \\ & \rightarrow \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \tilde{\rho}^\varepsilon(r))](x) d\tilde{\rho}^\varepsilon(r)(x), \\ & \tau \downarrow 0, \text{ almost every } r \in [0, T]. \end{aligned} \tag{3.17}$$

Passing to the limit  $\tau \downarrow 0$  on the right-hand side of (3.16) reduces to finding an  $L^1((s, t); dr)$  majorant, assuming (3.17) holds. By Young’s convolution inequality and Lemma A.3, we have

$$|\nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon(r))](x)| \leq \|\nabla V_\varepsilon\|_{L^1} \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})}.$$

Overall, this implies the uniform estimate in  $\tau > 0$

$$\left| \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon(r))](x) d\rho_\tau^\varepsilon(r)(x) \right| \leq \|\nabla\varphi\|_{L^\infty} \|\nabla V_\varepsilon\|_{L^1} \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})}.$$

Hence, we can pass to the limit  $\tau \downarrow 0$  in the right-hand side of (3.16) and conclude.

Let us prove (3.17). We fix  $r \in [0, T]$  and henceforth drop the explicit dependence on this variable. We add and subtract

$$\begin{aligned} & \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon)](x) d\rho_\tau^\varepsilon(x) - \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \tilde{\rho}^\varepsilon)](x) d\tilde{\rho}^\varepsilon(x) \\ &= \int_{\text{supp}\varphi} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon)](x) d\rho_\tau^\varepsilon(x) \end{aligned} \tag{3.18}$$

$$+ \int_{\text{supp}\varphi} \nabla\varphi(x) \cdot \nabla V_\varepsilon * [F'(V_\varepsilon * \tilde{\rho}^\varepsilon)](x) d[\rho_\tau^\varepsilon - \tilde{\rho}^\varepsilon](x). \tag{3.19}$$

Fix small  $\eta > 0$  and find  $R > 1$  large enough such that  $\int_{\mathbb{R}^d \setminus B_R} |\nabla V_\varepsilon(y)| dy < \eta$  where  $B_R$  denotes the open ball of radius  $R$  centred at the origin. We begin with the difference in (3.18) by expanding the convolution

$$\begin{aligned} |\nabla V_\varepsilon * [F'(V_\varepsilon * \rho_\tau^\varepsilon) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon)](x)| & \leq \int_{\mathbb{R}^d} |\nabla V_\varepsilon(y)| |F'(V_\varepsilon * \rho_\tau^\varepsilon(x - y)) \\ & \quad - F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x - y))| dy. \end{aligned}$$

Up to a further subsequence, Corollary A.1 gives

$$\sup_{x \in \text{supp}\varphi, y \in \bar{B}_R} |F'(V_\varepsilon * \rho_\tau^\varepsilon(x - y)) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x - y))| < \eta,$$

for  $\tau > 0$  sufficiently small. Hence,

$$\int_{B_R} |\nabla V_\varepsilon(y)| |F'(V_\varepsilon * \rho_\tau^\varepsilon(x - y)) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x - y))| dy < \|\nabla V_\varepsilon\|_{L^1} \eta. \tag{3.20}$$

Concerning the integral over  $\mathbb{R}^d \setminus B_R$ , we apply Lemma A.3 to obtain (uniformly in  $\tau > 0$ )

$$\int_{\mathbb{R}^d \setminus B_R} |\nabla V_\varepsilon(y)| |F'(V_\varepsilon * \rho_\tau^\varepsilon(x - y)) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x - y))| dy \leq 2 \|F'\|_{L^\infty([0, \|V_\varepsilon\|_{L^\infty})} \eta. \tag{3.21}$$

These inequalities imply that the integral in (3.18) can be made arbitrarily small in the limit  $\tau \downarrow 0$ .

Turning to the difference in (3.19), we only need to show that  $\nabla V_\varepsilon * [F'(V_\varepsilon * \tilde{\rho}^\varepsilon)]$  is continuous on  $\text{supp}\varphi$ . Then, we can appeal to the narrow convergence  $\rho_\tau^\varepsilon \rightarrow \tilde{\rho}^\varepsilon$  in duality with continuous and bounded functions. Suppose  $x^n \in \text{supp}\varphi$  is a sequence which converges to  $x \in \text{supp}\varphi$ , we compare the difference

$$\begin{aligned} & \nabla V_\varepsilon * F'(V_\varepsilon * \tilde{\rho}^\varepsilon)(x^n) - \nabla V_\varepsilon * F'(V_\varepsilon * \tilde{\rho}^\varepsilon)(x) \\ &= \int_{B_R} \nabla V_\varepsilon(y) [F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x^n - y)) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x - y))] dy \\ &+ \int_{\mathbb{R}^d \setminus B_R} \nabla V_\varepsilon(y) [F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x^n - y)) - F'(V_\varepsilon * \tilde{\rho}^\varepsilon(x - y))] dy. \end{aligned}$$

The integral over  $B_R$  can be made arbitrarily small as  $n \rightarrow \infty$  owing to the uniform continuity of  $F'(V_\varepsilon * \tilde{\rho}^\varepsilon(\cdot))$  from Corollary A.1 and integrability of  $\nabla V_\varepsilon$ . This is similar to what is done for (3.20). The other integral over  $\mathbb{R}^d \setminus B_R$  can be made arbitrarily small by the same argument for (3.21). □

### 4 Compactness in the limit $\varepsilon \downarrow 0$

This section discusses the construction of a limit  $\rho$  for a subsequence of  $\{\tilde{\rho}^\varepsilon\}_{\varepsilon>0}$ . The key estimate is Lemma 4.1 which we are able to prove for general functions  $F$  satisfying **(F1)**, **(F3)**, and the growth conditions **(F<sub>m</sub>)**.

Assumptions **(F1)**, **(F3)**, and **(F<sub>m</sub>)** cover all the power laws  $F(x) = \frac{1}{m-1}|x|^m$  for  $m > 1$  and  $F(x) = x \log x$  (corresponding to  $m = 1$ ). In the case  $m > 1$ , **(F<sub>m</sub>)** implies **(F2)** since  $F'$  can be extended to  $x = 0$ . More precisely, if  $F''$  satisfies the bounds in **(F<sub>m</sub>)** for some  $m > 1$ , then  $F''$  is locally integrable around 0. By the fundamental theorem of calculus,

$$F'(x) = F'(1) - \int_x^1 F''(t) dt, \quad \forall x > 0.$$

Owing to Lebesgue’s dominated convergence theorem, the right-hand side has a limit as  $x \downarrow 0$  and therefore so does the left-hand side which we call  $F'(0) := \lim_{x \downarrow 0} F'(x)$ .

**Remark 4.1** (Comments on **(F<sub>m</sub>)**) Combining **(F<sub>m</sub>)** with the assumption  $F(0) = 0$  from **(F1)** gives, for  $m > 1$ ,

$$\frac{c_1}{m(m-1)} x^m \leq F(x) - F'(0)x \leq \frac{c_2}{m(m-1)} x^m.$$

The inequalities above and the uniform bound for  $\mathcal{F}^\varepsilon[\tilde{\rho}^\varepsilon(t)]$  from Proposition 3.1 yield the following integrability estimate uniform in  $t \in [0, T]$  and  $\varepsilon > 0$

$$\begin{aligned} \|V_\varepsilon * \tilde{\rho}^\varepsilon(t)\|_{L^m(\mathbb{R}^d)}^m &\leq \frac{m(m-1)}{c_1} \mathcal{F}^\varepsilon[\tilde{\rho}^\varepsilon(t)] - \frac{m(m-1)}{c_1} F'(0) \\ &\leq \frac{m(m-1)}{c_1} \mathcal{F}^\varepsilon[\rho_0] - \frac{m(m-1)}{c_1} F'(0) \leq \frac{c_2}{c_1} \|V_\varepsilon * \rho_0\|_{L^m}^m \leq \frac{c_2}{c_1} \|\rho_0\|_{L^m}^m. \end{aligned} \tag{4.1}$$

Concerning the  $m = 1$  case, we directly estimate

$$\mathcal{H}[V_\varepsilon * \tilde{\rho}^\varepsilon(t)] = \mathcal{F}^\varepsilon[\tilde{\rho}^\varepsilon(t)] \leq \mathcal{F}^\varepsilon[\rho_0] = \mathcal{H}[V_\varepsilon * \rho_0] \leq \mathcal{H}[\rho_0]. \tag{4.2}$$

Here, we used Jensen’s inequality with the convex function  $x \log x$  and reference measure  $V_\varepsilon$  to obtain  $\mathcal{H}[V_\varepsilon * \rho_0] \leq \int V_\varepsilon * (\rho_0 \log \rho_0) = \int \rho \log \rho$  recalling  $\int V_\varepsilon = 1$  as well as Proposition 3.1.

The sequence of solutions  $\{\tilde{\rho}^\varepsilon\}_{\varepsilon>0}$  to (NLE) constructed in Sect. 3 is the candidate approximating *weak solution* of (DE). As  $\{\tilde{\rho}^\varepsilon\}_{\varepsilon>0}$  is in general a sequence of measures, it is useful to consider the regularised version,  $V_\varepsilon * \tilde{\rho}^\varepsilon$ . For brevity, we drop the tilde on  $\rho^\varepsilon$  from now on. First, we state compactness of  $\{\rho^\varepsilon\}_{\varepsilon>0}$  in  $C([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ .

**Proposition 4.1** *There exists an absolutely continuous curve  $\tilde{\rho} : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$  such that the sequence  $\{\rho^\varepsilon\}_{\varepsilon>0}$  admits a subsequence  $\{\rho^{\varepsilon_k}\}$  such that  $\rho^{\varepsilon_k}(t)$  narrow converges to  $\tilde{\rho}(t)$  for any  $t \in [0, T]$  as  $k \rightarrow +\infty$ .*

**Proof** The proof is exactly the same as in [7, Proposition 4.1] using a refined version of Ascoli-Arzelà [1, Proposition 3.3.1]. □

The narrow convergence proven in Proposition 4.1 is not sufficient to pass to the limit  $\varepsilon \downarrow 0$  from (NLE) to (DE). For this reason, we study the sequence  $v^\varepsilon(t) := V_\varepsilon * \rho^\varepsilon(t)$  for  $t \in [0, T]$  (we drop the subscript  $k$  for simplicity). We obtain higher regularity estimates uniform in  $\varepsilon$  by using the *flow interchange technique* developed by Matthes, McCann, and Savaré in [41]. The strategy is to compute the dissipation of  $\mathcal{F}^\varepsilon$  along a solution of an *auxiliary gradient flow*. This flow is chosen so that it satisfies an *Evolution Variational Inequality (EVI)* which allows us to obtain the desired estimate leading to compactness.

Since the seminal work of Jordan, Kinderlehrer, and Otto [37], it is known that the heat equation can be interpreted as the 2-Wasserstein gradient flow of the Boltzmann entropy  $\mathcal{H}$  (see below for the precise definition). Moreover the heat semigroup, denoted by  $S_{\mathcal{H}}$ , is a 0-flow in the following sense.

**Definition 4.1** ( $\lambda$ -flow) A semigroup  $S_\mathcal{E} : [0, +\infty] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathcal{P}_2(\mathbb{R}^d)$  is a  $\lambda$ -flow for a functional  $\mathcal{E} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{+\infty\}$  with respect to the distance  $d_W$  if, for an arbitrary  $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ , the curve  $t \mapsto S_\mathcal{E}^t \rho$  is absolutely continuous on  $[0, +\infty[$  and it satisfies the so-called Evolution Variational Inequality (EVI)

$$\frac{1}{2} \frac{d^+}{dt} d_W^2(S_\mathcal{E}^t \rho, \bar{\rho}) + \frac{\lambda}{2} d_W^2(S_\mathcal{E}^t \rho, \bar{\rho}) \leq \mathcal{E}(\bar{\rho}) - \mathcal{E}(S_\mathcal{E}^t \rho) \tag{4.3}$$

for all  $t \geq 0$ , with respect to every reference measure  $\bar{\rho} \in \mathcal{P}_2(\mathbb{R}^d)$  such that  $\mathcal{E}(\bar{\rho}) < \infty$ .

Below we use the flow interchange by considering the heat equation as an auxiliary flow with respect to the Boltzmann entropy

$$\mathcal{H}[\rho] = \begin{cases} \int_{\mathbb{R}^d} \rho(x) \log \rho(x) dx, & \rho \ll \text{Leb}(\mathbb{R}^d) \\ +\infty, & \text{otherwise} \end{cases}. \tag{4.4}$$

Again, when  $\rho$  is an absolutely continuous measure with respect to Lebesgue, we identify its density as  $\rho(x)$ .

**Remark 4.2** We remind the reader that  $\mathcal{H}[\rho]$  is bounded below by  $m_2(\rho)$ . This can be seen by looking at the *relative entropy* with respect to the standard Gaussian on  $\mathbb{R}^d$  denoted by  $\mathcal{M}(x) = (2\pi)^{-\frac{d}{2}} \exp\{-|x|^2/2\}$ . For any  $\rho \in \mathcal{P}_2^a(\mathbb{R}^d)$ , Jensen’s inequality with the convex function  $x \log x$  gives

$$\begin{aligned} \mathcal{H}[\rho | \mathcal{M}] &:= \int_{\mathbb{R}^d} \rho(x) \log \frac{\rho(x)}{\mathcal{M}(x)} dx = \int_{\mathbb{R}^d} \frac{\rho(x)}{\mathcal{M}(x)} \log \frac{\rho(x)}{\mathcal{M}(x)} \mathcal{M}(x) dx \\ &\geq \left( \int_{\mathbb{R}^d} \frac{\rho(x)}{\mathcal{M}(x)} \mathcal{M}(x) dx \right) \log \left( \int_{\mathbb{R}^d} \frac{\rho(x)}{\mathcal{M}(x)} \mathcal{M}(x) dx \right) = 0. \end{aligned}$$

This gives the lower bound for the entropy

$$\mathcal{H}[\rho] \geq \int_{\mathbb{R}^d} \rho(x) \log \mathcal{M}(x) dx = -\frac{d}{2} \log 2\pi - \frac{1}{2} m_2(\rho).$$

In the following, for any  $v \in \mathcal{P}_2(\mathbb{R}^d)$  such that  $\mathcal{H}(v) < +\infty$ , we denote by  $S_{\mathcal{H}}^t v$  the solution at time  $t$  of the heat equation coupled with an initial value  $v$  at  $t = 0$ . Moreover, for every  $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ , we define the dissipation of  $\mathcal{F}^\varepsilon$  along  $S_{\mathcal{H}}$  by

$$D_{\mathcal{H}} \mathcal{F}^\varepsilon(\rho) := \limsup_{s \downarrow 0} \left\{ \frac{\mathcal{F}^\varepsilon[\rho] - \mathcal{F}^\varepsilon[S_{\mathcal{H}}^s \rho]}{s} \right\}.$$

In order to prove stronger compactness, we begin with an  $L_t^2 H_x^1$  estimate on the  $\frac{m}{2}$  power of  $v_t^\varepsilon = V_\varepsilon * \rho_t^\varepsilon$ . This generalises Lemma 4.1 from [7].

