NUMERICAL STUDY OF A PARTICLE METHOD FOR GRADIENT FLOWS

J. A. CARRILLO, Y. HUANG, F. S. PATACCHINI, AND G. WOLANSKY

Abstract. We study the numerical behaviour of a particle method for gradient flows involving linear and nonlinear diffusion. This method relies on the discretisation of the energy via non-overlapping balls centred at the particles. The resulting scheme preserves the gradient flow structure at the particle level and enables us to obtain a gradient descent formulation after time discretisation. We give several simulations to illustrate the validity of this method, as well as a detailed study of one-dimensional aggregation-diffusion equations.

1. Introduction

In this work we introduce a new particle method for approximating the solutions to evolution equations of the form

$$\begin{cases} \rho_t = \nabla \cdot \big[\rho \nabla \big(H'(\rho(x)) + V(x) + W * \rho(x)\big)\big], \quad t > 0,\, x \in \mathbb{R}^d, \\ \rho(0,\cdot) = \rho_0(\cdot), \end{cases} \tag{1.1}$$

where $\rho(t,\cdot) \geq 0$ is the unknown probability measure and ρ_0 is a fixed element of $\mathcal{P}_2(\mathbb{R}^d)$, the set of Borel probability measures on $\mathbb{R}^d$ with bounded second moment. Note that we denote by the same symbol a probability measure and its density, whenever the latter exists. The operator $*$ denotes the convolution, $H\colon [0,\infty) \to \mathbb{R}$ is the *density of internal energy*, $V\colon \mathbb{R}^d \to \mathbb{R}$ is the *confinement potential*, and $W\colon \mathbb{R}^d \to \mathbb{R}$ is the *interaction potential*. These equations are ubiquitous in many applications, ranging from granular media and porous medium flows to collective behaviour models in mathematical biology and self-assembly, see [6, 8, 12, 36, 37] and the references therein.

Recent advances in the analysis of the equation (1.1) are mainly based on variational schemes using the natural gradient flow structure in the space of probability measures, see e.g. [26, 21, 31, 38, 1, 15]. We define the *continuum energy functional* $E\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{-\infty, +\infty\}$ by

$$E(\rho) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \big[H(\rho(x)) + V(x)\rho(x) + \tfrac{1}{2} W * \rho(x)\rho(x)\big]\,\mathrm{d}x & \text{if } \rho \in \mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d), \\ \displaystyle\int_{\mathbb{R}^d} \big[V(x) + \tfrac{1}{2} W * \rho(x)\big]\,\mathrm{d}\rho(x) & \text{if } \rho \notin \mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d) \\ & \quad \text{and } H = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$ is the subset of $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures which are absolutely continuous with respect to the Lebesgue measure. The functions H, V and W satisfy the following hypotheses.

Hypothesis 1. *V is a function in $C^1(\mathbb{R}^d)$ and W is a symmetric, locally integrable function in $C^1(\mathbb{R}^d \setminus \{0\})$ with $W(0) = 0$.*

Hypothesis 2. *H is a convex function in $C^1((0,\infty)) \cap C^0([0,\infty))$ with superlinear growth at infinity and $H(0) = 0$. Furthermore, $h(\lambda) := \lambda^d H\left(\lambda^{-d}\right)$ is convex and non-increasing on $(0,\infty)$.*

Date: 6 December 2016.

2010 *Mathematics Subject Classification.* 65M12, 35K05.

Key words and phrases. Particle method, diffusion, aggregation, gradient flow, discrete gradient flow, JKO scheme.

The assumption $W(0) = 0$ is made without loss of generality. Indeed, if $W(0)$ is finite, then W can be shifted "up" or "down" to get $W(0) = 0$; if W has a singularity at 0, then setting $W(0) := 0$ does not affect the physical behaviour of a system governed by the potential W. The assumptions that $H(0) = 0$ and h is convex and non-increasing imply that the energy E is displacement convex if, for example, $V = W = 0$, see [26], [25, Section 4] and [38, Theorem 5.15]. Also note that the classical cases $H(\rho) = \rho \log \rho$ and $H(\rho) = \frac{\rho^m}{m-1}$ $(m > 1)$ satisfy all the required assumptions.

The underlying topology on the probability measures in this paper is given by the *quadratic Wasserstein distance* $d_2(\rho, \mu)$, which is defined between two measures ρ and μ in $\mathcal{P}_2(\mathbb{R}^d)$ by

$$d_2(\rho, \mu) = \inf_{\gamma \in \Pi(\rho,\mu)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\gamma(x, y) \right]^{\frac{1}{2}},$$

where $\Pi(\rho, \mu)$ is the space of probability measures (also called transport plans) on $\mathbb{R}^d \times \mathbb{R}^d$ with first marginal ρ and second marginal μ. Let us fix a final time $T > 0$. We say that $\rho \colon [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$ is a *continuum gradient flow solution* with initial condition ρ_0 if

$$\begin{cases} \rho'(t) = -\nabla_{\mathcal{P}_2(\mathbb{R}^d)} E(\rho(t)), \\ \rho(0) = \rho_0, \end{cases} \tag{1.2}$$

holds in the sense of distributions on $[0, T] \times \mathbb{R}^d$, see [1, Equation (8.3.8)]. The operator $\nabla_{\mathcal{P}_2(\mathbb{R}^d)}$ is the classical *quadratic Wasserstein gradient* on $\mathcal{P}_2(\mathbb{R}^d)$, which takes the explicit form

$$\nabla_{\mathcal{P}_2(\mathbb{R}^d)} E(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\delta E}{\delta \rho} \right) \quad \text{for all } \rho \in \mathcal{P}_2(\mathbb{R}^d),$$

where $\frac{\delta E}{\delta \rho} = H'(\rho) + V + W * \rho$ is the first variation density of E at point ρ. As a by-product of the general theory developed for instance in [1], gradient flow solutions to (1.2) are weak solutions to (1.1). Note that if theoretical issues such as the existence and uniqueness of solutions to the continuum gradient flow (1.2) are of interest, appropriate additional assumptions must be imposed on H, V, W and ρ_0, see [21, 38, 1].

We propose below to approximate solutions to the continuum gradient flow (1.2) by finite atomic probability measures represented by finite numbers of particles. The basic idea is to restrict the continuum gradient flow to the discrete setting of atomic measures by performing the steepest descent of a suitable approximation of the energy E defined on finite numbers of Dirac masses, and apply a discrete analogue of the JKO scheme for the Fokker-Planck equation proposed by Jordan, Kinderlehrer and Otto [21]. The theoretical underpinning of this method is studied in the companion paper [17], where the convergence of the discrete gradient flow to the continuum one is proved in the framework proposed by Serfaty in [35, 34], in the special case of $V = W = 0$ in one dimension for equally-weighted particles and under additional appropriate hypotheses on H. The goal of the present paper is to give numerical evidence of such convergence and to motivate possible extensions of the theoretical result in [17] to nonzero confinement and interaction potentials (even with possible singularities), as well as to higher dimensions and unequally-weighted particles.