**Lemma 4.1** *Suppose  $F$  satisfies (F1), (F3), and (F<sub>m</sub>) for some  $m \geq 1$ . Let  $\rho_0 \in \mathcal{P}_2^a(\mathbb{R}^d) \cap L^m(\mathbb{R}^d)$ . In the case  $m = 1$ , assume further  $\mathcal{H}[\rho_0] < +\infty$ . Then, there exists a constant  $C = C(\rho_0, V_1, T) > 0$  such that*

$$\sup_{\varepsilon, \tau > 0} \left\| (v_\tau^\varepsilon)^{\frac{m}{2}} \right\|_{L^2(0, T; H^1(\mathbb{R}^d))} \leq C.$$

**Proof** If  $m = 1$ , then the  $L_t^2 L_x^2$  bound simply reads

$$\left\| (v_\tau^\varepsilon)^{\frac{1}{2}} \right\|_{L^2(0, T; L^2(\mathbb{R}^d))}^2 = \int_0^T \int_{\mathbb{R}^d} V_\varepsilon * \rho_\tau^\varepsilon(t, x) dx dt = T,$$

since both  $\|V_\varepsilon\|_{L^1} = \int_{\mathbb{R}^d} d\rho_\tau^\varepsilon(t)(x) = 1$ . For  $m > 1$ , the estimate is very similar to that of (4.1) applied to the pre-limit curves  $\rho_\tau^\varepsilon$ ,

$$\left\| (v_\tau^\varepsilon)^{\frac{m}{2}} \right\|_{L^2([0, T]; L^2(\mathbb{R}^d))}^2 = \int_0^T \|V_\varepsilon * \rho_\tau^\varepsilon\|_{L^m}^m dt \leq \frac{c_2 T}{c_1} \|\rho_0\|_{L^m}^m.$$

The rest of this proof focuses on the uniform bound for  $\nabla(v_\tau^\varepsilon)^{\frac{m}{2}}$ . For  $s > 0$ , we take  $S_{\mathcal{H}}^s \rho_{\tau, \varepsilon}^{n+1}$  as a competitor against  $\rho_{\tau, \varepsilon}^{n+1}$  in the minimisation problem (3.1). We thus have

$$\frac{1}{2\tau} d_W^2(\rho_{\tau, \varepsilon}^{n+1}, \rho_{\tau, \varepsilon}^n) + \mathcal{F}^\varepsilon[\rho_{\tau, \varepsilon}^{n+1}] \leq \frac{1}{2\tau} d_W^2(S_{\mathcal{H}}^s \rho_{\tau, \varepsilon}^{n+1}, \rho_{\tau, \varepsilon}^n) + \mathcal{F}^\varepsilon[S_{\mathcal{H}}^s \rho_{\tau, \varepsilon}^{n+1}],$$

which, dividing by  $s > 0$  and passing to  $\limsup_{s \downarrow 0}$ , gives

$$\tau D_{\mathcal{H}} \mathcal{F}^\varepsilon(\rho_{\tau, \varepsilon}^{n+1}) \leq \frac{1}{2} \frac{d^+}{dt} \Big|_{t=0} \left( d_W^2(S_{\mathcal{H}}^t \rho_{\tau, \varepsilon}^{n+1}, \rho_{\tau, \varepsilon}^n) \right) \stackrel{(E.V.I.)}{\leq} \mathcal{H}[\rho_{\tau, \varepsilon}^n] - \mathcal{H}[\rho_{\tau, \varepsilon}^{n+1}]. \tag{4.5}$$

In the last inequality we used that  $S_{\mathcal{J}_C}$  is a 0-flow, cf. Definition 4.1. Now, let us focus on the left hand side of (4.5). Firstly, note that

$$\begin{aligned}
 D_{\mathcal{J}_C} \mathcal{F}^\varepsilon(\rho_{\tau,\varepsilon}^{n+1}) &= \limsup_{s \downarrow 0} \left\{ \frac{\mathcal{F}^\varepsilon[\rho_{\tau,\varepsilon}^{n+1}] - \mathcal{F}^\varepsilon[S_{\mathcal{J}_C}^s \rho_{\tau,\varepsilon}^{n+1}]}{s} \right\} \\
 &= \limsup_{s \downarrow 0} \int_0^1 \left( -\frac{d}{dz} \Big|_{z=st} \mathcal{F}^\varepsilon[S_{\mathcal{J}_C}^z \rho_{\tau,\varepsilon}^{n+1}] \right) dt.
 \end{aligned}
 \tag{4.6}$$

Thus, we now compute the time derivative inside the above integral. Using integration by parts, the  $C^\infty$  regularity of the heat semigroup, and  $(\mathbf{F}_m)$ , we have

$$\begin{aligned}
 \frac{d}{dt} \mathcal{F}^\varepsilon[S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1}] &= - \int_{\mathbb{R}^d} F''(V_\varepsilon * S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1}) |\nabla V_\varepsilon * S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1}|^2 dx \\
 &\leq -c_1 \int_{\mathbb{R}^d} (V_\varepsilon * S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1})^{m-2} |\nabla V_\varepsilon * S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1}|^2 dx \\
 &= -\frac{4c_1}{m^2} \int_{\mathbb{R}^d} \left| \nabla (V_\varepsilon * S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1}) \right|^{\frac{m}{2}} dx.
 \end{aligned}
 \tag{4.7}$$

The previous computation is justified since  $S_{\mathcal{J}_C}^t \rho_{\tau,\varepsilon}^{n+1} > 0$  everywhere on  $\mathbb{R}^d$  so there is no division by zero. By substituting (4.7) into (4.6), from (4.5) we obtain

$$\tau \liminf_{s \downarrow 0} \int_0^1 \int_{\mathbb{R}^d} \left| \nabla (V_\varepsilon * S_{\mathcal{J}_C}^{st} \rho_{\tau,\varepsilon}^{n+1}) \right|^{\frac{m}{2}}(x) dx dt \leq \frac{m^2}{4c_1} (\mathcal{H}[\rho_{\tau,\varepsilon}^n] - \mathcal{H}[\rho_{\tau,\varepsilon}^{n+1}]).$$

In order to pass to the limit  $s \downarrow 0$  for  $m > 1$ , we first deduce  $V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1} \in L^m$  by  $(\mathbf{F}_m)$  and Proposition 3.1. Second, by standard properties of the heat semigroup, we obtain  $V_\varepsilon * S_{\mathcal{J}_C}^{st} \rho_{\tau,\varepsilon}^{n+1} \rightarrow V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1}$  in  $L^m$  as  $s \downarrow 0$ . Notice that the first and second steps are immediate for  $m = 1$ . Third, by the inequality

$$\left| (V_\varepsilon * S_{\mathcal{J}_C}^{st} \rho_{\tau,\varepsilon}^{n+1})^{\frac{m}{2}} - (V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})^{\frac{m}{2}} \right|^2 \leq 2 \left( (V_\varepsilon * S_{\mathcal{J}_C}^{st} \rho_{\tau,\varepsilon}^{n+1})^m + (V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})^m \right),$$

we can apply Theorem A.1 to deduce  $(V_\varepsilon * S_{\mathcal{J}_C}^{st} \rho_{\tau,\varepsilon}^{n+1})^{\frac{m}{2}} \rightarrow (V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1})^{\frac{m}{2}}$  in  $L^2$  as  $s \downarrow 0$ . Finally, the weak  $L^2$  lower semi-continuity of the  $H^1$  semi-norm gives

$$\tau \int_{\mathbb{R}^d} \left| \nabla |V_\varepsilon * \rho_{\tau,\varepsilon}^{n+1}|^{\frac{m}{2}}(x) \right|^2 dx \leq \frac{m^2}{4c_1} (\mathcal{H}[\rho_{\tau,\varepsilon}^n] - \mathcal{H}[\rho_{\tau,\varepsilon}^{n+1}]).$$

By summing up over  $n$  from 0 to  $N - 1$ , taking into account Remark 4.2 and that second order moments are uniformly bounded (see Proposition 3.1), we get

$$\begin{aligned}
 \int_0^T \int_{\mathbb{R}^d} \left| \nabla |V_\varepsilon * \rho_\tau^\varepsilon(t)|^{\frac{m}{2}}(x) \right|^2 dx dt &\leq \frac{m^2}{4c_1} (\mathcal{H}[\rho_0] - \mathcal{H}[\rho_{\tau,\varepsilon}^n]) \\
 &\leq \frac{m^2}{4c_1} (\mathcal{H}[\rho_0] + C(\rho_0, V_1, T)).
 \end{aligned}
 \tag{4.8}$$

For  $m = 1$ , the initial entropy is assumed to be bounded. For  $m > 1$ , since  $x \log x \leq Cx^m$  for any  $x \geq 0$ , and some constant  $C > 0$ , we always have  $\mathcal{H}[\rho_0] \leq C \|\rho_0\|_{L^m}^m$ . In both cases, the initial entropy is bounded, and this establishes the desired  $L^2_{t,x}$  bound for  $\nabla(v_\tau^\varepsilon)^{\frac{m}{2}}$ .  $\square$

The strong  $L^m$  compactness in time and space follows by applying a refined version of the Aubin-Lions Lemma due to Rossi and Savaré [50, Theorem 2]. For the reader’s convenience we recall the latter result below before presenting the compactness result for  $\{v^{\varepsilon_k}\}_k$ .

**Proposition 4.2** [50, Theorem 2] *Let  $X$  be a separable Banach space. Consider*

- *a lower semicontinuous functional  $\mathcal{F} : X \rightarrow [0, +\infty]$  with relatively compact sublevels in  $X$ ;*
- *a pseudo-distance  $g : X \times X \rightarrow [0, +\infty]$ , i.e.,  $g$  is lower semicontinuous and such that  $g(\rho, \eta) = 0$  for any  $\rho, \eta \in X$  with  $\mathcal{F}(\rho) < \infty, \mathcal{F}(\eta) < \infty$  implies  $\rho = \eta$ .*

*Let  $U$  be a set of measurable functions  $u : (0, T) \rightarrow X$ , with a fixed  $T > 0$ . Assume further that*

$$\sup_{u \in U} \int_0^T \mathcal{F}(u(t)) dt < \infty \quad \text{and} \quad \lim_{h \downarrow 0} \sup_{u \in U} \int_0^{T-h} g(u(t+h), u(t)) dt = 0. \tag{4.9}$$

*Then  $U$  contains an infinite sequence  $(u_n)_{n \in \mathbb{N}}$  that converges in measure, with respect to  $t \in (0, T)$ , to a measurable  $\tilde{u} : (0, T) \rightarrow X$ , i.e.*

$$\lim_{n \rightarrow \infty} |\{t \in (0, T) : \|u_n(t) - u(t)\|_X \geq \sigma\}| = 0, \quad \forall \sigma > 0.$$

The two conditions in (4.9) are called *tightness* and *weak integral equicontinuity*, respectively.

**Proposition 4.3** *Fix  $m \geq 1$  and consider the family  $\{v_\tau^\varepsilon\}_{\varepsilon \in (0, \varepsilon_0), \tau > 0}$  in Lemma 4.1. There is a subsequence  $\tau_k \downarrow 0$  such that for any  $\varepsilon > 0$ , we have*

$$v_{\tau_k}^\varepsilon \rightarrow v^\varepsilon = V_\varepsilon * \tilde{\rho}^\varepsilon, \quad \text{in } L^m([0, T] \times \mathbb{R}^d).$$

*Moreover, there is a subsequence  $\varepsilon_k \downarrow 0$  and a curve  $v \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d)) \cap L^m([0, T] \times \mathbb{R}^d)$  such that*

$$v^\varepsilon \rightarrow v, \quad \text{in } L^m([0, T] \times \mathbb{R}^d).$$

**Proof** The proof of the result is obtained by applying Proposition 4.2 to a subset of the sequence  $U := \{v_\tau^\varepsilon\}_{\varepsilon \in (0, \varepsilon_0), \tau > 0}$  for  $X := L^m(\mathbb{R}^d)$  and  $g := d_1$  being the 1-Wasserstein distance — extended to  $+\infty$  outside of  $\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ . As for the functional, we consider  $\mathcal{F} : L^m(\mathbb{R}^d) \rightarrow [0, +\infty]$  defined by

$$\mathcal{F}[v] = \begin{cases} \|v\|_{H^1(\mathbb{R}^d)}^2 + \int_{\mathbb{R}^d} |x|v(x) dx, & \text{if } v \in \mathcal{P}_1(\mathbb{R}^d) \text{ and } v^{\frac{m}{2}} \in H^1(\mathbb{R}^d); \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that elements in the domain of the functional  $\mathcal{F}$  belong to  $\mathcal{P}_1(\mathbb{R}^d)$ , thus  $0 = g(\rho, \eta) = d_1(\rho, \eta)$  implies  $\rho = \eta$ . Let us check that  $\mathcal{F}$  is an admissible functional.

Lower semicontinuity can be easily verified following, e.g., [7]. Let  $A_c := \{v \in L^m(\mathbb{R}^d) : \mathcal{F}[v] \leq c\}$  be a sublevel of  $\mathcal{F}$ , where  $c$  is a positive constant. We consider  $B_c := \{w = v^{\frac{m}{2}} : v \in A_c\}$  and prove that  $B_c$  is relatively compact in  $L^2(\mathbb{R}^d)$ , as the map  $w \in L^2(\mathbb{R}^d) \mapsto \iota(w) = w^{\frac{2}{m}} \in L^m(\mathbb{R}^d)$  is continuous and  $A_c = \iota(B_c)$ .

The Riesz-Fréchet-Kolmogorov theorem provides relative compactness in  $L^2(\mathbb{R}^d)$  of  $B_c$ . In fact, elements of  $B_c$  are bounded in  $L^2(\mathbb{R}^d)$  and it holds the uniform continuity estimate

$$\begin{aligned}
 & \int_{\mathbb{R}^d} |w(x+h) - w(x)|^2 dx \\
 &= \int_{\mathbb{R}^d} \left| \int_0^1 \frac{d}{d\tau} w(x+\tau h) d\tau \right|^2 dx = \int_{\mathbb{R}^d} \left| \int_0^1 h \cdot \nabla w(x+\tau h) d\tau \right|^2 dx \quad (4.10) \\
 &\leq |h|^2 \int_{\mathbb{R}^d} \int_0^1 |\nabla w(x+\tau h)|^2 d\tau dx = |h|^2 \|\nabla w\|_{L^2(\mathbb{R}^d)}^2,
 \end{aligned}$$

which implies  $\|w(\cdot+h) - w(\cdot)\|_{L^2(\mathbb{R}^d)} \rightarrow 0$  as  $h \rightarrow 0^+$ .

Before proceeding to the uniform integrability, we record the following improved estimates afforded to us by the fact that  $B_c$  is a bounded subset of  $H^1(\mathbb{R}^d)$ .

$$\sup_{w \in B_c} \|w\|_{L^q(\mathbb{R}^d)} \leq c, \quad q \in \begin{cases} \{+\infty\} & d = 1 \\ [2, +\infty) & d = 2 \\ [2, \frac{2d}{d-2}] & d > 2 \end{cases}. \quad (4.11)$$

In the case  $d = 1$ , for any  $m \geq 1$ , we set  $\delta = 1$  in the following estimate

$$\begin{aligned}
 \|w\|_{L^2(\mathbb{R}^d \setminus B_R)}^2 &= \int_{|x| \geq R} |v(x)|^m dx \leq \frac{1}{R^\delta} \int_{\mathbb{R}^d} |x|^\delta |v(x)|^m dx \\
 &\leq \frac{\|v\|_{L^\infty}^{m-1}}{R} \int_{\mathbb{R}^d} |x|v(x) dx \leq \frac{\|v\|_{L^\infty}^{m-1}}{R} \mathcal{F}[v] \leq \frac{\|v\|_{L^\infty}^{m-1}}{R} c. \quad (4.12)
 \end{aligned}$$

Hence, uniform integrability is proven in the case  $d = 1$  and  $m \geq 1$ . In fact, for any  $d \geq 2$  and  $m = 1$ , we can simply take  $\delta = 1$  again in (4.12) to establish uniform integrability in this case. For general  $d \geq 2$  and  $m > 1$ , we further develop (4.12) by Hölder’s inequality to obtain, for a particular choice of  $\delta \in (0, 1)$  which will be made clear,

$$\|w\|_{L^2(\mathbb{R}^d \setminus B_R)}^2 \leq \frac{1}{R^\delta} \left( \int_{\mathbb{R}^d} |x|v(x) dx \right)^\delta \left( \int_{\mathbb{R}^d} |v(x)|^{\frac{m-\delta}{1-\delta}} dx \right)^{1-\delta}. \quad (4.13)$$

The parameter  $\delta \in (0, 1)$  can be chosen to take advantage of the extra integrability from (4.11). For example, we can take

$$\delta = \frac{2}{d(m-1) + 2} \in (0, 1),$$

which is permissible in light of the Sobolev embedding (4.11)  $\int_{\mathbb{R}^d} |v(x)|^{\frac{m-\delta}{1-\delta}} dx \leq c$  recalling  $v = w^{\frac{2}{m}}$ . Thus, (4.13) yields the uniform integrability of  $w$  in  $L^2$ .