Let us mention that other numerical methods have been developed to conserve particular properties of solutions of the continuity equation (1.1). In [7, 12] the authors developed finite-volume methods preserving the decay of energy at the semi-discrete level, along with other important properties like non-negativity and mass conservation. Particle methods for these equations without the diffusion term are known to be convergent under suitable assumptions on the potentials V and W since these results are connected to the question of the mean-field limit and rigorous derivation of the equations from particle trajectories, see [13] for instance. In the case of diffusion equations, particle methods based on suitable regularisations of the flux of the continuity equation (1.1) have been proposed in [32, 18, 23, 24]; note that an early particle method was derived for collisional

kinetic equations in [33]. Our steepest descent method is purely variational and is based on regularising the internal part of the energy E by substituting particles by non-overlapping blobs, as we discuss next. Let us mention that the numerical approximation of the JKO variational scheme has already been tackled by different methods using pseudo-inverse distributions in one dimension [19, 8, 10], diffeomorphisms [16], or solving for the optimal map in a JKO step [5]. Our method avoids these computationally intensive procedures by the approximation of the energy in the discrete setting. Finally, note that gradient-flow-based Lagrangian methods for higher-order, drift diffusion and Fokker-Planck equations have recently been proposed in [29, 30, 28, 22].

The paper is structured as follows. In Section 2 we give a summary of the method and the derivation of the discrete gradient flow in a more general setting than that considered in [17]. Section 3 is dedicated to the derivation of the numerical scheme used to approximate the continuum gradient flow (1.2), that is an explicit version of the JKO scheme, as well as to a numerical validation study of the scheme via diffusion equations. We stress that, from the scientific computing point of view, this is a preliminary study whose aim is to motivate the presented particle method. Finally, in Section 4 we give the numerical results for various one-dimensional aggregation-diffusion equations—we emphasise that this particle method, despite its simplicity, is able to capture the critical mass for the modified one-dimensional Keller-Segel model.

2. Particle method and discrete gradient flows

In this method, the underlying probability measure is characterised by the particles' positions $(x_1, \dots, x_N) \in (\mathbb{R}^d)^N = \mathbb{R}^{Nd}$ and the associated weights (or masses) $w := (w_1, \cdots, w_N) \in \mathbb{R}^N$, where N is the total number of particles considered. Throughout this paper, the positions $(x_1, \dots, x_N)$ are evolving in time but the weights w are fixed and such that $w_i > 0$ and $\sum_{i=1}^N w_i = 1$. Also, we denote by $\mathbb{R}^{Nd}_w$ the space of particles with weights w, that is, $\boldsymbol{x} := (x_1, \dots, x_N) \in \mathbb{R}^{Nd}_w$ means that each particle x_i is in $\mathbb{R}^d$ and is associated with the weight w_i. Notice the boldface font when referring to elements of $\mathbb{R}^{Nd}_w$.

Remark 2.1. As an important convention in the rest of the paper, whenever particles $\boldsymbol{x} \in \mathbb{R}^{Nd}_w$ are considered, they are assumed to be distinct, i.e., $x_i \neq x_j$ if $i \neq j$. Moreover, in one dimension, the particles are assumed to be sorted increasingly, i.e., $x_{i+1} > x_i$ for all $i \in \{1, \dots, N-1\}$.

2.1. Discrete gradient flow.

Consider N particles $\boldsymbol{x} := (x_1, \dots, x_N) \in \mathbb{R}^{Nd}_w$ with fixed weights w. The most natural representation of the underlying probability measure is the empirical measure

$$\boldsymbol{x} \mapsto \mu_N = \sum_{i=1}^N w_i \delta_{x_i}, \tag{2.1}$$

which belongs to the space of *atomic measures*

$$\mathcal{A}_{N,w}(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}_2(\mathbb{R}^d) \mid \exists\, \boldsymbol{x} \in \mathbb{R}^{Nd}_w, \mu = \sum_{i=1}^N w_i \delta_{x_i} \right\}.$$

Definition 2.2 (Discrete energy)**.** We define the *discrete energy* $E_N \colon \mathcal{A}_{N,w}(\mathbb{R}^d) \to \mathbb{R}$ by

$$E_N(\mu_N) = \sum_{i=1}^N |B_i| H\left(\frac{w_i}{|B_i|}\right) + \sum_{i=1}^N w_i V(x_i) + \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N w_i w_j W(x_i - x_j), \tag{2.2}$$

where B_i is the open ball of centre x_i and radius $\frac{1}{2} \min_{j \neq i} |x_i - x_j|$, and $|B_i|$ is the volume of B_i.

Note that E_N is finite over the whole $\mathcal{A}_{N,w}(\mathbb{R}^d)$ since, by the Hypotheses 1 and 2, H, V and W are pointwise finite. The essence of this discrete approximation E_N on $\mathcal{A}_{N,w}(\mathbb{R}^d)$ of the continuum

energy E on $\mathcal{P}_2(\mathbb{R}^d)$ lies in the treatment of the internal part $\int_{\mathbb{R}^d} H\big(\rho(x)\big)\,\mathrm{d}x$ of the energy E, which becomes infinity on point masses; the point mass of each particle is uniformly spread to circumvent this problem. To this end, consider

$$\boldsymbol{x} \mapsto \rho_N = \sum_{i=1}^{N} w_i \frac{\chi_{B_i}}{|B_i|}, \tag{2.3}$$

where χ_{B_i} is the characteristic function of the ball B_i. Clearly ρ_N is in $\mathcal{P}_{\mathrm{ac},2}(\mathbb{R}^d)$, and thus the internal part of the energy is well-defined for ρ_N. Note that the representation ρ_N does not involve overlapping of balls, but contains "gaps" between balls whose sizes are expected to decrease as the number of particles increases. Then, using (2.3) for the internal part of the energy and (2.1) for the confinement and interaction parts, E_N defined in (2.2) is exactly

$$\int_{\mathbb{R}^d} H(\rho_N(x))\,\mathrm{d}x + \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\mu_N(x) + \int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\,\mathrm{d}\mu_N(y)\,\mathrm{d}\mu_N(x).$$

Here the diagonal terms in the interaction potential vanish since $W(0)=0$ by Hypothesis 1. This choice of non-overlapping particles has the main advantage of reducing the computational cost of the internal part of the discrete energy functional. One can already notice that this approach allows us to treat diffusive effects simultaneously with confinement and interaction ones in a very natural way; this is another advantage of our method, as it becomes clearer throughout this paper.

Since the expression above depends essentially on $\boldsymbol{x} \in \mathbb{R}^{Nd}_w$, we can define the discrete energy equivalently as a function of $\boldsymbol{x} \in \mathbb{R}^{Nd}_w$ instead of $\mu_N \in \mathcal{A}_{N,w}(\mathbb{R}^d)$:

$$\widetilde{E}_N(\boldsymbol{x}) := E_N(\mu_N) \quad \text{for all } \mu_N \in \mathcal{A}_{N,w}(\mathbb{R}^d) \text{ with particles } \boldsymbol{x} \in \mathbb{R}^{Nd}_w. \tag{2.4}$$

Definition 2.3 (Subdifferential in Hilbert spaces)**.** Let X be a Hilbert space with inner product $\langle\cdot,\cdot\rangle_X$ and $\phi\colon X \to \mathbb{R}$. We define the *subdifferential* $\partial\phi\colon X \to 2^X$ of ϕ, for all $x \in X$, by the subset

$$\partial\phi(x) = \left\{ z \in X \mid \liminf_{y\to x} \frac{\phi(y) - \phi(x) - \langle z, y-x\rangle_X}{|y-x|_X} \geq 0 \right\}.$$

If $\partial\phi(x)$ is not empty, then $\partial^0\phi(x)$ is defined to be the unique element in $\partial\phi(x)$ with minimal norm.

Now, we say that $\boldsymbol{x}\colon [0,T] \to \mathbb{R}^{Nd}_w$ is a *discrete gradient flow solution* with initial condition $\boldsymbol{x^0} \in \mathbb{R}^{Nd}$ if the differential inclusion

$$w \cdot \boldsymbol{x}'(t) \in -\partial\widetilde{E}_N(\boldsymbol{x}(t)) \qquad \text{for almost all } t \in (0,T] \tag{2.5}$$

is satisfied, and $\boldsymbol{x}(0) = \boldsymbol{x^0}$. Here $\boldsymbol{x}'(t)$ is the velocity of the curve $\boldsymbol{x}(t)$ and $w\cdot\boldsymbol{x}'$ is the element-wise product $(w_1x_1', \dots, w_Nx_N')$. This formulation is not a standard differential inclusion because of the presence of the weights. To cope with this, we introduce the following inner product on $\mathbb{R}^{Nd}_w$.

Definition 2.4 (Weighted inner product on $\mathbb{R}^{Nd}_w$)**.** For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{Nd}_w$ we define the *weighted inner product* between $\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle_w = \sum_{i=1}^{N} w_i \langle x_i, y_i\rangle_{\mathbb{R}^d}.$$

From now on, the Euclidean space $\mathbb{R}^{Nd}_w$ is endowed with this inner product. This definition clearly induces the weighted norm

$$|\boldsymbol{x}|_w := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle_w} = \sqrt{\textstyle\sum_{i=1}^{N} w_i |x_i|^2_{\mathbb{R}^d}} \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^{Nd}_w. \tag{2.6}$$

It also induces a slightly more general subdifferential structure: for any functional $\phi\colon \mathbb{R}^{Nd}_w \to \mathbb{R}$,

$$\partial_w\phi(\boldsymbol{x}) := \left\{ \boldsymbol{z} \in \mathbb{R}^{Nd}_w \mid w\cdot\boldsymbol{z} \in \partial\phi(\boldsymbol{x}) \right\} \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^{Nd}_w,$$

where again $w \cdot \boldsymbol{z}$ is the element-wise product between w and $\boldsymbol{z}$; and we can then define the element $\partial_w^0 \phi(\boldsymbol{x})$ with minimal norm accordingly. The discrete gradient flow inclusion (2.5) can now be rewritten more naturally as

$$\begin{cases} \boldsymbol{x}'(t) \in -\partial_w \widetilde{E}_N(\boldsymbol{x}(t)) & \text{for almost all } t \in (0,T], \\ \boldsymbol{x}(0) = \boldsymbol{x^0}. \end{cases} \tag{2.7}$$

This is the discrete gradient flow structure that is used from now on.

Remark 2.5. In [17] the authors showed that in one dimension with $V = W = 0$ the discrete energy E_N Γ-converges (in the d_2-topology, see Definition A.3) to the continuum one E if adequate additional assumptions are imposed on H.

2.2. **p-approximated discrete gradient flow.** The gradient flow (2.7) is written as a differential inclusion instead of an ordinary differential equation, primarily because the radius $\frac{1}{2}\min_{j\neq i}|x_i - x_j|$ of the balls B_i is not a smooth function of the positions in the definition of the energy $\widetilde{E}_N$. As a result, the energy $\widetilde{E}_N$ is not a smooth function of the positions $\boldsymbol{x}$ when the radius of the ball B_i is determined by more than one particle.

If $\widetilde{E}_N$ is lower semi-continuous and convex, and $\boldsymbol{x^0} \in \{\boldsymbol{y} \in \mathbb{R}_w^{Nd} \mid \partial_w \widetilde{E}_N(\boldsymbol{y}) \neq \emptyset\}$, it is known by [4, Theorem 1 of Section 3.2] that the discrete gradient flow (2.7) is well-posed and the equation $\boldsymbol{x}'(t) = -\partial_w^0 \widetilde{E}_N(\boldsymbol{x}(t))$ is satisfied for almost every $t \in [0,T]$. Note that, in our case, the lower semi-continuity of $\widetilde{E}_N$ is trivial to check by the assumptions on H, whereas the convexity is proved in Proposition A.6 for convex potentials V and W in one dimension. Although the element of minimal norm $\partial_w^0 \widetilde{E}_N(\boldsymbol{x}(t))$ can be computed in some situations as in [17] (for the special case of equally-weighted particles with internal energy only and in one dimension), the detailed procedure is usually involved and is not convenient for numerical implementation. In theory, one can approximate $\widetilde{E}_N$ by a differentiable function usually characterised by a parmeter, where the minimal norm element of the subdifferential is recovered as the limit of the gradient of the approximation as the parameter goes to infinity. A common choice of such an approximation is

$$Y_N^p(\boldsymbol{x}) := \inf_{\boldsymbol{y} \in \mathbb{R}_w^{Nd}} \left(\widetilde{E}_N(\boldsymbol{y}) + \tfrac{p}{2} |\boldsymbol{x} - \boldsymbol{y}|_w^2 \right),$$

depending on the parameter $p > 0$. Then Y_N^p is differentiable almost everywhere in $\mathbb{R}_w^{Nd}$, ∇Y_N^p is called the *Yosida approximation* of $\partial_w \widetilde{E}_N$, and

$$\nabla Y_N^p(\boldsymbol{x}) \xrightarrow[p\to\infty]{} \partial_w^0 \widetilde{E}_N(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in \mathbb{R}_w^{Nd},$$

see [4, Theorem 2 of Section 3.1 and Theorem 4 of Section 3.4]. However, the Yosida approximation requires another optimisation, and is therefore not very well adapted to the computation of $\partial_w^0 \widetilde{E}_N(\boldsymbol{x})$. An alternative is to approximate the radius of the balls B_i in the definition of $\widetilde{E}_N$ by a smooth function.

Definition 2.6 (p-approximation of the minimum function)**.** For $p > 0$ and integer $s \geq 2$, the function $\min_p \colon (0,\infty)^s \to (0,\infty)$ is defined as

$$\min_p(x_1, \ldots, x_s) = \begin{cases} \left(\dfrac{1}{s} \displaystyle\sum_{i=1}^{s} x_i^{-p} \right)^{-\frac{1}{p}} & \text{if } x_1, \ldots, x_s \in (0,\infty), \\ 0 & \text{otherwise.} \end{cases}$$

We call $\min_p(x_1, \ldots, x_s)$ the *p-approximation* of the minimum function $\min(x_1, \cdots, x_s)$.

The following proposition, whose proof is straightforward and left to the reader, justifies the use of the p-approximation of the minimum function.

Proposition 2.7. *Let $s \geq 2$. The following statements hold.*

(1) Let $p > 0$. Then $\min_p \in C^\infty((0,\infty)^s)$.

(2) $\min_p(x) \to \min(x)$ as $p \to \infty$ for all $x \in (0,\infty)^s$.

(3) $p \mapsto \min_p(x)$ is non-increasing on $(0,\infty)$ for all $x \in (0,\infty)^s$.