We now check tightness and weak integral equicontinuity, i.e. conditions (4.9). Let us set  $U := \{v_\tau^\varepsilon\}_{0 < \varepsilon \leq \varepsilon_0, 0 < \tau,}$  being  $v_\tau^\varepsilon : [0, T] \rightarrow L^m(\mathbb{R}^d)$  the sequence defined above by  $v_\tau^\varepsilon = V_\varepsilon * \rho_\tau^\varepsilon$ , which satisfies Lemma 4.1. For any  $0 < \varepsilon \leq \varepsilon_0$  and  $\tau > 0$ , it holds

$$\begin{aligned}
 \int_0^T \mathcal{F}[v_\tau^\varepsilon(t)] dt &= \int_0^T \left\| (v_\tau^\varepsilon)^{\frac{m}{2}} \right\|_{H^1(\mathbb{R}^d)}^2 dt + \int_0^T \int_{\mathbb{R}^d} |x|v_\tau^\varepsilon(x) dx dt \\
 &\leq C(\rho_0, V_1, T) + \varepsilon_0 T \int_{\mathbb{R}^d} V_1(z)|z| dz < +\infty,
 \end{aligned}$$

where we used

$$\begin{aligned} \int_{\mathbb{R}^d} |x|v_\tau^\varepsilon(x) dx &= \iint_{\mathbb{R}^{2d}} |x|V_\varepsilon(x-y) d\rho_\tau^\varepsilon(y) dx \\ &\leq \iint_{\mathbb{R}^{2d}} V_\varepsilon(x-y)|x-y| d\rho_\tau^\varepsilon(y) dx + \iint_{\mathbb{R}^{2d}} V_\varepsilon(x-y)|y| d\rho_\tau^\varepsilon(y) dx \\ &= \varepsilon \int_{\mathbb{R}^d} V_1(z)|z| dz + \int_{\mathbb{R}^d} V_1(z) dz \int_{\mathbb{R}^d} |y| d\rho_\tau^\varepsilon(y) \\ &\leq \varepsilon_0 \int_{\mathbb{R}^d} V_1(z)|z| dz + \sqrt{m_2(\rho_\tau^\varepsilon)} \int_{\mathbb{R}^d} V_1(z) dz < +\infty. \end{aligned}$$

due to Remark 3.1 and Proposition 3.1. Taking the supremum in  $U$  we have tightness. For the weak integral equicontinuity, we fix  $\varepsilon, h > 0$  and consider the  $\tau \leq h$  and  $\tau > h$  cases separately. Starting with  $\tau \leq h$ , we use the almost Hölder continuity of  $\rho_\tau^\varepsilon$  proven in (3.8) of Proposition 4.1. More precisely, it holds

$$\begin{aligned} \int_0^{T-h} d_1(v_\tau^\varepsilon(t+h), v_\tau^\varepsilon(t)) dt &\leq \int_0^{T-h} d_W(v_\tau^\varepsilon(t+h), v_\tau^\varepsilon(t)) dt \\ &\leq \int_0^{T-h} d_W(\rho_\tau^\varepsilon(t+h), \rho_\tau^\varepsilon(t)) dt \\ &\leq c \int_0^{T-h} (\sqrt{h} + \sqrt{\tau}) dt \leq 2c(T-h)\sqrt{h}. \end{aligned}$$

where in the intermediate inequalities we used (3.8) for some constant  $c > 0$  (independent of  $\varepsilon, \tau, h$ ) as well as standard properties of Wasserstein distances, c.f. for example [54, Section 5.1]. The equicontinuity follows by sending  $h \downarrow 0$ . In the case  $\tau > h$ , we use (3.7) instead to estimate

$$\begin{aligned} \int_0^{T-h} d_1(v_\tau^\varepsilon(t+h), v_\tau^\varepsilon(t)) dt &\leq \int_0^{T-h} d_W(\rho_\tau^\varepsilon(t+h), \rho_\tau^\varepsilon(t)) dt \leq h \sum_{n=0}^{N-1} d_W(\rho_{\tau,\varepsilon}^{n+1}, \rho_{\tau,\varepsilon}^n) \\ &\leq hN^{\frac{1}{2}} \left( \sum_{n=0}^{N-1} d_W^2(\rho_{\tau,\varepsilon}^{n+1}, \rho_{\tau,\varepsilon}^n) \right)^{\frac{1}{2}} \leq cT^{\frac{1}{2}}h, \end{aligned}$$

where the constant  $c$  is defined when proving (3.7).

We are left to prove the relative compactness in  $L^m([0, T]; L^m(\mathbb{R}^d))$  for all  $m \geq 1$ . We start with the limit  $\tau \downarrow 0$  for fixed  $\varepsilon > 0$ . Remember that the estimates we have proven so far are uniform in  $\varepsilon$  and  $\tau$  so there is no dependence on  $\varepsilon$  as  $\tau \downarrow 0$ . We begin with the  $m > 1$  case. The first part in the proof of Lemma 4.1 showed that  $\|v_\tau^\varepsilon\|_{L^m([0, T] \times \mathbb{R}^d)}$  is uniformly bounded. Thus there exists a subsequence  $\tau_k \downarrow 0$  such that  $v_{\tau_k}^\varepsilon \rightharpoonup v^\varepsilon$  in  $L^m([0, T] \times \mathbb{R}^d)$  for some  $v^\varepsilon \in L^m([0, T] \times \mathbb{R}^d)$ . By Proposition 3.1, we know that  $\rho_{\tau_k}^\varepsilon$  narrowly converges to  $\tilde{\rho}^\varepsilon$  along a subsequence uniformly in  $[0, T]$ . By testing against smooth functions, we must have agreement between these limits  $v^\varepsilon = V_\varepsilon * \tilde{\rho}^\varepsilon$ . Moreover, along a further subsequence which we just label  $\tau \downarrow 0$ , we can apply Proposition 4.2 giving

$$\lim_{\tau \downarrow 0} \left| \{t \in (0, T) : \|v_\tau^\varepsilon(t) - v^\varepsilon(t)\|_{L^m(\mathbb{R}^d)} \geq \sigma\} \right| = 0, \quad \forall \sigma > 0.$$



Let us denote the set above by  $A_\sigma(\tau)$ . For arbitrary  $\sigma > 0$ , we have

$$\begin{aligned} \|v_\tau^\varepsilon - v^\varepsilon\|_{L^m([0, T] \times \mathbb{R}^d)}^m &= \int_0^T \int_{\mathbb{R}^d} |v_\tau^\varepsilon - v^\varepsilon|^m = \left( \int_{A_\sigma(\tau)} + \int_{[0, T] \setminus A_\sigma(\tau)} \right) \int_{\mathbb{R}^d} |v_\tau^\varepsilon - v^\varepsilon|^m \\ &\leq \sup_{s \in [0, T]} 2^{(m-1)} (\|v_\tau^\varepsilon(s)\|_{L^m}^m + \|v^\varepsilon(s)\|_{L^m}^m) |A_\sigma(\tau)| + \sigma^m T. \end{aligned}$$

Similar to (4.1), we can insert

$$\sup_{\varepsilon, \tau > 0, t \in [0, T]} \|v_\tau^\varepsilon\|_{L^m}^m \leq \frac{C_2}{c_1} \|\rho_0\|_{L^m}^m, \quad \text{and} \quad \sup_{\varepsilon > 0, t \in [0, T]} \|v^\varepsilon\|_{L^m}^m \leq \frac{C_2}{c_1} \|\rho_0\|_{L^m}^m$$

into the previous estimate to obtain

$$\|v_\tau^\varepsilon - v^\varepsilon\|_{L^m([0, T] \times \mathbb{R}^d)}^m \leq 2^m \frac{C_2}{c_1} \|\rho_0\|_{L^m}^m |A_\sigma(\tau)| + \sigma^m T.$$

Passing to  $\tau \downarrow 0$  and using  $\lim_{\tau \downarrow 0} |A_\sigma(\tau)| = 0$ , we arrive at

$$\limsup_{\tau \downarrow 0} \|v_\tau^\varepsilon - v^\varepsilon\|_{L^m([0, T] \times \mathbb{R}^d)} \leq \sigma T^{\frac{1}{m}}.$$

Since  $\sigma > 0$  was arbitrary, this implies the strong  $L^m$  convergence from  $v_\tau^\varepsilon$  to  $v^\varepsilon = V_\varepsilon * \tilde{\rho}^\varepsilon$ .

In the case  $m = 1$ , we need to argue differently. We apply [50, Proposition 1.10] which asserts that relative compactness in  $L^1((0, T); X)$  is implied by uniform integrability and relative compactness in measure as a function with values in  $X$  ( $X \equiv L^1(\mathbb{R}^d)$  in this proof). Compactness in measure has just been proven as an application of Proposition 4.2. Following [50, Remark 1.11], uniform integrability is a consequence of the strong integral equicontinuity

$$\lim_{h \rightarrow 0} \sup_{\tau} \int_0^{T-h} \left| \|v_\tau^\varepsilon(t+h)\|_{L^1(\mathbb{R}^d)} - \|v_\tau^\varepsilon(t)\|_{L^1(\mathbb{R}^d)} \right| dt = 0,$$

where we used that  $\|v_\tau^\varepsilon(t)\|_{L^1(\mathbb{R}^d)} = 1$  for any  $t \in [0, T]$  and  $\varepsilon > 0$ .

Strong compactness  $\varepsilon \downarrow 0$ : We first claim that the estimate in Lemma 4.1 also holds uniformly for  $v^\varepsilon$ , namely

$$\sup_{\varepsilon > 0} \left\| (v^\varepsilon)^{\frac{m}{2}} \right\|_{L^2(0, T; H^1(\mathbb{R}^d))} \leq C(\rho_0, V_1, T). \tag{4.14}$$

This can be seen by the fact that, up to a further subsequence,  $(v_\tau^\varepsilon)^{\frac{m}{2}}$  converges to  $(v^\varepsilon)^{\frac{m}{2}}$  strongly in  $L^2([0, T] \times \mathbb{R}^d)$ . Indeed, by standard results in  $L^p$  integration theory and the fact that  $v_\tau^\varepsilon \rightarrow v^\varepsilon$  strongly in  $L^m$ , there exists  $w^\varepsilon \in L^m([0, T] \times \mathbb{R}^d)$  such that, along a subsequence,  $|v_\tau^\varepsilon| \leq w^\varepsilon$  for almost every  $(t, x) \in [0, T] \times \mathbb{R}^d$ . Moreover, we have  $v_\tau^\varepsilon \rightarrow v^\varepsilon$  pointwise almost everywhere in  $[0, T] \times \mathbb{R}^d$ . Using Lebesgue’s dominated convergence theorem, we obtain

$$\int_0^T \int_{\mathbb{R}^d} \left| (v_\tau^\varepsilon)^{\frac{m}{2}} - (v^\varepsilon)^{\frac{m}{2}} \right|^2 \rightarrow 0,$$

since the integrand converges to 0 pointwise almost everywhere and it is majorised, uniformly in  $\tau$ , by

$$\left| (v_\tau^\varepsilon)^{\frac{m}{2}} - (v^\varepsilon)^{\frac{m}{2}} \right|^2 \leq 2((w^\varepsilon)^m + (v^\varepsilon)^m) \in L^1([0, T] \times \mathbb{R}^d).$$

Owing to the (weak  $L^2$ ) lower semicontinuity of the  $H^1$  seminorm, the estimate in Lemma 4.1 passes to the limit (along a subsequence)  $\tau \downarrow 0$  and (4.14) is established.

At this point, we can repeat all of the previous argument for  $U = \{v^\varepsilon\}_{\varepsilon \in (0, \varepsilon_0)}$ . We take the same space  $X = L^m(\mathbb{R}^d)$  and  $g = d_1$ . The same functional  $\mathcal{F}$  is still admissible. Tightness and weak integral equicontinuity can be analogously proven.

Proposition 4.2 applies and we have convergence in measure for  $v^\varepsilon$  to some curve  $v$  described in the statement of this result. By the same arguments as before, this convergence is strong in  $L^m([0, T]; L^m(\mathbb{R}^d))$ . □

### 5 Convergence of solutions

This section addresses the proof of Theorem 2.2. We cover the case  $m \geq 2$  in Sect. 5.1 while the case  $1 < m < 2$  is treated in Sect. 5.2. To simplify the presentation, we focus on functionals  $\mathcal{F} = \mathcal{H}_m$  but we also discuss (see Remark 5.2) the extension to general energies satisfying (F1), (F2), (F3), and (F<sub>m</sub>) for convergence from (NLE) to (DE).

#### 5.1 The case $m \geq 2$

Building on the previous discussions from Sects. 3 and 4, we denote  $\rho^\varepsilon$  the weak measure solutions to (NLE-m) constructed from the JKO scheme in Proposition 3.1. Moreover, we focus on the subsequence such that  $v^\varepsilon = V_\varepsilon * \rho^\varepsilon$  converges to  $v \in C([0, T]; \mathcal{P}_2(\mathbb{R}^d)) \cap L^m([0, T] \times \mathbb{R}^d)$  in  $L^m([0, T] \times \mathbb{R}^d)$  from Proposition 4.3. Starting from the definition of weak measure solution to (NLE-m) we can reformulate the right-hand side as follows:

$$\begin{aligned}
 & \int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) - \int_{\mathbb{R}^d} \varphi(x) d\rho_0(x) \\
 &= -\frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \cdot \nabla V_\varepsilon * (V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) d\rho_r^\varepsilon(x) dr \\
 &= -\frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} (V_\varepsilon * \rho_r^\varepsilon \nabla \varphi)(x) \nabla (V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) dx dr \\
 &= -2 \int_0^t \int_{\mathbb{R}^d} (V_\varepsilon * \rho_r^\varepsilon \nabla \varphi)(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}-1} \nabla (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}} dx dr \\
 &= -2 \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) \nabla (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr \\
 &\quad - 2 \int_0^t \int_{\mathbb{R}^d} z_r^\varepsilon(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}-1} \nabla (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr,
 \end{aligned} \tag{5.1}$$

being for any  $r \in [0, T]$  and  $x \in \mathbb{R}^d$ , the error term

$$z_r^\varepsilon(x) := (V_\varepsilon * \rho_r^\varepsilon \nabla \varphi)(x) - \nabla \varphi(x) (V_\varepsilon * \rho_r^\varepsilon)(x). \tag{5.2}$$

The product  $(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}} \nabla (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}$  in the last line of (5.1) is a weak-strong convergence pair in  $L^2_{t,x}$ . Indeed, recall the uniform  $L^2_t H^1_x$  bound on  $(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}$  and Proposition 4.2 from Sect. 4. Hence, the first integral in the last line of (5.1) passes well in the limit (along a subsequence)  $\varepsilon \rightarrow 0$ . This is precised later in the full proof of Theorem 2.2 so we dedicate much of this section to estimates proving that the error vanishes as  $\varepsilon \rightarrow 0$ .

**Remark 5.1** If  $V_1$  is compactly supported, the argument that the last term in (5.1) vanishes as  $\varepsilon \downarrow 0$  can be simplified based on the arguments in Sect. 5.2. In the rest of this subsection however, we present a general argument allowing for  $V_1$  with unbounded support.

Notice that the last term in the last equality of (5.1) can be estimated as

$$\|z^\varepsilon (v^\varepsilon)^{\frac{m}{2}-1} \nabla (v^\varepsilon)^{\frac{m}{2}}\|_{L^1([0,t] \times \mathbb{R}^d)} \leq \|z^\varepsilon\|_{L^m([0,t] \times \mathbb{R}^d)} \|(v^\varepsilon)^{\frac{m}{2}-1}\|_{L^q([0,t] \times \mathbb{R}^d)} \|\nabla (v^\varepsilon)^{\frac{m}{2}}\|_{L^2([0,t] \times \mathbb{R}^d)},$$

for  $q = \frac{2m}{m-2}$ , so that  $\frac{1}{m} + \frac{1}{q} + \frac{1}{2} = 1$  and

$$\|(v^\varepsilon)^{\frac{m}{2}-1}\|_{L^q([0,t] \times \mathbb{R}^d)} = \|v^\varepsilon\|_{L^m([0,t] \times \mathbb{R}^d)}^{m\left(\frac{m-2}{2}\right)} \leq c(T, V_1, \rho_0, m).$$

Notice that the exponent  $q$  is only valid for  $m \geq 2$  based on the computations above. In order to obtain a solution of (PME) in the  $\varepsilon \rightarrow 0^+$  limit, we need to prove that  $z^\varepsilon \rightarrow 0$  in  $L^m([0, T] \times \mathbb{R}^d)$ , for any  $t \in [0, T]$ . In turn, this will imply the error term in (5.1) vanishes as  $\varepsilon \rightarrow 0$ , as a consequence of the  $L^2$  version of Lebesgue dominated convergence theorem and weak- $L^2$  convergence. More precisely, we note that the product  $z^\varepsilon (v^\varepsilon)^{\frac{m}{2}-1} \in L^2([0, t] \times \mathbb{R}^d)$  and it converges to 0 strongly in  $L^2$ .

**Lemma 5.1** *There exists a subsequence  $\varepsilon_k \downarrow 0$  such that the error term  $z^{\varepsilon_k}$  converges to zero in  $L^m([0, T] \times \mathbb{R}^d)$  as  $k \rightarrow \infty$ .*

**Proof** First we notice that for any  $t \in [0, T]$  and  $\varphi \in C_c^2(\mathbb{R}^d)$  it holds

$$\begin{aligned} \int_{\mathbb{R}^d} |z_t^\varepsilon(x)| dx &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} V_\varepsilon(x-y) |\nabla \varphi(y) - \nabla \varphi(x)| d\rho_t^\varepsilon(y) dx \\ &\leq \|D^2 \varphi\|_\infty \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} V_\varepsilon(x-y) |y-x| d\rho_t^\varepsilon(y) dx \\ &= \varepsilon \|D^2 \varphi\|_\infty \int_{\mathbb{R}^d} |z| V_1(z) dz, \end{aligned}$$

by means of the change of variable  $z = \frac{x-y}{\varepsilon}$ . Therefore, there exists a constant  $C(V_1, \varphi)$  such that  $\|z^\varepsilon\|_{L^\infty([0,T]; L^1(\mathbb{R}^d))} \leq \varepsilon C(V_1, \varphi)$ , whence, up to passing to a subsequence,  $z_t^\varepsilon(x) \rightarrow 0$  for a.e.  $(t, x) \in [0, T] \times \mathbb{R}^d$ . We now find a majorant to apply the  $L^p$  version of the generalised Lebesgue dominated convergence theorem.

For almost every  $x \in \mathbb{R}^d$  and  $t \in [0, T]$ , for  $i = 1, \dots, d$ , the non-negativity of  $V_1$  and  $\rho_t^\varepsilon$  gives

$$\left| \int_{\mathbb{R}^d} V_\varepsilon(x-y) \partial_{x_i} \varphi(y) d\rho_t^\varepsilon(y) \right| \leq \int_{\mathbb{R}^d} V_\varepsilon(x-y) |\partial_{x_i} \varphi(y)| d\rho_t^\varepsilon(y) \leq \|\partial_{x_i} \varphi\|_\infty v_t^\varepsilon(x),$$

whence

$$|z_t^\varepsilon(x)| \leq 2 \|\nabla \varphi\|_\infty |v_t^\varepsilon(x)|.$$

Since  $v^\varepsilon \in L^m([0, t] \times \mathbb{R}^d)$  and it converges strongly in  $L^m$ , c.f. Proposition 4.3, we are able to conclude the result, as aforementioned.  $\square$

**Lemma 5.2** *For any  $t \in [0, T]$  and any  $\varphi \in C_c^1(\mathbb{R}^d)$  it holds*

$$\lim_{\varepsilon \rightarrow 0^+} \int_{\mathbb{R}^d} \varphi(x) v_t^\varepsilon(x) dx = \int_{\mathbb{R}^d} \varphi(x) d\tilde{\rho}(t).$$

**Proof** For any  $t \in [0, T]$  and any  $\varphi \in C_c^1(\mathbb{R}^d)$ , by using the definition of  $v_t^\varepsilon$  we obtain:

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \varphi(x) v_t^\varepsilon(x) dx - \int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) \right| &= \left| \int_{\mathbb{R}^d} \varphi(x) (V_\varepsilon * \rho_t^\varepsilon)(x) dx - \int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) \right| \\ &= \left| \int_{\mathbb{R}^d} (\varphi * V_\varepsilon)(x) d\rho_t^\varepsilon(x) - \int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) \right| \\ &= \left| \int_{\mathbb{R}^d} [(\varphi * V_\varepsilon)(x) - \varphi(x)] d\rho_t^\varepsilon(x) \right| \\ &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\varphi(x-y) - \varphi(x)| V_\varepsilon(y) dy d\rho_t^\varepsilon(x) \\ &\leq \|\nabla\varphi\|_\infty \int_{\mathbb{R}^d} |y| V_\varepsilon(y) dy \\ &= \varepsilon \|\nabla\varphi\|_\infty \int_{\mathbb{R}^d} |x| V_1(x) dx, \end{aligned}$$

which converges to 0 as  $\varepsilon \rightarrow 0^+$  since  $\int_{\mathbb{R}^d} |x| V_1(x) dx < +\infty$ . In the second last estimate, we used the mean-value inequality  $|\varphi(x-y) - \varphi(x)| \leq \|\nabla\varphi\|_\infty |y|$ .  $\square$

We now have all the information to prove Theorem 2.2 in the case  $\mathcal{F} = \mathcal{H}_m$  for  $m \geq 2$ . Proof of Theorem 2.2 for  $\mathcal{F} = \mathcal{H}_m$  and  $m \geq 2$  Since  $\rho^\varepsilon$  is a weak solution to (NLE-m), for any  $\varphi \in C_c^1(\mathbb{R}^d)$  and  $t \in [0, T]$  it satisfies

$$\begin{aligned} &\int_{\mathbb{R}^d} \varphi(x) d\rho_t^\varepsilon(x) - \int_{\mathbb{R}^d} \varphi(x) d\rho_0(x) \\ &= -2 \int_0^t \int_{\mathbb{R}^d} \nabla\varphi(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) \nabla(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr \\ &\quad - 2 \int_0^t \int_{\mathbb{R}^d} z_r^\varepsilon(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}-1} \nabla(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr, \end{aligned}$$

as explained in (5.1), Proposition 4.1, Lemma 4.1, Lemma 5.2, and Lemma 4.3 infer existence of a subsequence of  $\rho^\varepsilon(t)$  narrowly converging to  $\tilde{\rho} \in L^m([0, T]; L^m(\mathbb{R}^d))$ , and, in particular,  $\{v^\varepsilon\}_\varepsilon$  admits a subsequence such that

$$\begin{aligned} v^{\varepsilon_k} &\rightarrow \tilde{\rho} \quad \text{in } L^m([0, T]; L^m(\mathbb{R}^d)); \\ \nabla(v^{\varepsilon_k})^{\frac{m}{2}} &\rightharpoonup w \quad \text{in } L^2([0, T]; L^2(\mathbb{R}^d)). \end{aligned}$$

By a standard argument one can show that  $(v^{\varepsilon_k})^{\frac{m}{2}} \rightarrow (\tilde{\rho})^{\frac{m}{2}}$  in  $L^2([0, T]; L^2(\mathbb{R}^d))$ , whence  $w \equiv \nabla(\tilde{\rho})^{\frac{m}{2}}$ . Before letting  $\varepsilon \rightarrow 0^+$  and obtaining the result we need to further regularise the test function,  $\varphi$ , since in Lemma 5.1 we make use of test functions in  $C_c^2(\mathbb{R}^d)$ . In this regard, we consider a standard mollifier  $\eta \in C_c^\infty(\mathbb{R}^d)$  and the corresponding sequence  $\varphi^\sigma := \eta^\sigma * \varphi \in C_c^\infty(\mathbb{R}^d)$ , being  $\eta^\sigma(x) = \sigma^{-d} \eta(x/\sigma^d)$  for any  $x \in \mathbb{R}^d$  and  $\sigma > 0$ . As a consequence of the observations above and Lemma 5.1, by letting  $\varepsilon \rightarrow 0^+$  we obtain, for any  $\sigma > 0$  and  $t \in [0, T]$ ,

$$\begin{aligned} &\int_{\mathbb{R}^d} \varphi^\sigma(x) \tilde{\rho}(t, x) dx \\ &= \int_{\mathbb{R}^d} \varphi^\sigma(x) \rho_0(x) dx - 2 \int_0^t \int_{\mathbb{R}^d} [\tilde{\rho}(s, x)]^{\frac{m}{2}} \nabla\varphi^\sigma(x) \cdot \nabla[\tilde{\rho}(s, x)]^{\frac{m}{2}} dx ds \\ &= \int_{\mathbb{R}^d} \varphi^\sigma(x) \rho_0(x) dx - \frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} \tilde{\rho}(s, x) \nabla\varphi^\sigma(x) \cdot \nabla[\tilde{\rho}(s, x)]^{m-1} dx ds, \end{aligned}$$

where in the last equality we are using  $m \geq 2$ , hence the chain rule holds true, cf. Remark 2.3. More precisely, we re-write  $\rho^{m-1} = G \circ u$ , for  $u = \rho^{\frac{m}{2}}$  and  $G(x) = x^{\frac{2(m-1)}{m}}$  since  $\frac{2(m-1)}{m} \geq 1$ . As pointed out in Remark 2.3, the usual definition of weak solution holds by identifying  $\nabla \rho^m = 2\rho^{\frac{m}{2}} \nabla \rho^{\frac{m}{2}}$  (in the weak sense) — write  $\rho^m = G \circ u$ , for  $u = \rho^{\frac{m}{2}}$  and  $G(x) = x^2$ . Since  $\varphi^\sigma$  converges uniformly to  $\varphi$  on compact sets, we can let  $\sigma \rightarrow 0$  and obtain that  $\tilde{\rho}$  is a weak solution to (PME) in the sense of Definition 2.3. Uniqueness of weak solutions of (PME) is a known result, c.f. e.g. [28, 58]. Hence, we obtain convergence of the whole sequence  $\rho^\varepsilon$  narrowly converges to  $\tilde{\rho}$ , and  $\tilde{\rho}^{-\frac{1}{2}} |\nabla \tilde{\rho}^m| \in L^1([0, T]; L^2(\mathbb{R}^d))$ , by comparison with the theory in [1].  $\square$

**Remark 5.2** (Theorem 2.2 for general functionals) For a general integrand  $F$  satisfying (F1), (F2), and (F3), the RHS of (5.1) becomes

$$- \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \nabla F'(V_\varepsilon * \rho_r^\varepsilon)(x) (V_\varepsilon * \rho_r^\varepsilon)(x) dx dr - \int_0^t \int_{\mathbb{R}^d} z_r^\varepsilon(x) \nabla F'(V_\varepsilon * \rho_r^\varepsilon) dx dr,$$

being  $z^\varepsilon$  as in (5.2). Supposing  $F$  also satisfies (F<sub>m</sub>) for some  $m \geq 2$ , we have the estimate  $|F''(x)| \leq cx^{m-2}$ , hence  $F \in C^2([0, \infty))$  (origin included). Notice that the power laws for  $m \geq 2$  satisfy all of (F1), (F2), (F3), and (F<sub>m</sub>). The error term can be estimated as

$$\begin{aligned} & \left| \int_0^t \int_{\mathbb{R}^d} z_r^\varepsilon(x) \nabla F'(V_\varepsilon * \rho_r^\varepsilon) dx dr \right| \\ & \lesssim \int_0^t \int_{\mathbb{R}^d} |z_r^\varepsilon(x)| |v_r^\varepsilon(x)|^{\frac{m}{2}-1} |\nabla(v_r^\varepsilon(x))^{\frac{m}{2}}| dx dr \\ & \lesssim \|z^\varepsilon\|_{L^m([0,t] \times \mathbb{R}^d)} \|(v^\varepsilon)^{\frac{m}{2}-1}\|_{L^q([0,t] \times \mathbb{R}^d)} \|\nabla(v^\varepsilon)^{\frac{m}{2}}\|_{L^2([0,t] \times \mathbb{R}^d)}, \end{aligned}$$

for  $q = \frac{2m}{m-2}$ , so that  $\frac{1}{m} + \frac{1}{q} + \frac{1}{2} = 1$ ; thus it vanishes as  $\varepsilon \rightarrow 0$  similar to Lemma 5.1. As for the first term, note that it can be rewritten as

$$\begin{aligned} & \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) \nabla F'(v_r^\varepsilon)(x) v_r^\varepsilon(x) dx dr \\ & = \frac{2}{m} \int_0^t \int_{\mathbb{R}^d} F''(v_r^\varepsilon(x)) (v_r^\varepsilon(x))^{2-\frac{m}{2}} \nabla(v_r^\varepsilon)^{\frac{m}{2}} \nabla \varphi(x) dx dr, \end{aligned}$$

where  $F''(x)x^{2-\frac{m}{2}}$  is extended by zero when  $x = 0$  owing to (F<sub>m</sub>), and we applied the chain rule twice on the set  $\{v_r^\varepsilon > 0\}$

$$v_r^\varepsilon \nabla v_r^\varepsilon = \frac{1}{2} \nabla [(v_r^\varepsilon)^{\frac{m}{2}}]^{\frac{4}{m}} = \frac{2}{m} (v_r^\varepsilon(x))^{2-\frac{m}{2}} \nabla(v_r^\varepsilon)^{\frac{m}{2}}.$$