(4) Let $p > 0$. Then $\min_p(x) \geq \min(x)$ for all $x \in (0,\infty)^s$, with equality if $x_i = x_j$ for all i, j.

(5) Let $p \geq 1$. Then $\min_p(x) \leq s^{1/p} \min(x)$ for all $x \in (0,\infty)^s$.

Now the discrete energy is further approximated using a smoothed radius as below.

Definition 2.8 (p-approximated discrete energy)**.** Let $p > 0$. The *p-approximated discrete energy* $E_N^p \colon \mathcal{A}_{N,w}(\mathbb{R}^d) \to \mathbb{R}$ is defined by

$$E_N^p(\mu_N) = \sum_{i=1}^N |B_i^p| H\left(\frac{w_i}{|B_i^p|}\right) + \sum_{i=1}^N w_i V(x_i) + \frac{1}{2}\sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j W(x_i - x_j), \tag{2.8}$$

for all $\mu_N \in \mathcal{A}_{N,w}(\mathbb{R}^d)$ with particles $\boldsymbol{x} \in \mathbb{R}_w^{Nd}$. Here B_i^p is the ball with the new radius

$$\tfrac{1}{2}\min_p\{|x_i - x_j| \mid j \neq i\} = \tfrac{1}{2}\left(\tfrac{1}{N-1}\textstyle\sum_{\substack{j=1 \\ j\neq i}}^N |x_i - x_j|^{-p}\right)^{-\frac{1}{p}}.$$

As in (2.4), we can define the p-approximated discrete energy on $\mathbb{R}_w^{Nd}$ rather than on $\mathcal{A}_{N,w}(\mathbb{R}^d)$:

$$\widetilde{E}_N^p(\boldsymbol{x}) := E_N^p(\mu_N) \quad \text{for all } \mu_N \in \mathcal{A}_{N,w}(\mathbb{R}^d) \text{ with particles } \boldsymbol{x} \in \mathbb{R}_w^{Nd}. \tag{2.9}$$

We say that $\boldsymbol{x} \colon [0,T] \to \mathbb{R}_w^{Nd}$ is a *p-approximated discrete gradient flow solution* with initial condition $\boldsymbol{x^0} \in \mathbb{R}_w^{Nd}$ if

$$\begin{cases} \boldsymbol{x}'(t) = -\nabla_w \widetilde{E}_N^p(\boldsymbol{x}(t)) & \text{for almost all } t \in (0,T], \\ \boldsymbol{x}(0) = \boldsymbol{x^0}, \end{cases} \tag{2.10}$$

where the weighted gradient ∇_w is naturally defined on $\mathbb{R}_w^{Nd}$ by

$$\nabla_w := \left(\frac{1}{w_1}\frac{\partial}{\partial x_1}, \dots, \frac{1}{w_N}\frac{\partial}{\partial x_N}\right).$$

Remark 2.9. Whenever $d = 1$ and V and W are convex, we know by Proposition A.6 that $\widetilde{E}_N^p$ is convex, and therefore the p-approximated discrete gradient flow (2.10) is well-posed.

Remark 2.10. The interest of the p-approximation of the discrete gradient flow only lies in the numerical simplicity of coding a gradient descent on an ODE system such as (2.10) rather than on a differential inclusion such as (2.7). The p-approximation is indeed not needed for the theoretical proof given in [17] of the convergence of the discrete gradient flow to the continuum one, since, as already mentioned, in that specific case the minimal norm element of $\partial\widetilde{E}_N$ is explicitly computable.

Remark 2.11. Approximations of the minimum function other than the one given in Definition 2.6 are possible. One example is

$$\min_p(x_1, \dots, x_s) = \begin{cases} -\frac{1}{p}\log\left(\frac{1}{s}\sum_{i=1}^s e^{-px_i}\right) & \text{if } x_1, \dots, x_s \in (0,\infty), \\ 0 & \text{otherwise.} \end{cases}$$

The properties of $\min_p$ given in Proposition 2.7 are not enough to say that the p-approximated discrete gradient flow (2.10) converges to the discrete one (2.7) as $p \to \infty$ in some sense. This still needs to be checked if we want to justify the numerical use of the p-approximated gradient flow. In the next section and in Appendix A we prove that this is the case in dimension one since we can there exploit the convexity of the discrete energies.

2.3. **One-dimensional case.**

Definition 2.12 (Inter-particle distance)**.** For any particles $\boldsymbol{x} \in \mathbb{R}^N_w$ we denote the *inter-particle distance* by the positive quantity (eventually $+\infty$ by convention)

$$\Delta x_i := x_i - x_{i-1} \quad \text{for } i \in \{1, \dots, N+1\},$$

with the convention $x_0 = -\infty$ and $x_{N+1} = +\infty$. We also write, for $p > 0$,

$$r_i = \min(\Delta x_i, \Delta x_{i+1}) \quad \text{and} \quad r_i^p = \min_p(\Delta x_i, \Delta x_{i+1}) \quad \text{for all } i \in \{1, \dots, N\}.$$

Note that $r_i = |B_i|$ and $r_i^p = |B_i^p|$.

Since particles are sorted increasingly by convention, the p-approximated discrete energy (2.8) in dimension one can be defined in the simpler form

$$E_N^p(\mu_N) = \sum_{i=1}^N r_i^p H\left(\frac{w_i}{r_i^p}\right) + \sum_{i=1}^N w_i V(x_i) + \frac{1}{2}\sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j W(x_i - x_j), \tag{2.11}$$

for all $\mu_N \in \mathcal{A}_{N,w}(\mathbb{R})$ with particles $\boldsymbol{x} \in \mathbb{R}^{Nd}_w$. The equivalent formulation of the energy $\widetilde{E}_N^p$ on the particles is defined in dimension one accordingly, see (2.9).

In order to check that the p-approximated discrete gradient flow defined above indeed approximates the discrete one, we need a few elements of maximal monotone operator theory (see [3, 4, 9] for a detailed overview of the theory). These notions and results being quite abstract, they are only given in Appendix A. There we show how to use this theory to prove the convergence of the p-approximated discrete gradient flow to the discrete one in a precise sense, in the case when V and W are convex; we therefore justify the numerical use of the p-approximated gradient flow (2.10) (at least in the case when V and W are convex and $d = 1$).

Remark 2.13. In this method, one reason why we decide to discretise according to non-overlapping balls, rather than, for example, Voronoi cells (see for instance [20] for a detailed account on Voronoi cells), is to allow for simpler computations of the discrete energies, which in turn gives rise to less costly simulations. In fact, this is not very relevant in one dimension since then the implementation of Voronoi cells is actually not more costly than that of non-overlapping balls; the one-dimensional simulations given below, which are run according to the non-overlapping balls described above, should therefore be seen as an initial validation of the method and as test cases which need to be extended to higher dimensions in a future work.

3. Numerical scheme and validation

Here a general variational formulation is applied to the discrete setting described above, leading to our numerical scheme giving the particles' positions at discrete time steps (see [1, 2, 21] for a Fokker-Planck motivation and its generalisation to curves in probability spaces). For the sake of generality the derivation of the scheme is partly done for the discrete gradient flow, rather than the p-approximated one. The simulations given later were, however, performed in the p-approximated setting only. Note also that, in this section, some indices N and p are dropped for clarity.