When multiplied with  $F''(v_r^\varepsilon)$ , the integrand on the right-hand side makes sense in  $L^1$  owing to (F<sub>m</sub>) since  $F''(x) \leq cx^{m-2}$ . Then, we are left to show  $g(v^\varepsilon) := F''(v^\varepsilon)(v^\varepsilon)^{2-\frac{m}{2}}$  strongly converges in  $L^2([0, T] \times \mathbb{R}^d)$ . This is indeed achieved by bounding  $|F''(v^\varepsilon)(v^\varepsilon)^{2-\frac{m}{2}}| \leq c(v^\varepsilon)^{\frac{m}{2}}$  and applying the generalised version of the Lebesgue dominated convergence theorem. Therefore, in the  $\varepsilon \rightarrow 0^+$  limit we obtain (up to pass to a subsequence)

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(x) \tilde{\rho}(t, x) dx & = \int_{\mathbb{R}^d} \varphi(x) \rho_0(x) dx - \frac{2}{m} \int_0^t \int_{\mathbb{R}^d} F''(\tilde{\rho}(s, x)) (\tilde{\rho}(s, x))^{2-\frac{m}{2}} \\ & \quad \nabla(\tilde{\rho}(s, x))^{\frac{m}{2}} \nabla \varphi(x) dx dr. \end{aligned}$$

Defining  $G$  such that  $G'(x^{\frac{m}{2}}) = F''(x)x^{2-\frac{m}{2}}$ , or  $G'(x) = F''(x^{\frac{2}{m}})x^{\frac{2}{m}(2-\frac{m}{2})}$ , and  $G(0) = 0$ , we can apply the chain rule to  $P(x) = G(x^{\frac{m}{2}})$  to obtain that  $\nabla P(\rho) \in L^1([0, T] \times \mathbb{R}^d)$ ,

thus

$$\int_{\mathbb{R}^d} \varphi(x) \tilde{\rho}(t, x) dx = \int_{\mathbb{R}^d} \varphi(x) \rho_0(x) dx - \int_0^t \int_{\mathbb{R}^d} \nabla P(\tilde{\rho}(s, x)) \nabla \varphi(x) dx ds.$$

Note that the chain rule and the construction of the pressure  $P$  holds for all  $m > 1$ . Uniqueness of distributional solutions of (DE) is proven in [5] for bounded solutions, which is actually the case for  $L^1$  solutions — see [15, 59] for further details on the so-called  $L^1$ - $L^\infty$  regularising effect. In particular, from [1, Theorem 11.2.5] we infer that our solution is a 2-Wasserstein gradient flow satisfying

$$\int_0^T \int_{\mathbb{R}^d} \frac{|\nabla P(\rho)|^2}{\rho} dx dt < \infty.$$

Furthermore, uniqueness of solutions implies convergence of the whole sequence  $\rho^\varepsilon$ , as for (PME).

**Remark 5.3** Our result can be also interpreted in the context of *generalised gradient flows* or *gradient structures*, following the dynamical interpretation of the Wasserstein distance, c.f. [3, 33]. More precisely, we know

$$d_W^2(\rho_0, \rho_1) = \inf \left\{ \int_0^1 \mathcal{A}(\rho_t, j_t) dt : (\rho, j) \text{ solves } \begin{cases} \partial_t \rho + \nabla \cdot j = 0, \\ \rho(0) = \rho_0, \quad \rho(1) = \rho_1 \end{cases} \right\},$$

where, for any  $\lambda \in \mathcal{M}(\mathbb{R}^d; \mathbb{R}^d)$  such that  $\rho, j \ll |\lambda|$ ,

$$\mathcal{A}(\rho, j) = \int_{\mathbb{R}^d} \alpha \left( \frac{dj}{d|\lambda|}, \frac{d\rho}{d|\lambda|} \right) d|\lambda|, \quad \text{and} \quad \alpha(j, r) := \begin{cases} \frac{(j)^2}{r} & \text{if } r > 0, \\ 0 & \text{if } j \leq 0 \text{ and } r = 0, \\ \infty & \text{if } j > 0 \text{ and } r = 0. \end{cases}$$

Upon using a careful regularisation and cut-off argument one can prove the following chain rule for any absolutely continuous curve with respect to the Wasserstein distance

$$\mathcal{F}^\varepsilon[\rho(t)] - \mathcal{F}^\varepsilon[\rho_0] = - \int_0^t \int_{\mathbb{R}^d} \nabla \frac{\delta \mathcal{F}^\varepsilon}{\delta \rho}(x) \cdot dj_t(x) dt,$$

hence re-interpret weak (measure) solutions of (NLE) as the zero level set of the De Giorgi functional

$$\mathcal{F}^\varepsilon[\rho(t)] - \mathcal{F}^\varepsilon[\rho_0] + \frac{1}{2} \int_0^t \int_{\mathbb{R}^d} \left| \nabla \frac{\delta \mathcal{F}^\varepsilon}{\delta \rho}(x) \right|^2 d\rho_t(x) dt + \frac{1}{2} \int_0^t \mathcal{A} \left( \rho, -\rho \nabla \frac{\delta \mathcal{F}^\varepsilon}{\delta \rho} \right) dt = 0.$$

### 5.2 The case $1 < m < 2$

The key idea here is to estimate the error  $z^\varepsilon$  from (5.2) differently by exploiting the compact support of  $V_1$ . We take  $R > 0$  such that  $\text{supp} V_1 \subset B_R$ . In the case  $m \geq 2$ , negative powers of  $v^\varepsilon = V_\varepsilon * \rho^\varepsilon$  never appeared in (5.1) but these computations can be recycled by cautiously avoiding the 0 level set of  $v^\varepsilon$ . We define

$$A_r^\varepsilon := \{x \in \mathbb{R}^d \mid v_r^\varepsilon(x) = V_\varepsilon * \rho_r^\varepsilon(x) > 0\}$$

and alter the computations in (5.1) carefully

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi d\rho_t^\varepsilon - \int_{\mathbb{R}^d} \varphi d\rho_0 &= -\frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \nabla V_\varepsilon * (V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) d\rho_r^\varepsilon(x) dr \\ &= \frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} (V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) [\nabla V_\varepsilon * (\rho_r^\varepsilon \nabla\varphi)](x) dx dr \quad (5.3) \\ &= \frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} (V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) \nabla[V_\varepsilon * (\rho_r^\varepsilon \nabla\varphi)](x) dx dr. \end{aligned}$$

In the second line, we swapped the convolution against  $\nabla V_\varepsilon$  and picked up a minus sign because it is an odd function (remember from (V) that  $V_1$  is even). At this point, we would like to perform integration by parts and apply the gradient onto  $(V_\varepsilon * \rho_r^\varepsilon)^{m-1}$ . In contrast to (5.1), when  $1 < m < 2$  we need to avoid the zero set of  $V_\varepsilon * \rho_r^\varepsilon$ ; as smooth as this convolution may be, the function  $(V_\varepsilon * \rho_r^\varepsilon)^{m-1}$  is not differentiable on  $\mathbb{R}^d \setminus A_r^\varepsilon$ . However, due to Lemma 5.3 (see below), we can justify the integration by parts and develop (5.3) to get

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi d\rho_t^\varepsilon - \int_{\mathbb{R}^d} \varphi d\rho_0 &= \frac{m}{m-1} \int_0^t \int_{\mathbb{R}^d} (V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) \nabla[V_\varepsilon * (\rho_r^\varepsilon \nabla\varphi)](x) dx dr \\ &= -\frac{m}{m-1} \int_0^t \int_{A_r^\varepsilon} \nabla(V_\varepsilon * \rho_r^\varepsilon)^{m-1}(x) V_\varepsilon * (\rho_r^\varepsilon \nabla\varphi)(x) dx dr \quad (\text{Lemma 5.3}) \\ &= -2 \int_0^t \int_{A_r^\varepsilon} (V_\varepsilon * \rho_r^\varepsilon \nabla\varphi)(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}-1} \nabla(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}} dx dr \\ &= -2 \int_0^t \int_{\mathbb{R}^d} \nabla\varphi(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) \nabla(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr \\ &\quad - 2 \int_0^t \int_{A_r^\varepsilon} z_r^\varepsilon(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}-1} \nabla(V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr. \end{aligned} \tag{5.4}$$

The last line is nearly identical to the end result of (5.1). Here, we integrate over  $A_r^\varepsilon = \{V_\varepsilon * \rho_r^\varepsilon > 0\}$  which is justified by Lemma 5.3. As was the case in Sect. 5.1, we need to show that the error term in the last line vanishes as  $\varepsilon \downarrow 0$ .

Proof of Theorem 2.2 for  $1 < m < 2$  Convergence in the first term on the right-hand side of (5.4) can be treated as for the case  $m \geq 2$ , due to Lemma 4.1 and Proposition 4.3. Hence we focus on the error term. For simplicity, let us assume  $\varphi \in C_c^2(\mathbb{R}^d)$  since it can be approximated in such a way as described in Sect. 5.1. We estimate the error by first expressing it as

$$\begin{aligned} z^\varepsilon(x) &= V_\varepsilon * (\rho^\varepsilon \nabla\varphi)(x) - (V_\varepsilon * \rho^\varepsilon)(x) \nabla\varphi(x) \\ &= \int_{\mathbb{R}^d} V_\varepsilon(x-y) (\nabla\varphi(y) - \nabla\varphi(x)) d\rho^\varepsilon(y). \end{aligned}$$

Next, we apply the Mean Value Theorem to the difference  $|\nabla\varphi(y) - \nabla\varphi(x)| \leq \|D^2\varphi\|_{L^\infty} |x - y|$  and obtain

$$|z^\varepsilon| \leq \|D^2\varphi\|_{L^\infty} \int_{\mathbb{R}^d} |x-y| V_\varepsilon(x-y) d\rho^\varepsilon(y).$$

Now, we exploit the compact support of the generator  $V_1$ . Since  $V_1$  is supported within  $B_R$ , then  $V_\varepsilon$  is supported within  $B_{R\varepsilon}$  which leads to

$$\begin{aligned} |z^\varepsilon(x)| &\leq \|D^2\varphi\|_{L^\infty} \int_{\{y \in \mathbb{R}^d \mid |x-y| \leq R\varepsilon\}} |x-y| V_\varepsilon(x-y) d\rho^\varepsilon(y) \\ &\leq R\varepsilon \|D^2\varphi\|_{L^\infty} \int_{\mathbb{R}^d} V_\varepsilon(x-y) d\rho^\varepsilon(y) = R\varepsilon \|D^2\varphi\|_{L^\infty} (V_\varepsilon * \rho^\varepsilon)(x), \quad \forall x \in \mathbb{R}^d. \end{aligned} \tag{5.5}$$

The last integral in (5.4) can be estimated with (5.5) as follows

$$\begin{aligned} &\left| \int_0^t \int_{A_r^\varepsilon} z_r^\varepsilon(x) (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}-1} \nabla (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) dx dr \right| \\ &\leq R\varepsilon \|D^2\varphi\|_{L^\infty} \int_0^t \int_{A_r^\varepsilon} (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}} \left| \nabla (V_\varepsilon * \rho_r^\varepsilon)^{\frac{m}{2}}(x) \right| dx dr \\ &\leq R\varepsilon \|D^2\varphi\|_{L^\infty} \| (v^\varepsilon)^{\frac{m}{2}} \|_{L^2([0,T] \times \mathbb{R}^d)} \| \nabla (v^\varepsilon)^{\frac{m}{2}} \|_{L^2([0,T] \times \mathbb{R}^d)}. \end{aligned}$$

We have suggestively recalled the notation  $v^\varepsilon = V_\varepsilon * \rho^\varepsilon$  precisely with Lemma 4.1 in mind;  $(v^\varepsilon)^{\frac{m}{2}}$  is uniformly bounded in  $L^2([0, T]; H^1(\mathbb{R}^d))$ . Hence, the last integral is uniformly bounded in  $\varepsilon$ . Moreover, the prefactor of vanishing  $\varepsilon$  implies that the last term of (5.4) converges to zero in the limit, thus recovering

$$\int_{\mathbb{R}^d} \varphi(x) d\rho_t(x) = \int_{\mathbb{R}^d} \varphi(x) d\rho_0(x) - 2 \int_0^t \int_{\mathbb{R}^d} \nabla \varphi(x) (\rho_r)^{\frac{m}{2}}(x) \nabla (\rho_r)^{\frac{m}{2}}(x) dx dr.$$

By means of the chain rule for Sobolev spaces, one can prove  $\nabla \rho^m = 2\rho^{\frac{m}{2}} \nabla \rho^{\frac{m}{2}}$  (in the weak sense). More precisely, we can see  $\rho^m = G \circ u$ , for  $u = \rho^{\frac{m}{2}}$  and  $G(x) = x^2$ . The work by Dahlberg and Kenig [28] establishes uniqueness of very weak solutions for (PME), for  $m > 1$ . As a byproduct, we also infer that our solution is a gradient flow in the sense of [1, Theorem 11.2.5], meaning

$$\int_0^T \int_{\mathbb{R}^d} \frac{|\nabla \rho^m|^2}{\rho} dx dt < \infty.$$

As for general energies induced by  $F$  satisfying (F1), (F2), (F3), and (F<sub>m</sub>), a combination of the same estimates from Remark 5.2 (disregarding the Hölder estimate) and this new technique for treating  $z^\varepsilon$  yield the full result. □

We now justify the integration by parts step going from (5.3) to (5.4). Recall the notation  $v_r^\varepsilon = V_\varepsilon * \rho_r^\varepsilon$ .

**Lemma 5.3** *For fixed  $\varepsilon$ ,  $t > 0$ , and  $\varphi \in C_c^1(\mathbb{R}^d)$ , there holds*

$$\int_0^t \int_{\mathbb{R}^d} (v_r^\varepsilon)^{m-1}(x) \nabla V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x) dx dr = - \int_0^t \int_{A_r^\varepsilon} \nabla (v_r^\varepsilon)^{m-1}(x) V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x) dx dr. \tag{5.6}$$

*In particular, both integrals converge absolutely.*



**Proof** We begin by proving both integrals in (5.6) converge absolutely. For the integral on the left-hand side of (5.6), we estimate

$$\begin{aligned}
 & \int_0^t \int_{\mathbb{R}^d} |(v_r^\varepsilon)^{m-1}(x) \nabla V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x)| dx dr \\
 & \leq \int_0^t \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} V_\varepsilon(x-y) d\rho(y) \right)^{m-1} \left( \int_{\mathbb{R}^d} |\nabla V_\varepsilon(x-z) \nabla \varphi(z)| d\rho_r^\varepsilon(z) \right) dx dr \\
 & \leq \int_0^t \int_{\mathbb{R}^d} \|V_\varepsilon\|_{L^\infty}^{m-1} \|\nabla \varphi\|_{L^\infty} \int_{\mathbb{R}^d} |\nabla V_\varepsilon(x-z)| d\rho_r^\varepsilon(z) dx dr \\
 & \leq \int_0^t \|V_\varepsilon\|_{L^\infty}^{m-1} \|\nabla \varphi\|_{L^\infty} \|\nabla V_\varepsilon\|_{L^1} dr \leq \|V_\varepsilon\|_{L^\infty}^{m-1} \|\nabla \varphi\|_{L^\infty} \|\nabla V_\varepsilon\|_{L^1} t,
 \end{aligned}
 \tag{5.7}$$

where the last line is obtained by Fubini’s theorem. Therefore, the integral on the left-hand side of (5.6) is absolutely convergent. Turning to the integral on the right-hand side of (5.6), we first record

$$\begin{aligned}
 |V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x)| & \leq \int |V_\varepsilon(x-y) \nabla \varphi(y)| d\rho_r^\varepsilon(y) \leq \|\nabla \varphi\|_{L^\infty} \int V_\varepsilon(x-y) d\rho(y) \\
 & = \|\nabla \varphi\|_{L^\infty} v_r^\varepsilon(x) = \|\nabla \varphi\|_{L^\infty} v_r^\varepsilon(x) \chi_{A_r^\varepsilon}(x),
 \end{aligned}
 \tag{5.8}$$

where  $\chi_{A_r^\varepsilon}(x)$  is the indicator function on the set  $A_r^\varepsilon$ . With (5.8), we obtain

$$\begin{aligned}
 |\nabla(v_r^\varepsilon)^{m-1}(x)| |V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x)| & \leq \|\nabla \varphi\|_{L^\infty} |\nabla(v_r^\varepsilon)^{m-1}(x)| v_r^\varepsilon(x) \chi_{A_r^\varepsilon}(x) \\
 & = (m-1) \|\nabla \varphi\|_{L^\infty} |\nabla v_r^\varepsilon| (v_r^\varepsilon)^{m-2} v_r^\varepsilon \chi_{A_r^\varepsilon} \\
 & = \frac{2(m-1)}{m} \|\nabla \varphi\|_{L^\infty} |\nabla(v_r^\varepsilon)|^{\frac{m}{2}} (v_r^\varepsilon)^{\frac{m}{2}} \chi_{A_r^\varepsilon}.
 \end{aligned}$$

Recalling Lemma 4.1, we conclude by comparison that

$$|\nabla(v_r^\varepsilon)^{m-1}(x)| |V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x)| \in L^1((0, t) \times \mathbb{R}^d),$$

which shows that the integral on the right-hand side of (5.6) is absolutely convergent.