3.1. **The scheme.** Take $(t_n)_{n=0}^M \subset [0, T]$ a subdivision of the time interval $[0, T]$ and a time-step size $\Delta t = \Delta_n t$ (eventually adaptive) such that $t_n = t_0 + \sum_{i=0}^{n-1} \Delta_i t$ for all $n \in \{1, \dots, M\}$. Suppose that, for some $n \in \{0, \dots, M-1\}$, we know $\mu^n \in \mathcal{A}_{N,w}(\mathbb{R}^d)$, where μ^n is the empirical measure of an approximated discrete gradient flow solution at time t_n. Then we want to get an approximated

discrete gradient flow solution at the time t_{n+1}, that is μ^{n+1}. To this end we use the JKO scheme, i.e., we choose μ^{n+1} as

$$\mu^{n+1} := \operatorname*{argmin}_{\mu \in \mathcal{A}_{N,w}(\mathbb{R}^d)} \left\{ \frac{1}{2\Delta t} d_2^2(\mu^n, \mu) + E_N(\mu) \right\}, \tag{3.1}$$

where we recall that d_2 is the quadratic Wasserstein distance. In the following any object with sub- or superscript n or $n+1$ is associated with the time step n or $n+1$, respectively.

From now on, apart from a two-dimensional test in Section 3.4.5, the setting is *one-dimensional*.

3.1.1. *Computation of the Wasserstein distance.* Let us compute the Wasserstein distance in (3.1). For each $n \in \{0, \dots, M-1\}$ and approximation μ^n and $\mu \in \mathcal{A}_{N,w}(\mathbb{R})$, we directly get

$$d_2^2(\mu^n, \mu) = \sum_{i=1}^N w_i (x_i^n - x_i)^2, \tag{3.2}$$

where $\boldsymbol{x}^n = (x_1^n, \dots, x_N^n)$ and $\boldsymbol{x} = (x_1, \dots, x_N)$ are the particles of μ^n and μ, respectively, which leads to the scheme

$$\mu^{n+1} := \operatorname*{argmin}_{\mu \in \mathcal{A}_{N,w}(\mathbb{R})} \left(\sum_{i=1}^N w_i \frac{(x_i^n - x_i)^2}{2\Delta t} + E_N(\mu) \right). \tag{3.3}$$

Clearly the scheme (3.3) on the empirical measures can be equivalently rewritten on the particles:

$$w \cdot \boldsymbol{x}^{n+1} := \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}_w^N} \left(\sum_{i=1}^N w_i \frac{(x_i^n - x_i)^2}{2\Delta t} + \widetilde{E}_N(\boldsymbol{x}) \right), \tag{3.4}$$

where $w \cdot \boldsymbol{x}^{n+1}$ is the element-wise multiplication between the vectors w and $\boldsymbol{x}^{n+1}$. The minimisation problem (3.4) is characterised by

$$0 \in \partial_w \left(\sum_{i=1}^N w_i \frac{(x_i^n - x_i^{n+1})^2}{2\Delta t} + \widetilde{E}_N(\boldsymbol{x}^{n+1}) \right). \tag{3.5}$$

When $\widetilde{E}_N$ is convex, (3.5) is precisely the backward Euler scheme of the differential inclusion (2.7).

3.1.2. *Minimisation and final form of the scheme.* Let $p > 0$. We now return to the p-approximated setting, i.e., we consider (3.4) (and equivalently (3.3)) with $\widetilde{E}_N^p$ instead of $\widetilde{E}_N$ (and equivalently E_N^p instead of E_N). We want to minimise, over the whole set $\mathbb{R}_w^N$, the functional in the argmin operator in (3.4) (and (3.3)) to find our approximation $\boldsymbol{x}^{n+1}$ (and μ^{n+1}) at time step $n+1$. First note that, for each $n \in \{0, \dots, M-1\}$, we can write the discrete energy $\widetilde{E}_N^p$ as

$$\widetilde{E}_N^p(\boldsymbol{x}^{n+1}) = \sum_{i=1}^N w_i E_i^{n+1},$$

where

$$E_i^{n+1} := r_i^{n+1} H\left(\frac{w_i}{r_i^{n+1}} \right) + w_i V(x_i^{n+1}) + \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^N w_j W(x_i^{n+1} - x_j^{n+1}),$$

with $r_i^{n+1} := r_i^{p,n+1}$, see Definition 2.12. For any $n \in \{0, \dots, M-1\}$, (3.5) becomes

$$x_i^{n+1} = x_i^n - \frac{\Delta t}{w_i} \sum_{j=1}^N w_j \frac{\partial E_j^{n+1}}{\partial x_i^{n+1}}, \tag{3.6}$$

where the uninspiring computation of the derivatives in the sum terms is left to the reader. The scheme (3.6) is the implicit Euler scheme of the ODE (2.10), or equivalently, of the ODE

$$w_i \frac{\mathrm{d}x_i}{\mathrm{d}t} = -\frac{\partial \widetilde{E}_N^p}{\partial x_i}(\boldsymbol{x}), \quad i = 1, 2, \dots, N,$$

and it coincides with the JKO scheme for $\widetilde{E}_N^p$, as in (3.1). Since the implicit scheme (3.6) is difficult and costly to solve, its explicit Euler version was used in the numerical examples of this paper:

$$x_i^{n+1} = x_i^n - \frac{\Delta t}{w_i} \sum_{j=1}^{N} w_j \frac{\partial E_j^n}{\partial x_i^n}. \tag{3.7}$$

We remind the reader that the convergence of the implicit Euler version of the scheme is presented in [21] in one dimension. The stability analysis of the explicit scheme (3.7) is not dealt with as it is not the purpose of the present paper—we are not worried about time stepping stability issues for large number of particles in this preliminary stage of validation of our new approach. We can nevertheless indicate that, intuitively, a time step satisfying $\Delta t \leq \frac{C}{N^2}$ for some constant $C > 0$ should suffice to ensure stability; indeed, this would be in line with the classical CFL condition for diffusion equations where the mesh size is of order $\frac{1}{N}$. Note that this condition was respected in the simulations presented below. Let us also mention that higher-order time discretisations could be used in place of the explicit Euler scheme, and would indeed lead to a better time and space accuracy of the method.

Remark 3.1. It is worth pointing out that our particle method does not aim at being competitive against classical finite-volume and finite-difference schemes for purely diffusive equations. In fact, it aims primarily at being simple to implement—even in higher dimensions—and flexible when considering additional terms to diffusion such as confinement and interaction. In terms of complexity at each time step, the method is of order N^2 regardless of the space dimension, as it actually is for other finite-volume methods [12]. Note that, however, if we time-discretise (2.7) rather than its regularised form (2.10), the complexity becomes significantly higher in more than one dimension (at most of order N^3) since then we are required to find the closest neighbour to each particle; in dimension one the order of complexity is still N^2 thanks to the increasing ordering of the particles.

3.2. **Initialisation of the scheme.** Below we give two different ways of approximating the initial profile. Let us first introduce the notion of pseudo-inverse.

Definition 3.2 (Pseudo-inverse)**.** Let $F\colon \mathbb{R} \to [0,1]$ be a non-decreasing and right-continuous function. The *pseudo-inverse* $\Phi\colon [0,1] \to \mathbb{R} \cup \{-\infty, +\infty\}$ of F is the non-decreasing and right-continuous function defined by

$$\Phi(\varepsilon) = \inf\{x \in \mathbb{R} \mid F(x) > \varepsilon\} \quad \text{for all } \varepsilon \in [0,1].$$

3.2.1. *Initially equally-weighted particles.* If we want to approximate the initial profile $\rho_0 \in \mathcal{P}_2(\mathbb{R})$ with equally-weighted particles, we need to start with unequally-spaced particles. Consider the pseudo-inverse Φ_0 of the cumulative distribution $x \mapsto \rho_0((-\infty, x])$ of the initial profile. Suppose $w_i = 1/N$ for all $i \in \{1, \dots, N\}$. Then choose the initial particles $\boldsymbol{x^0} := (x_1^0, \dots, x_N^0) \in \mathbb{R}_w^N$ as

$$x_i^0 = \Phi_0\left(\frac{2i-1}{2N}\right) \quad \text{for all } 1 \leq i \leq N.$$

3.2.2. *Initially equally-spaced particles.* If now we want to approximate the initial profile $\rho_0 \in \mathcal{P}_2(\mathbb{R})$ with equally-spaced particles, we need to assign a different weight to each particle. Consider F_0,

the cumulative distribution of the initial profile, and $\boldsymbol{x^0} \in \mathbb{R}^N_w$, some chosen equally-spaced initial particles. Let us define

$$x^0_{i\pm\frac{1}{2}} = \tfrac{1}{2}\left(x^0_i + x^0_{i\pm1}\right) \quad \text{for all } 1 \le i \le N.$$

Then we choose the weights w as

$$\begin{cases} w_1 = F_0\left(x^0_{1+\frac{1}{2}}\right), \\ w_i = F_0\left(x^0_{i+\frac{1}{2}}\right) - F_0\left(x^0_{i-\frac{1}{2}}\right) \quad \text{for all } 1 < i < N, \\ w_N = 1 - F_0\left(x^0_{N-\frac{1}{2}}\right). \end{cases} \tag{3.8}$$

The weight w_i as given in (3.8) for each $1 < i < N$ is the mass that ρ_0 "assigns" to the interval $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, that is, to the Voronoi cell of x_i. Note that the choice we have on the initial positions of the particles $\boldsymbol{x^0}$ is not completely free of constraints: they must be chosen inside the support of ρ_0 or the resulting weights are zero. Finally, since the initial particles are chosen to be equally spaced, only the first and last particles' locations are needed to determine the locations of all the others; for the following numerical simulations we give this information under the form $I_{\text{init}} = [x_1, x_N]$.

Remark 3.3. There are two main advantages in the second initialisation approach. The first one is that it allows a better approximation of the initial profile when there is a strong variation in the density; indeed one has the freedom to place initial particles in less populated regions. The second one is that there are no gaps between the initial balls of centres x^0_i and diameters $\min_p(\Delta x^0_i, \Delta x^0_{i+1}) = \min(\Delta x^0_i, \Delta x^0_{i+1})$, see Proposition 2.7(4), allowing again a better approximation of the initial profile.

3.3. **Computation of the error.** The natural error to compute for our scheme (3.7) is the quadratic Wassertein error.

3.3.1. *Error with respect to an exact solution.* If we want to get the error between the numerical and exact solutions we proceed as follows. Let $\rho \in \mathcal{P}_2(\mathbb{R})$ be a continuum gradient flow solution at the final time T. Also write $\boldsymbol{x} \in \mathbb{R}^N_w$, the approximation of the p-approximated discrete gradient flow solution obtained with the JKO scheme (3.7) at the final time step M with associated empirical measure $\mu := \mu^M$. Then we define the *quadratic Wasserstein error* by

$$e_{d_2} = d_2(\rho, \mu).$$

To compute this, one can use the one-dimensional pseudo-inverse definition of the quadratic Wasserstein distance. Let us write F and G the cumulative distributions of ρ and μ, respectively, and Φ and Ψ their respective pseudo-inverses. Then

$$e_{d_2} = \left(\int_0^1 \left(\Phi(\varepsilon) - \Psi(\varepsilon)\right)^2 \mathrm{d}\varepsilon\right)^{\frac{1}{2}}.$$

We have, for all $i \in \{1, \dots, N\}$,

$$\Psi(\varepsilon) = x_i \quad \text{if} \quad \varepsilon \in [\Omega_{i-1}, \Omega_i),$$

with $\Omega_i := \sum_{j=1}^i w_j$ and the convention $\Omega_0 = 0$. Therefore

$$e_{d_2} = \left(\sum_{i=1}^N \int_{\Omega_{i-1}}^{\Omega_i} \left(x_i - \Phi(\varepsilon)\right)^2 \mathrm{d}\varepsilon\right)^{\frac{1}{2}}. \tag{3.9}$$

The error can thus be easily determined if the inverse of the cumulative distribution Φ of the exact solution at the final time T is known.

3.3.2. *Error with respect to a discrete steady state.* If we know that the considered gradient flow has a steady state and we are interested in the stabilisation behaviour of the scheme (3.7) we proceed as follows. Let $\boldsymbol{x^*} := (x_1^*, \dots, x_N^*) \in \mathbb{R}_w^N$ be a discrete steady state (obtained by running a simulation for a "large" time) with associated empirical measure μ^*, and let $\boldsymbol{x} \in \mathbb{R}_w^N$ be the approximation of the p-approximated discrete gradient flow solution obtained with the JKO scheme (3.7) at some final time step M with empirical measure $\mu := \mu^M$. Then we define the error by

$$e_{d_2}^* = d_2(\mu^*, \mu) = \left(\sum_{i=1}^N w_i (x_i^* - x_i)^2 \right)^{\frac{1}{2}}, \tag{3.10}$$

which is nothing but the *Euclidean error* on the particle space $\mathbb{R}_w^N$, see (2.6).

3.4. **Numerical validation: Diffusions.** Before giving any simulation results, let us summarise practical implementation aspects of the scheme.

- As already justified in Section 3.1, all the following simulations were obtained by implementing the explicit scheme (3.7).
- The parameter p needed for the iteration of such scheme was chosen to be 10 for every simulation. This choice is only justified by the fact that it gave sensible results. Note that taking p "large" is a bad idea. Indeed, if p increases, then the p-approximated discrete gradient flow "approaches" the discrete one, for which the energy $\widetilde{E}_N$ is not differentiable whenever two neighbouring inter-particle distances are equal. This may cause numerical instabilities as we would expect the inter-particle distances at the centre of a symmetric profile (like a Gaussian) to be indeed equal; this is what we observed for large p under the form of particle oscillations.
- All one-dimensional simulations were initialised in the way described in Section 3.2.2, up to a slight modification for end particles in Section 4.3. In each case, the initial continuum profile and the interval I_{init} are explicitly given.
- The choice of the time-step size is explicitly given for each simulation.
- All the solution profiles given in the figures for the one-dimensional simulations were drawn by linking linearly the centres of every constant piece of the function ρ_N defined in (2.3). Note that this is only one way of representating the discrete solution and other choices are also possible.