Integration by parts: For brevity, we drop the subscript  $r \in (0, t)$  and the superscript  $\varepsilon > 0$  so we consider  $v = V * \rho$  in place of  $v_r^\varepsilon = V_\varepsilon * \rho_r^\varepsilon$ . In order to verify (5.6), we fix  $\sigma > 0$  and a direction  $e \in \mathbb{S}^{d-1}$  and look at the following difference quotient

$$I_\sigma := \int_0^t \int_{\mathbb{R}^d} v^{m-1}(x) \frac{V * (\rho \nabla \varphi)(x + \sigma e) - V * (\rho \nabla \varphi)(x)}{\sigma} dx dr.$$

Owing to the uniform bound from (5.8) and Lebesgue’s Dominated Convergence Theorem, we have

$$\lim_{\sigma \rightarrow 0} I_\sigma = e \cdot \int_0^t \int_{\mathbb{R}^d} (v_r^\varepsilon)^{m-1}(x) \nabla V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x) dx dr,$$

recovering the left-hand side of (5.6) (along any arbitrary direction  $e \in \mathbb{S}^{d-1}$ ). On the other hand, by changing variables we also have

$$\begin{aligned}
 I_\sigma & = \int_0^t \int_{\mathbb{R}^d} \frac{v^{m-1}(x - \sigma e) - v^{m-1}(x)}{\sigma} V * (\rho \nabla \varphi)(x) dx dr \\
 & = - \int_0^t \int_{A_r^\varepsilon} \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) dx dr.
 \end{aligned}$$

We are allowed to restrict the integration region to  $A_r^\varepsilon$  due to (5.8); if  $x \notin A_r^\varepsilon$ , then  $|V * (\rho \nabla \varphi)(x)| \leq \|\nabla \varphi\|_{L^\infty} v(x) = 0$ . In order to prove (5.6), we wish to show

$$\begin{aligned} \lim_{\sigma \rightarrow 0} I_\sigma &= \lim_{h \rightarrow 0} \left( - \int_0^t \int_{A_r^\varepsilon} \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) dx dr \right) \\ &= -e \cdot \int_0^t \int_{A_r^\varepsilon} \nabla (v_r^\varepsilon)^{m-1}(x) V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x) dx dr. \end{aligned} \tag{5.9}$$

The strategy is to apply the extended Dominated Convergence Theorem (Theorem A.1) by exhibiting an appropriate sequence of majorants to the integrand in the first line of (5.9).

The first step is to remember (5.8) and estimate the integrand of  $I_\sigma$  as follows

$$\begin{aligned} &\left| \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) \right| \\ &\leq \|\nabla \varphi\|_{L^\infty} \left| \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} \right| |V * \rho(x)| \\ &= \|\nabla \varphi\|_{L^\infty} \left| \frac{v^m(x) - v^{m-1}(x - \sigma e)v(x)}{\sigma} \right|. \end{aligned} \tag{5.10}$$

In the last line, we have distributed  $v = V * \rho$  into the difference quotient. We wish to re-express the term  $v^{m-1}(x - \sigma e)v(x)$  as  $v^m(x)$ + error term. For this, we use the Mean Value Theorem to write

$$v(x) = v(x - \sigma e) - \int_0^1 \frac{d}{ds} v(x - s\sigma e) ds = v(x - \sigma e) + \sigma e \cdot \int_0^1 \nabla v(x - s\sigma e) ds.$$

Substituting this into  $v^m(x) - v^{m-1}(x - \sigma e)v(x)$  gives

$$\begin{aligned} &v^m(x) - v^{m-1}(x - \sigma e)v(x) \\ &= v^m(x) - v^{m-1}(x - \sigma e) \left( v(x - \sigma e) + \sigma e \cdot \int_0^1 \nabla v(x - s\sigma e) ds \right). \end{aligned}$$

Inserting this into (5.10) yields the estimate

$$\begin{aligned} &\left| \frac{v^{m-1}(x) - v^{m-1}(x - h\sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) \right| \\ &\leq \|\nabla \varphi\|_{L^\infty} \frac{1}{\sigma} \left| v^m(x) - v^m(x - \sigma e) - v^{m-1}(x - \sigma e)\sigma e \right. \\ &\quad \left. \cdot \int_0^1 \nabla V * (\rho \nabla \varphi)(x - s\sigma e) ds \right|. \end{aligned} \tag{5.11}$$

We use the Mean-Value Theorem again with the initial difference (valid since  $m > 1$ )

$$\begin{aligned} &v^m(x) - v^m(x - \sigma e) \\ &= - \int_0^1 \frac{d}{ds} v^m(x - s\sigma e) ds = m\sigma e \\ &\quad \cdot \int_0^1 v^{m-1}(x - s\sigma e) \nabla v(x - s\sigma e) ds \end{aligned}$$

and insert this into (5.11) to obtain

$$\begin{aligned}
 & \left| \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) \right| \\
 & \leq \|\nabla \varphi\|_{L^\infty} \frac{1}{\sigma} \left| m \sigma e \cdot \int_0^1 v^{m-1}(x - s \sigma e) \nabla v(x - s \sigma e) ds - v^{m-1}(x - \sigma e) \sigma e \right. \\
 & \quad \left. \cdot \int_0^1 \nabla V * (\rho \nabla \varphi)(x - s \sigma e) ds \right| \\
 & \leq \|\nabla \varphi\|_{L^\infty} \left| m \int_0^1 v^{m-1}(x - s \sigma e) \nabla v(x - s \sigma e) - v^{m-1}(x - \sigma e) \nabla V * (\rho \nabla \varphi)(x - s \sigma e) ds \right| \\
 & \leq \|\nabla \varphi\|_{L^\infty} \int_0^1 \left| m v^{m-1}(x - s \sigma e) + \|\nabla \varphi\|_{L^\infty} v^{m-1}(x - \sigma e) \right| |\nabla v(x - s \sigma e)| ds. \\
 & \leq C_{m,\varphi} \|V\|_{L^\infty}^{m-1} \int_0^1 |\nabla v(x - s \sigma e)| ds.
 \end{aligned}
 \tag{5.12}$$

In the third line of (5.12), we eliminated the common factor of  $\sigma$  together with the trivial estimate  $|e| = 1$ . In the fourth line of (5.12), we estimated similar to (5.8) the convolution

$$|\nabla V * (\rho \nabla \varphi)| \leq \|\nabla \varphi\|_{L^\infty} |\nabla V * \rho| = \|\nabla \varphi\|_{L^\infty} |\nabla v|.$$

In the final line of (5.12), we used the following inequality

$$v^{m-1}(x) = \left( \int V(x - y) d\rho(y) \right)^{m-1} \leq \|V\|_{L^\infty}^{m-1} \left( \int d\rho(y) \right)^{m-1} = \|V\|_{L^\infty}^{m-1}.$$

We are now in a position to apply Theorem A.1 with  $X = (0, t) \times \mathbb{R}^d$  where  $y = (r, x) \in X$  and

$$\begin{aligned}
 f^\sigma(r, x) &= \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) \chi_{A_r^\varepsilon}(x), \\
 g^\sigma(r, x) &= C_{m,\varphi} \|V\|_{L^\infty}^{m-1} \int_0^1 |\nabla v(x - s \sigma e)| ds.
 \end{aligned}$$

Here,  $\chi_{A_r^\varepsilon}(x)$  is the indicator function of the set  $A_r^\varepsilon$ . Remember that we have suppressed the dependence on  $r \in (0, t)$  in  $\rho$ . The first assumption of Theorem A.1 has been verified by the estimate of (5.12). We can verify the second assumption of Theorem A.1 since the pointwise limits of  $f^\sigma$  and  $g^\sigma$  as  $\sigma \rightarrow 0$  are

$$f(r, x) = e \cdot \nabla(v)^{m-1}(x) V * (\rho \nabla \varphi)(x) \chi_{A_r^\varepsilon}(x), \quad g(r, x) = C_{m,\varphi} \|V\|_{L^\infty}^{m-1} |\nabla v(x)|.$$

The pointwise limit for  $f^\sigma$  is justified since  $A_r^\varepsilon$  is an open set. As for the pointwise limit of  $g^\sigma$ , the usual Dominated Convergence Theorem suffices.

It remains to check the third assumption of Theorem A.1 which we do with Fubini;

$$\begin{aligned}
 \int_X g^\sigma(r, x)dy &= C_{m,\varphi} \|V\|_{L^\infty}^{m-1} \int_0^t \int_{\mathbb{R}^d} \int_0^1 |\nabla v(x - s\sigma e)| ds dx dr \\
 &= C_{m,\varphi} \|V\|_{L^\infty}^{m-1} \int_0^t \int_0^1 \int_{\mathbb{R}^d} |\nabla v(x - s\sigma e)| dx ds dr \\
 &= C_{m,\varphi} \|V\|_{L^\infty}^{m-1} \int_0^t \int_0^1 \int_{\mathbb{R}^d} |\nabla v(x)| dx ds dr \\
 &= C_{m,\varphi} \|V\|_{L^\infty}^{m-1} \int_0^t \int_{\mathbb{R}^d} |\nabla v(x)| dx dr = \int_X g(r, x)dy.
 \end{aligned}$$

Therefore, by Theorem A.1, we have  $\int_X f^\sigma(r, x)dy \rightarrow \int_X f(r, x)dy$  as  $\sigma \rightarrow 0$  which is precisely (5.9);

$$\begin{aligned}
 \lim_{\sigma \rightarrow 0} I_\sigma &= \lim_{\sigma \rightarrow 0} \left( - \int_0^t \int_{A_\varepsilon^\varepsilon} \frac{v^{m-1}(x) - v^{m-1}(x - \sigma e)}{\sigma} V * (\rho \nabla \varphi)(x) dx dr \right) \\
 &= - \lim_{\sigma \rightarrow 0} \int_X f^\sigma(r, x)dy = - \lim_{\sigma \rightarrow 0} \int_X f(r, x)dy \\
 &= -e \cdot \int_0^t \int_{A_\varepsilon^\varepsilon} \nabla (v_r^\varepsilon)^{m-1}(x) V_\varepsilon * (\rho_r^\varepsilon \nabla \varphi)(x) dx dr.
 \end{aligned}$$

□

### 6 Convexity, uniqueness, and particle approximation

In this section, we sketch the argument adapted from [27] to prove Theorem 2.3 and Corollary 2.1. Recall that we assume  $(F_m)$  with  $m > 1$ . To simplify the exposition, we set  $F'(0) = 0$ . In view of the assumptions needed for the kernel  $V_1$ , see  $(V)$ , it is not reasonable to choose  $V_1$  convex, as we require finite second order moment. However, this does not prohibit  $\lambda$ -convexity of the functional  $\mathcal{F}^\varepsilon$  along geodesics. Indeed, differentiability of  $\mathcal{F}^\varepsilon$ , as in [10, Proposition 3.10], holds in our case assuming  $F$  satisfies  $(F1)$ ,  $(F2)$ ,  $(F3)$ , and it is convex (convexity is ensured by  $(F_m)$ ). Furthermore, we need  $V_1 \in C^2(\mathbb{R}^d)$  and  $D^2V_1 \in L^\infty(\mathbb{R}^d)$ . In order to give a few explanations in this direction, fix  $\rho_1, \rho_2 \in \mathcal{P}_2(\mathbb{R}^d)$  and consider the geodesic connecting  $\rho_1$  to  $\rho_2$  defined by

$$\rho_\alpha := ((1 - \alpha)\pi^1 + \alpha\pi^2)_\# \gamma, \quad \alpha \in [0, 1],$$

where  $\gamma \in \Gamma(\mathbb{R}^d \times \mathbb{R}^d)$  satisfies

$$\pi_i^\# \gamma = \rho_i \text{ for } i = 1, 2.$$

Owing to the regularisation by  $V_\varepsilon$ , one can show (generalising and adapting the computations in [27, Propositions 3.4 and 3.6]) that the functional  $\mathcal{F}^\varepsilon$  satisfies a geodesic ‘above the tangent line’ inequality [25, Proposition 2.8] in the sense that

$$\mathcal{F}^\varepsilon(\rho_2) - \mathcal{F}^\varepsilon(\rho_1) - \frac{d}{d\alpha} \Big|_{\alpha=0} \mathcal{F}^\varepsilon(\rho_\alpha) \geq -\frac{\lambda_F^\varepsilon}{2} d_W^2(\rho_1, \rho_2), \tag{6.1}$$

where

$$\lambda_F^\varepsilon := -\frac{c_2 \|D^2 V_\varepsilon\|_{L^\infty} \|V_\varepsilon\|_{L^\infty}^{m-2}}{m-1}.$$

The technical assumption  $F'(0) = 0$  enters here in the computation of (6.1). As a consequence we infer  $\lambda_F^\varepsilon$ -convexity similar to [10, Proposition 3.11] as well as a characterisation of the subdifferential, [10, Proposition 3.12] — adapted to our functional, meaning that

$$\frac{\delta \mathcal{F}^\varepsilon}{\delta \rho} = V_\varepsilon * F'(V_\varepsilon * \rho).$$

As for the modulus of convexity  $\lambda_F^\varepsilon$ , similarly to [27], using  $(\mathbf{F}_m)$  and fixed  $m > 1$ , we have

$$\lambda_F^\varepsilon \approx -\varepsilon^{-2-d(m-1)},$$

meaning that  $\lambda_F^\varepsilon \rightarrow -\infty$  as  $\varepsilon \rightarrow 0$ . The information above is enough to prove existence of a unique gradient flow of  $\mathcal{F}^\varepsilon$ , for  $\varepsilon > 0$  fixed, following [1] and [10, Section 5] for regularised energies, thus proving Theorem 2.3. We omit the details and refer the interested reader again to similar computations in [27].

The regularisation of the energy by mollifiers  $V_1$  satisfying  $(\mathbf{V})$  used in this manuscript, as well as in [7], allows to extend stability of gradient flows to the case  $m > 1$  — further assuming  $V_1$  is compactly supported for  $1 < m < 2$  (c.f. Sect. 5.2). The advantage of using convex energies is given by the possibility of using stability estimates with the 2-Wasserstein distance so that one obtains a particle approximation when the number of particles involved depends on  $\varepsilon$ , i.e.  $N = N(\varepsilon)$ . This is a qualitative result, recently proved rigorously in [27, Theorem 1.4], for  $m = 2$ . In our setting, we consider (NLE) as a continuity equation with velocity given by  $-\nabla V_\varepsilon * F'(V_\varepsilon * \rho)$ . Then, under mild assumptions on the mollifier  $V_1$  and the function  $F$ , the empirical measure  $\rho_\varepsilon^N(t) = \frac{1}{N} \sum_{j=1}^N \delta_{x_\varepsilon^j(t)}$  is a weak solution to (NLE) provided the particles satisfy the following ODE system

$$\dot{x}_\varepsilon^i(t) = -\nabla \int_{\mathbb{R}^d} V_\varepsilon(x_\varepsilon^i(t) - y) F' \left( \frac{1}{N} \sum_{j=1}^N V_\varepsilon(y - x_\varepsilon^j(t)) \right) dy \quad \forall i = 1, \dots, N.$$

The regularisation in (NLE) is done in the same spirit as [27], therefore an analogous version of [27, Theorem 1.4] also holds true in our setting, with  $\lambda_F^\varepsilon$  specified above and  $m > 1$ . More precisely, as a consequence of the usual stability estimate for  $\lambda$ -gradient flows, [1, Theorem 11.2.1], we know

$$d_W(\rho^\varepsilon(t), \rho_\varepsilon^N(t)) \leq e^{-\lambda_F^\varepsilon t} d_W(\rho_\varepsilon^N(0), \rho(0)).$$

Therefore, if we assume that for  $\varepsilon \rightarrow 0$  there exists  $N = N(\varepsilon) \rightarrow +\infty$  such that

$$e^{-\lambda_F^\varepsilon t} d_W(\rho_\varepsilon^N(0), \rho(0)) \rightarrow 0,$$

we infer the mean field limit since  $\rho^\varepsilon$  is converging to a weak solution of (DE).

**Acknowledgements** The authors would like to thank Martin Burger, David Gómez-Castro, and Giuseppe Savaré for helpful discussions on the topics of the manuscript. They are also grateful to the anonymous reviewers for their suggestions. The authors were supported by the Advanced Grant Nonlocal-CPD (Nonlocal PDEs for Complex Particle Dynamics: Phase Transitions, Patterns and Synchronization) of the European Research Council Executive Agency (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 883363). JAC was also partially supported by the EPSRC grant numbers EP/T022132/1 and EP/V051121/1. JW was supported by the Mathematical Institute Award at the University of Oxford.

**Data availability statement** Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## 7 Technical proofs

We begin with an extension of the Dominated Convergence Theorem [51, Chapter 4, Theorem 17]. This is quoted with less generality for our purposes.

**Theorem A.1** (*Extended Dominated Convergence Theorem*) *Let  $(f^\sigma)_{\sigma>0}$  and  $(g^\sigma)_{\sigma>0}$  be sequences of measurable functions on a (Lebesgue) measurable set  $X \subset \mathbb{R}^n$  such that  $g^\sigma \geq 0$  and suppose that there exist measurable functions  $f, g$  satisfying the following assumptions.*

- (1)  $|f^\sigma(y)| \leq g^\sigma(y)$  for all  $\sigma > 0$  and pointwise almost every  $y \in X$ .
- (2)  $f^\sigma(y) \rightarrow f(y)$  and  $g^\sigma(y) \rightarrow g(y)$  pointwise almost every  $y \in X$  as  $\sigma \rightarrow 0$ .
- (3)  $\int_X g^\sigma(y)dy \rightarrow \int_X g(y)dy$  as  $\sigma \rightarrow 0$ .

Then, we have  $\int_X f^\sigma(y)dy \rightarrow \int_X f(y)dy$  as  $\sigma \rightarrow 0$ .

Next, we turn to estimates related to the JKO scheme with the regularised energy  $\mathcal{F}^\varepsilon$  and regularity of  $V$  and  $F$  under various assumptions.

**Remark A.1** From the definition of the 2-Wasserstein distance and the inequality  $|y|^2 \leq 2|x|^2 + 2|x - y|^2$  it follows

$$m_2(\rho_1) \leq 2m_2(\rho_0) + 2d_W^2(\rho_0, \rho_1), \quad \forall \rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d).$$

For a function  $F : [0, +\infty) \rightarrow (-\infty, +\infty]$  satisfying **(F1)**, its negative part can be estimated (c.f. [1, Remark 9.3.7]) by

$$F^-(s) \leq c_1s + c_2s^\alpha, \quad \forall s \in [0, +\infty),$$

for some constants  $c_1, c_2 > 0$  and we can take, without loss of generality,  $\alpha \in \left(\frac{d}{d+2}, 1\right)$ . With this notation, we have the following result.

**Lemma A.1** *Suppose  $F : [0, +\infty) \rightarrow (-\infty, +\infty]$  satisfies **(F1)** for some  $\alpha \in \left(\frac{d}{d+2}, 1\right)$ . Then, whenever  $\rho \in \mathcal{P}_2(\mathbb{R}^d)$  has a density with respect to Lebesgue,  $F^-(\rho) \in L^1(\mathbb{R}^d)$ . In particular, the functional  $\mathcal{F}$  admits the lower bound*

$$\mathcal{F}[\rho] \geq -c_1 - c_2 C_{d,\alpha} (1 + m_2(\rho))^\alpha > -\infty,$$

where  $C_{d,\alpha} > 0$  is a constant depending only on  $d$  and  $\alpha$ . Moreover, for any  $\varepsilon > 0$  and  $V_1$  with finite second moment as in **(V)**, the functional  $\mathcal{F}^\varepsilon$  admits the lower bound

$$\mathcal{F}^\varepsilon[\rho] \geq -c_1 - c_2 \tilde{C}_{d,\alpha} (1 + \varepsilon^2 m_2(V_1) + m_2(\rho))^\alpha, \tag{A.1}$$

where  $\tilde{C}_{d,\alpha} = 4C_{d,\alpha}$  but we suppress multiplicative constants and abuse notation by reusing  $C_{d,\alpha}$ .

**Proof** The proof follows [1, Remark 9.3.7]. We have  $\int_{\mathbb{R}^d} F^-(\rho(x))dx \leq c_1 + c_2 \int_{\mathbb{R}^d} \rho^\alpha(x)dx$ . Using Hölder’s inequality and  $\frac{2\alpha}{1-\alpha} > d$ , we have

$$\begin{aligned} \int_{\mathbb{R}^d} \rho^\alpha(x)dx &= \int_{\mathbb{R}^d} \rho^\alpha(x)(1 + |x|)^{2\alpha}(1 + |x|)^{-2\alpha}dx \\ &\leq \left( \int_{\mathbb{R}^d} (1 + |x|)^2 \rho(x)dx \right)^\alpha \left( \int_{\mathbb{R}^d} (1 + |x|)^{\frac{-2\alpha}{1-\alpha}} dx \right)^{1-\alpha} \\ &\leq C_{d,\alpha}(2 + 2m_2(\rho))^\alpha < +\infty. \end{aligned}$$

The lower bound for  $\mathcal{F}$  readily follows. It remains to establish the lower bound for  $\mathcal{F}^\varepsilon$ . Recall that since  $\mathcal{F}^\varepsilon[\rho] = \mathcal{F}[V_\varepsilon * \rho]$ , we have

$$\mathcal{F}^\varepsilon[\rho] \geq -c_1 - c_2 C_{d,\alpha} (1 + m_2(V_\varepsilon * \rho))^\alpha.$$

It suffices to estimate  $m_2(V_\varepsilon * \rho)$ . Using Fubini, we have

$$\begin{aligned} m_2(V_\varepsilon * \rho) &= \int_{\mathbb{R}^d} |x|^2 \int_{\mathbb{R}^d} V_\varepsilon(x - y) d\rho(y) dx \\ &\leq 2 \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} |x - y|^2 V_\varepsilon(x - y) dx \right) d\rho(y) \\ &\quad + 2 \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} V_\varepsilon(x - y) dx \right) |y|^2 d\rho(y) \\ &= 2 \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} |z|^2 \varepsilon^{-d} V_1(z/\varepsilon) dz \right) d\rho(y) + 2m_2(\rho) = 2\varepsilon^2 m_2(V_1) + 2m_2(\rho). \end{aligned}$$

□

The lower bound for  $\mathcal{F}^\varepsilon$  is used to prove that one step of the JKO scheme (3.1) is well-defined.

**Lemma A.2** Fix  $\varepsilon > 0$  and suppose  $F : [0, +\infty) \rightarrow (-\infty, +\infty]$  satisfies (F1). For fixed  $\bar{\rho} \in \mathcal{P}_2(\mathbb{R}^d)$ , the functional  $\rho \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \frac{d_W^2(\rho, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho] \in (-\infty, +\infty]$  admits minimisers whenever  $\tau > 0$  is chosen independently of  $\varepsilon > 0$  to satisfy

$$\tau \leq \frac{1}{2c_2 C_{d,\alpha}}.$$

**Proof** This proof follows the direct method.

Step 1 - Uniform lower bound: We apply (A.1) from Lemma A.1 to begin with

$$\frac{d_W^2(\rho, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho] \geq \frac{d_W^2(\rho, \bar{\rho})}{2\tau} - c_1 - c_2 C_{d,\alpha} (1 + \varepsilon^2 m_2(V_1) + m_2(\rho))^\alpha.$$

Using Remark A.1, we can replace  $m_2(\rho)$  to further estimate

$$\frac{d_W^2(\rho, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho] \geq \frac{d_W^2(\rho, \bar{\rho})}{2\tau} - c_1 - c_2 C_{d,\alpha} (1 + \varepsilon^2 m_2(V_1) + m_2(\bar{\rho}) + d_W^2(\rho, \bar{\rho})),$$

where we have used  $(1 + |x|)^\alpha \leq 1 + |x|$  for  $\alpha \in (0, 1)$ . Finally, the only dependence on  $\rho$  is through  $d_W^2(\rho, \bar{\rho})$ . The coefficient of this term is positive when

$$\frac{1}{2\tau} - c_2 C_{d,\alpha} \geq 0 \iff 0 < \tau \leq \frac{1}{2c_2 C_{d,\alpha}}.$$

With this restriction on  $\tau$ , the lower bound reads

$$\frac{d_W^2(\rho, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho] \geq -c_1 - c_2 C_{d,\alpha}(1 + \varepsilon^2 m_2(V_1) + m_2(\bar{\rho})).$$

**Step 2 - The Direct Method:** From the previous step, we can find an infimising sequence  $\rho_n \in \mathcal{P}_2(\mathbb{R}^d)$  such that

$$\frac{d_W^2(\rho_n, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho_n] \rightarrow \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \frac{d_W^2(\mu, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\mu] \right\} > -\infty.$$

As a convergent sequence, we also have

$$\sup_{n \in \mathbb{N}} \left\{ \frac{d_W^2(\rho_n, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho_n] \right\} \leq C < +\infty,$$

for some constant  $C > 0$ . By Remark A.1, the second moments are uniformly bounded by

$$\sup_{n \in \mathbb{N}} m_2(\rho_n) \leq 2m_2(\bar{\rho}) + C\tau.$$

Therefore, by Prokhorov’s theorem, there exists a subsequence, still labelled  $\rho_n$ , such that  $\rho_n \rightharpoonup \rho$  (narrow convergence) for  $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ . Moreover, it is known (c.f. Remark 2.1) that (F1) provides the required lower semicontinuity with respect to narrow convergence so that

$$\frac{d_W^2(\rho, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho] \leq \liminf_{n \rightarrow \infty} \left\{ \frac{d_W^2(\rho_n, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\rho_n] \right\} = \inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \frac{d_W^2(\mu, \bar{\rho})}{2\tau} + \mathcal{F}^\varepsilon[\mu] \right\}.$$

□

These results are enough to prove Proposition 3.1.

**Lemma A.3** *Suppose  $F : [0, +\infty) \rightarrow \mathbb{R}$  satisfies (F2) and for  $\varepsilon > 0$ , define  $V_\varepsilon(x) = \varepsilon^{-d} V_1(x/\varepsilon)$  where  $V_1$  satisfies (V). Then, for any non-negative measure  $\rho \in \mathcal{M}(\mathbb{R}^d)$  such that  $\rho(\mathbb{R}^d) = C < +\infty$ , we have the estimate*

$$\|F'(V_\varepsilon * \rho(\cdot))\|_{L^\infty(\mathbb{R}^d)} \leq \|F'(\cdot)\|_{L^\infty([0, C\|V_\varepsilon\|_{L^\infty})} < +\infty.$$

**Proof** We estimate the convolution of  $V_\varepsilon$  against  $\rho$  by

$$V_\varepsilon * \rho(x) = \int_{\mathbb{R}^d} V_\varepsilon(x - y) d\rho(y) \leq C\|V_\varepsilon\|_{L^\infty}.$$

Being  $V_1 \geq 0$  and  $\rho$  a positive measure, the image of  $V_\varepsilon * \rho(\cdot)$  is therefore contained in  $[0, C\|V_\varepsilon\|_{L^\infty}]$ . Being  $F'$  continuous on  $[0, +\infty)$ , it is bounded on  $[0, C\|V_\varepsilon\|_{L^\infty}]$ , and the result is proved. □

**Lemma A.4** *For any  $\varepsilon > 0$ , define  $V_\varepsilon(x) = \varepsilon^{-d} V_1(x/\varepsilon)$  where  $V_1$  satisfies (V). Then, for any compact set  $K \subset \mathbb{R}^d$  and  $\rho \in \mathcal{P}(\mathbb{R}^d)$ , there is a constant  $C > 0$  depending only on  $\varepsilon, K, V_1$ , and  $m_1(\rho)$  such that*

$$|V_\varepsilon * \rho(x^1) - V_\varepsilon * \rho(x^2)| \leq C|x^1 - x^2|, \quad \forall x^1, x^2 \in K.$$



Moreover, suppose  $\rho^n \in \mathcal{P}(\mathbb{R}^d)$  is a sequence such that  $\sup_{n \in \mathbb{N}} m_1(\rho^n) < +\infty$ . Then, there is a subsequence of  $\rho^n$  (still labelled  $\rho^n$ ) such that  $\rho^n \rightarrow \rho \in \mathcal{P}(\mathbb{R}^d)$  and the following limit holds

$$\lim_{n \rightarrow \infty} \sup_{x \in K} |V_\varepsilon * \rho^n(x) - V_\varepsilon * \rho(x)| = 0.$$

**Proof** We use the mean-value form of Taylor’s theorem to estimate

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} (V_\varepsilon(x^1 - y) - V_\varepsilon(x^2 - y)) d\rho(y) \right| \\ & \leq |x^1 - x^2| \int_{\mathbb{R}^d} \int_0^1 |\nabla V_\varepsilon(tx^1 + (1-t)x^2 - y)| dt d\rho(y) \\ & \leq C_{\varepsilon, v_1} |x^1 - x^2| \int_{\mathbb{R}^d} (1 + |x^1| + |x^2| + |y|) d\rho(y) \leq C_{\varepsilon, v_1} C_K |x^1 - x^2| (1 + m_1(\rho)). \end{aligned}$$

In the last line, the constant  $C_{\varepsilon, v_1}$  comes from the linear growth assumption of  $\nabla V_1$  scaled with  $\varepsilon$ , and the constant  $C_K$  absorbs bounds for  $|x^1|$  and  $|x^2|$  since  $x^1, x^2 \in K$ .