In this section we validate our scheme (3.7) by showing test simulations run on diffusion equations: the heat and porous medium equations.

3.4.1. *The heat equation.* The (linear) heat equation $\rho_t = \Delta\rho$ is a continuum gradient flow (1.2) for $H(\rho) = \rho \log \rho$ and $V = W = 0$. The evolution of the solution starting from the initial data

$$\rho_0^{\text{heat}}(x) = \frac{1}{\sqrt{4\pi t_0}} e^{-\frac{x^2}{4t_0}} \quad \text{with } t_0 = 0.25 \tag{3.11}$$

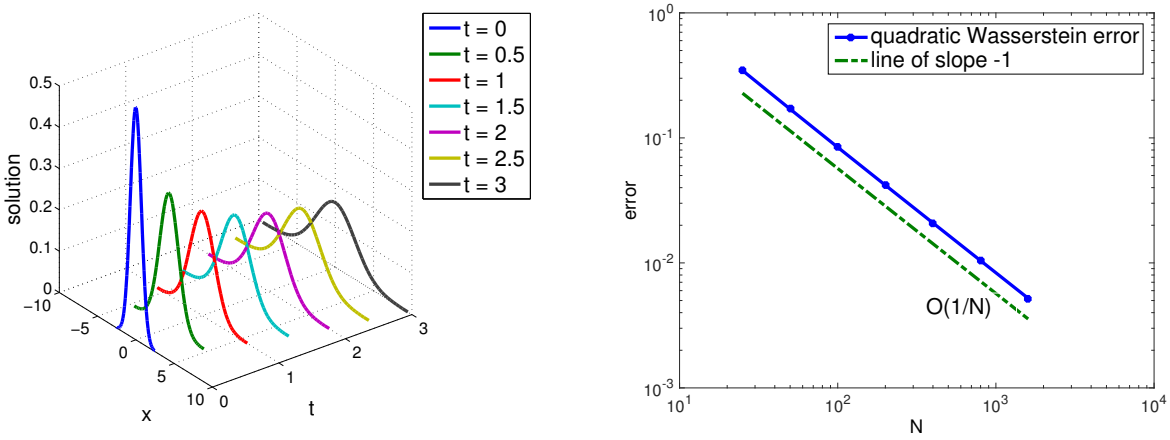
is shown in Figure 1a. The exact solution is given, for all $t > 0$, by

$$\rho(t,x) = \frac{1}{\sqrt{4\pi(t+t_0)}} e^{-\frac{x^2}{4(t+t_0)}} \quad \text{for all } x \in \mathbb{R},$$

whose cumulative distribution F and its pseudo-inverse Φ are

$$\begin{cases} F(t,x) = \dfrac{1}{2}\left(1 + \operatorname{erf}\left(\dfrac{x}{\sqrt{4(t+t_0)}}\right)\right) & \text{for all } x \in \mathbb{R}, \\ \Phi(t,\varepsilon) = \sqrt{4(t+t_0)}\,\operatorname{erf}^{-1}(2\varepsilon - 1) & \text{for all } \varepsilon \in [0,1), \end{cases}$$

where erf is the error function. The error at the final time T can then be found using (3.9) and some quadrature form to approximate the integrals therein, see Figure 1b. In Figure 1a we chose $I_{\text{init}} = [-2.5, 2.5]$, whereas in Figure 1b we chose $I_{\text{init}} = [-4, 4]$, see Section 3.2.2.



(a) Evolution for $N = 50$ with $\Delta t = 10^{-5}$.

(b) Error with N at $T = 3$ with $\Delta t = 5 \cdot 10^{-7}$.

FIGURE 1. The heat equation.

From Figure 1b we can see that the quadratic Wasserstein error with respect to the exact solution ρ evolves linearly with the number of particles on a log-log plot. From this plot, it looks fair to say that the error of our scheme (3.7) is $e_{d_2} = \mathcal{O}\left(\frac{1}{N}\right)$.

3.4.2. *The porous medium equation.* The porous medium equation $\rho_t = \Delta \rho^m$ is a continuum gradient flow (1.2) for $H(\rho) = \frac{\rho^m}{m-1}, m > 1$, and $V = W = 0$. The evolution of the solution starting from the initial data

$$\rho_0^{\text{por}}(x) = \frac{1}{t_0^\alpha} \psi\left(\frac{x}{t_0^\alpha}\right) \quad \text{with } t_0 = 0.25 \tag{3.12}$$

is shown in Figure 2a. Here $\alpha = \frac{1}{m+1}$, $\psi \colon \xi \mapsto (K - \kappa \xi^2)_+^{1/(m-1)}$, where the subscript $_+$ stands for the positive part, $\kappa = \frac{m-1}{2m(m+1)}$ and K determined by the conservation of mass. Indeed, since the total conserved mass is one, then the constant K can be expressed as

$$K = \left[\Gamma\left(\tfrac{1}{m-1} + \tfrac{3}{2}\right) \sqrt{\kappa} \Big/ \left(\Gamma\left(\tfrac{m}{m-1}\right) \Gamma\left(\tfrac{1}{2}\right)\right)\right]^{\frac{2(m-1)}{m+1}},$$

where Γ is the Gamma-function. Then one can verify that, for all $t > 0$,

$$\rho(t, x) = \frac{1}{(t + t_0)^\alpha} \psi\left(\frac{x}{(t + t_0)^\alpha}\right) \quad \text{for all } x \in \mathbb{R},$$

is a solution, see [37, Section 4.4]. For $x \in \left(0, (t + t_0)^\alpha \sqrt{\frac{K}{\kappa}}\right]$, the cumulative distribution is

$$F(t, x) = \frac{1}{2} + \int_0^x \frac{1}{(t + t_0)^\alpha} \left(K - \frac{\kappa x^2}{(t + t_0)^{2\alpha}}\right)_+^{\frac{1}{m-1}} \mathrm{d}x = \frac{1}{2} + \frac{1}{2} I\left(\frac{\kappa}{K} \frac{x^2}{(t + t_0)^{2\alpha}}; \frac{1}{2}, \frac{m}{m-1}\right),$$

where $I(x; a, b) := \frac{B(x;a,b)}{B(1;a,b)}$ and B is the incomplete Beta-function, that is,

$$B(x; a, b) = \int_0^x z^{a-1}(1 - z)^{b-1} \, \mathrm{d}z \quad \text{for all } x \geq 0 \text{ and } a, b > 0.$$

Similarly, for $x \in \left(-(t+t_0)^\alpha \sqrt{\frac{K}{\kappa}}, 0\right]$,

$$F(t,x) = \frac{1}{2} - \frac{1}{2} I\left(\frac{\kappa}{K}\frac{x^2}{(t+t_0)^{2\alpha}}; \frac{1}{2}, \frac{m}{m-1}\right).$$