Turning to the sequence  $\rho^n$ , standard tightness arguments using the uniform first moment bound yield  $\rho \in \mathcal{P}(\mathbb{R}^d)$  such that  $\rho^n \rightarrow \rho$  along a subsequence. As for the uniform convergence over  $K$  along a subsequence, properties of the mollification by  $V_\varepsilon$  yield that

$$x \mapsto |V_\varepsilon * \rho^n(x) - V_\varepsilon * \rho(x)|$$

is a continuous function for every  $n \in \mathbb{N}$ . Therefore, for every  $n \in \mathbb{N}$ , there is  $x_K^n \in K$  such that

$$\sup_{x \in K} |V_\varepsilon * \rho^n(x) - V_\varepsilon * \rho(x)| = |V_\varepsilon * \rho^n(x_K^n) - V_\varepsilon * \rho(x_K^n)|.$$

Since  $K$  is compact, there is a subsequence, still labelled  $x_K^n$ , such that  $x_K^n \rightarrow x_K \in K$  as  $n \rightarrow \infty$ . Along this subsequence, we continue the previous estimate

$$\begin{aligned} \sup_{x \in K} |V_\varepsilon * \rho^n(x) - V_\varepsilon * \rho(x)| &= |V_\varepsilon * \rho^n(x_K^n) - V_\varepsilon * \rho(x_K^n)| \\ &\leq |V_\varepsilon * \rho^n(x_K^n) - V_\varepsilon * \rho^n(x_K)| + |V_\varepsilon * \rho^n(x_K) - V_\varepsilon * \rho(x_K)| \\ &\quad + |V_\varepsilon * \rho(x_K) - V_\varepsilon * \rho(x_K^n)|. \end{aligned}$$

The first and third differences can be made arbitrarily small by the previous Lipschitz estimate owing to  $\sup_{n \in \mathbb{N}} m_1(\rho^n) < +\infty$ . For the second difference, remember that  $V_\varepsilon(\cdot) \in C_b(\mathbb{R}^d)$  so certainly  $V_\varepsilon(x_K - \cdot) \in C_b(\mathbb{R}^d)$  and  $\rho^n \rightarrow \rho$  in duality against  $C_b$ .  $\square$

**Corollary A.1** *In the setting of Lemma A.4, suppose  $F$  satisfies (F2). Then, the composition  $F'(V_\varepsilon * \rho(\cdot))$  is uniformly continuous on  $K$ . Moreover,*

$$\lim_{n \rightarrow \infty} \sup_{x \in K} |F'(V_\varepsilon * \rho^n(x)) - F'(V_\varepsilon * \rho(x))| = 0.$$

**Proof** Fix  $\eta > 0$ , then since  $F' \in C([0, +\infty))$ , it is uniformly continuous on  $[0, \|V_\varepsilon\|_{L^\infty}]$  so there exists  $\delta_1 > 0$  such that

$$\forall V^1, V^2 \in [0, \|V_\varepsilon\|_{L^\infty}], \quad |V^1 - V^2| < \delta_1 \implies |F'(V^1) - F'(V^2)| < \eta. \quad (\text{A.2})$$

Starting with the uniform continuity of  $F'(V_\varepsilon * \rho(\cdot))$ , for the  $\delta_1$  in (A.2), define  $\delta := \delta_1/C$  where  $C > 0$  is the Lipschitz constant in Lemma A.4. Then, whenever  $x^1, x^2 \in K$  satisfy

$|x^1 - x^2| < \delta$ , the Lipschitz estimate from Lemma A.4 implies

$$|V_\varepsilon * \rho(x^1) - V_\varepsilon * \rho(x^2)| < \delta_1.$$

As well, we can take  $V^i = V_\varepsilon * \rho(x^i)$  for  $i = 1, 2$  in (A.2) to conclude the uniform continuity. Note that  $V^i \in [0, \|V_\varepsilon\|_{L^\infty}]$ .

Turning to the limit, the uniform convergence from Lemma A.4 gives the existence of  $N \in \mathbb{N}$  such that for every  $n \geq N$  and  $x \in K$ , we have

$$|V_\varepsilon * \rho^n(x) - V_\varepsilon * \rho(x)| < \delta_1.$$

We would like to apply (A.2) and set  $V^1 = V_\varepsilon * \rho^n(x)$  and  $V^2 = V_\varepsilon * \rho(x)$ . We need to verify that  $V_\varepsilon * \rho^n(x), V_\varepsilon * \rho(x) \in [0, \|V_\varepsilon\|_{L^\infty}]$  for every  $x \in K$  which is easily proven.  $\square$

## References

1. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd edn. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2008)
2. Barenblatt, G.: On one class of solutions of the one-dimensional problem of non-stationary filtration of a gas in a porous medium. Prikl. Mat. i Mekh. **17**, 739–742 (1953)
3. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000)
4. Boi, S., Capasso, V., Morale, D.: Modeling the aggregative behavior of ants of the species *polyergus rufescens*. volume 1, pp. 163–176. 2000. Spatial Heterogeneity in Ecological Models (Alcalá de Henares, 1998)
5. Brézis, H., Crandall, M.G.: Uniqueness of solutions of the initial-value problem for  $u_t - \Delta\varphi(u) = 0$ . J. Math. Pures Appl. (9) **58**(2), 153–163 (1979)
6. Burger, M., Capasso, V., Morale, D.: On an aggregation model with long and short range interactions. Nonlinear Anal. Real World Appl. **8**(3), 939–958 (2007)
7. Burger, M., Esposito, A.: Porous medium equation and cross-diffusion systems as limit of nonlocal interaction. Nonlinear Anal., 235. Paper No. 113347, 30 (2023)
8. Calvez, V., Carrillo, J.A.: Volume effects in the Keller-Segel model: energy estimates preventing blow-up. Journal de Mathématiques Pures et Appliquées **86**(2), 155–175 (2006)
9. Carrillo, J.A., Choi, Y.-P., Hauray, M.: The derivation of swarming models: mean-field limit and Wasserstein distances. In: Collective Dynamics from Bacteria to Crowds, pp. 1–46. Springer (2014)
10. Carrillo, J.A., Craig, K., Patacchini, F.S.: A blob method for diffusion. Calc. Var. Partial Differential Equations, **58**(2), 53. Paper No. 53 (2019)
11. Carrillo, J.A., Craig, K., Yao, Y.: Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In: Active Particles. Vol. 2. Advances in Theory, Models, and Applications, Model. Simul. Sci. Eng. Technol., pp. 65–108. Birkhäuser/Springer, Cham (2019)
12. Carrillo, J.A., Delgadino, M.G., Desvillettes, L., Wu, J.: The Landau Equation as a Gradient Flow (2022)
13. Carrillo, J.A., Delgadino, M.G., Wu, J.: Boltzmann to Landau from the gradient flow perspective. Nonlinear Anal., **219**, 49. Paper No. 112824 (2022)
14. Carrillo, J.A., Di Francesco, M., Figalli, A., Laurent, T., Slepcev, D.: Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. Duke Math. J. **156**(2), 229–271 (2011)
15. Carrillo, J.A., Di Francesco, M., 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)
16. Carrillo, J.A., Hu, J., Wang, L., Wu, J.: A particle method for the homogeneous Landau equation. J. Comput. Phys. **X**, 7, 100066 (2020)
17. Carrillo, J.A., Jüngel, A., Markowich, P.A., Toscani, G., Unterreiter, A.: Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities. Monatsh. Math. **133**(1), 1–82 (2001)
18. Carrillo, J.A., Matthes, D., Wolfram, M.-T.: Lagrangian schemes for Wasserstein gradient flows. In: Geometric Partial Differential Equations. Part II, Volume 22 of Handb. Numer. Anal., pp. 271–311. Elsevier/North-Holland, Amsterdam, [2021] (2021)

19. Carrillo, J.A., McCann, R.J., Villani, C.: Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana* **19**(3), 971–1018 (2003)
20. 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)
21. Carrillo, J.A., Moll, J.S.: Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms. *SIAM J. Sci. Comput.*, **31**(6):4305–4329 (2009/10)
22. Carrillo, J.A., Murakawa, H., Sato, M., Togashi, H., Trush, O.: A population dynamics model of cell-cell adhesion incorporating population pressure and density saturation. *J. Theor. Biol.* **474**, 14–24 (2019)
23. Carrillo, J.A., Ranetbauer, H., Wolfram, M.-T.: Numerical simulation of nonlinear continuity equations by evolving diffeomorphisms. *J. Comput. Phys.* **327**, 186–202 (2016)
24. Chen, L., Daus, E.S., Holzinger, A., Jüngel, A.: Rigorous derivation of population cross-diffusion systems from moderately interacting particle systems. *J. Nonlinear Sci.* **31**(6), 38. Paper No. 94 (2021)
25. Craig, K.: Nonconvex gradient flow in the Wasserstein metric and applications to constrained nonlocal interactions. *Proc. Lond. Math. Soc.* (3) **114**(1), 60–102 (2017)
26. Craig, K., Bertozzi, A.L.: A blob method for the aggregation equation. *Math. Comp.* **85**(300), 1681–1717 (2016)
27. Craig, K., Elamvazhuthi, K., Haberland, M., Turanova, O.: A blob method for inhomogeneous diffusion with applications to multi-agent control and sampling. *Math. Comp.* **92**(344), 2575–2654 (2023)
28. Dahlberg, B.E.J., Kenig, C.E.: Nonnegative solutions of the porous medium equation. *Comm. Partial Differ. Equ.* **9**(5), 409–437 (1984)
29. Daneri, S., Radici, E., Runa, E.: Deterministic particle approximation of aggregation-diffusion equations on unbounded domains. *J. Differ. Equ.* **312**, 474–517 (2022)
30. Degond, P., Mustieles, F.-J.: A deterministic approximation of diffusion equations using particles. *SIAM J. Sci. Statist. Comput.* **11**(2), 293–310 (1990)
31. Di Francesco, M., Esposito, A., Schmidtchen, M.: Many-particle limit for a system of interaction equations driven by Newtonian potentials. *Calc. Var. Partial Differ. Equ.* **60**(2), 44. Paper No. 68 (2021)
32. Di Francesco, M., Fagioli, S.: Measure solutions for nonlocal interaction pdes with two species. *Nonlinearity* **26**, 2777–2808 (2013)
33. Dolbeault, J., Nazaret, B., Savaré, G.: A new class of transport distances between measures. *Calc. Var. Partial Differ. Equ.* **34**(2), 193–231 (2009)
34. Figalli, A., Philipowski, R.: Convergence to the viscous porous medium equation and propagation of chaos. *ALEA Lat. Am. J. Probab. Math. Stat.*, **4**, 185–203 (2008)
35. Gamba, A., Ambrosi, D., Coniglio, A., de Candia, A., Di Talia, S., Giraudo, E., Serini, G., Preziosi, L., Bussolino, F.: Percolation, Morphogenesis, and Burgers Dynamics in Blood Vessels Formation. *Phys. Rev. Lett.* **90**(11), 118101 (2003)
36. Gosse, L., Toscani, G.: Lagrangian numerical approximations to one-dimensional convolution-diffusion equations. *SIAM J. Sci. Comput.* **28**(4), 1203–1227 (2006)
37. Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29**(1), 1–17 (1998)
38. Kalašnikov, A.S.: The differential properties of the generalized solutions of equations of the type of nonstationary filtration. *Vestnik Moskov. Univ. Ser. I Mat. Meh.*, 29(1):62–68, (1974). Collection of articles dedicated to the memory of Ivan Georgievič Petrovskii. (Russian)
39. Lifshitz, E.M.: Perspectives in theoretical physics. Pergamon Press, Oxford, (1992). The collected papers of E. M. Lifshitz [E. M. Lifshits], Edited by L. P. Pitaevskii, With an introduction by D. ter Haar, With a biography of Lifshitz by Ya. B. Zel'dovich and M. I. Kaganov, translated by J. B. Sykes
40. Lions, P.-L., Mas-Gallic, S.: Une méthode particulière déterministe pour des équations diffusives non linéaires. *C. R. Acad. Sci. Paris Sér. I Math.*, **332**(4), 369–376 (2001)
41. Matthes, D., McCann, R., Savaré, G.: A family of fourth order equations of gradient flow type. *Comm. P.D.E.*, **34**(11), 1352–1397 (2009)
42. McCann, R.J.: A convexity principle for interacting gases. *Adv. Math.* **128**(1), 153–179 (1997)
43. Morale, D., Capasso, V., Oelschläger, K.: An interacting particle system modelling aggregation behavior: from individuals to populations. *J. Math. Biol.* **50**(1), 49–66 (2005)
44. Oelschläger, K.: Large systems of interacting particles and the porous medium equation. *J. Differ. Equ.* **88**(2), 294–346 (1990)
45. Oelschläger, K.: Large systems of interacting particles and the porous medium equation. *J. Differ. Equ.* **88**(2), 294–346 (1990)
46. Oelschläger, K.: A sequence of integro-differential equations approximating a viscous porous medium equation. *Z. Anal. Anwendungen* **20**(1), 55–91 (2001)

47. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differ. Equ.* **26**(1–2), 101–174 (2001)
48. Painter, K., Hillen, T.: Volume-filling and quorum-sensing in models for chemosensitive movement. *Can. Appl. Math. Q.* **10**(4), 501–544 (2002)
49. Philipowski, R.: Interacting diffusions approximating the porous medium equation and propagation of chaos. *Stochas. Process. Appl.* **117**(4), 526–538 (2007)
50. Rossi, R., Savaré, G.: Tightness, integral equicontinuity and compactness for evolution problems in banach spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, **2**(2), 395–431 (2003)
51. Royden, H.L.: *Real Analysis*. Macmillan Publishing Company, New York, third edition (1988)
52. Russo, G.: Deterministic diffusion of particles. *Comm. Pure Appl. Math.* **43**(6), 697–733 (1990)
53. Sandier, E., Serfaty, S.: Gamma-convergence of gradient flows with applications to Ginzburg-Landau. *Comm. Pure Appl. Math.* **57**(12), 1627–1672 (2004)
54. Santambrogio, F.: *Optimal Transport for Applied Mathematicians*. Progress in Nonlinear Differential Equations and Their Applications, vol. 87. Birkhäuser Verlag, Basel (2015)
55. Serfaty, S.: Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *Discrete Contin. Dyn. Syst.* **31**(4), 1427–1451 (2011)
56. Topaz, C.M., Bertozzi, A.L., Lewis, M.A.: A nonlocal continuum model for biological aggregation. *Bull. Math. Biol.* **68**(7), 1601–1623 (2006)
57. Vázquez, J.L.: An introduction to the mathematical theory of the porous medium equation. In *Shape optimization and free boundaries* (Montreal, PQ, 1990), vol. 380 of NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci., pp. 347–389. Kluwer Acad. Publ., Dordrecht (1992)
58. Vázquez, J.L.: *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, (2007). Mathematical theory
59. Véron, L.: Effets régularisants de semi-groupes non linéaires dans des espaces de Banach. *Ann. Fac. Sci. Toulouse Math. (5)* **1**(2), 171–200 (1979)
60. Villani, C.: *Topics in optimal transportation*. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI (2003)
61. Villani, C.: *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin (2009)
62. Zel’dovich, Y.B., Kompaneets, A.: Towards a theory of heat conduction with thermal conductivity depending on the temperature. *Collection of Papers Dedicated to 70th Birthday of Academician AF Ioffe*, *Izd. Akad. Nauk SSSR, Moscow*, pp. 61–71 (1950)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.