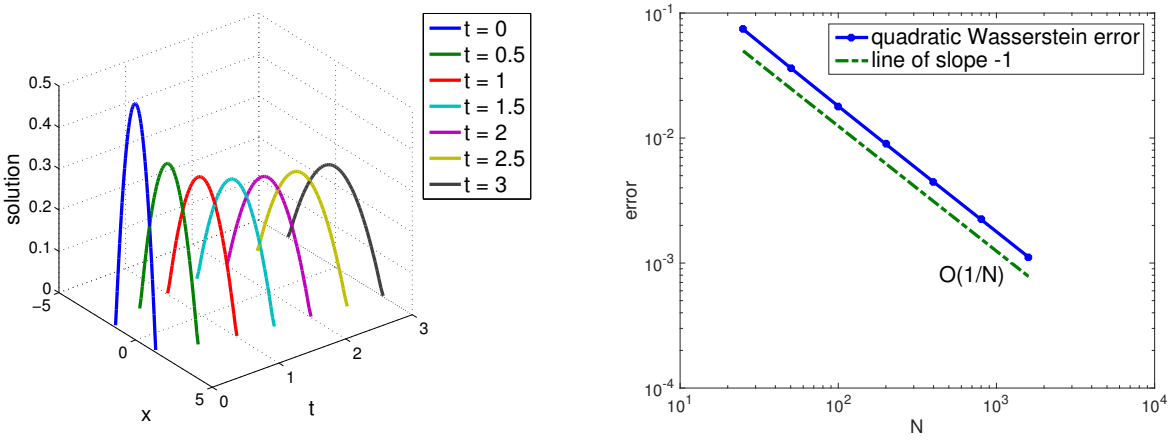
Also, noticing that

$$K^{\frac{1}{m-1}+\frac{1}{2}}\kappa^{-\frac{1}{2}} = \left(B\left(1; \tfrac{1}{2}, \tfrac{m}{m-1}\right)\right)^{-1},$$

we get that the pseudo-inverse of F is

$$\Phi(t,\varepsilon) = \begin{cases} -(t+t_0)^\alpha \sqrt{\frac{K}{\kappa}} \sqrt{I^{-1}\left(1-2\varepsilon; \frac{1}{2}, \frac{m}{m-1}\right)} & \text{if } 0 \le \varepsilon < \frac{1}{2}, \\ (t+t_0)^\alpha \sqrt{\frac{K}{\kappa}} \sqrt{I^{-1}\left(2\varepsilon-1; \frac{1}{2}, \frac{m}{m-1}\right)} & \text{if } \frac{1}{2} \le \varepsilon < 1. \end{cases}$$

The error can then be found using (3.9) and approximating the integrals therein, see Figure 2b. In Figure 2 we chose $I_{\text{init}} = [-k_0, k_0]$, where $k_0 := t_0^\alpha \sqrt{K/\kappa}$, see Section 3.2.2.



(a) Evolution for $N = 50$ with $\Delta t = \frac{0.1}{N^2}$.

(b) Error with N at $T = 3$ with $\Delta t = 10^{-7}$.

FIGURE 2. The porous medium equation with $m = 2$.

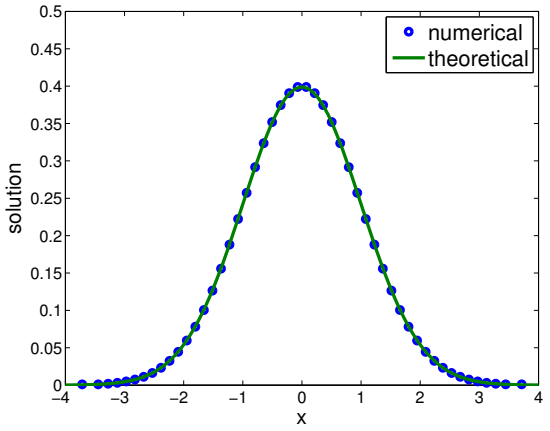
From the plot in Figure 2b, as for the one in Figure 1b for the heat equation, it looks fair to say again that the error of our scheme (3.7) is $e_{d_2} = \mathcal{O}\left(\frac{1}{N}\right)$.

3.4.3. *The linear Fokker-Planck equation.* Let us consider the heat equation with quadratic confinement potential, i.e., the heat equation with $V(x) = \frac{x^2}{2}$. In this case, regardless of the initial condition, there is a steady state:

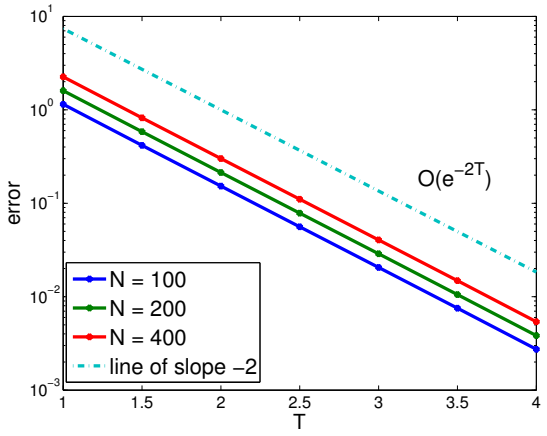
$$\rho_\infty(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{for all } x \in \mathbb{R}.$$

Figure 3b was obtained from the continuum initial profile ρ_0^{heat}, see (3.11), with $I_{\text{init}} = [-2.5, 2.5]$ for Figure 3a and $I_{\text{init}} = [-4, 4]$ for Figure 3b.

In Figure 3b, we can see that the stabilisation of the scheme towards the discrete steady state (which we arbitrarily define as being the discrete solution obtained at $T = 6$) is linear on a semi-log plot with a slope very close to -2, which seems not to depend on the number of particles N. We can therefore write the error as $e_{d_2}^* = \mathcal{O}\left(e^{-2T}\right)$, see (3.10).



(a) Comparison at $T = 4$ for $N = 50$.



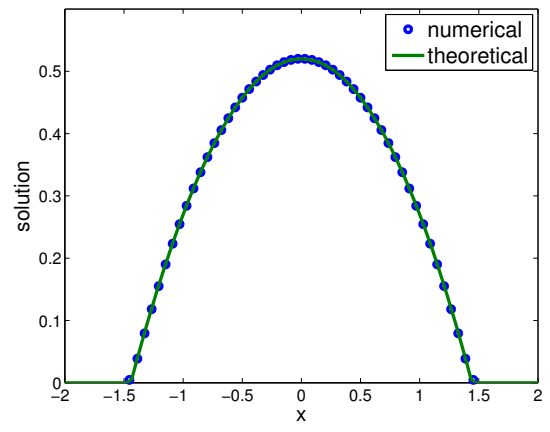
(b) Error with final time T for different N's with respect to the discrete steady state.

FIGURE 3. The linear Fokker-Planck equation with $\Delta t = 10^{-5}$ – Stabilisation in time of the scheme (rate of convergence to the discrete steady state).

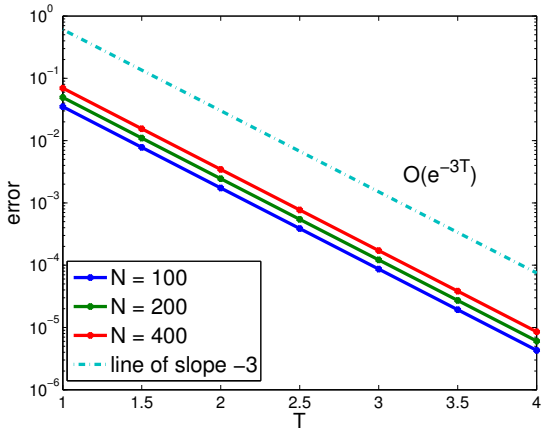
3.4.4. *The nonlinear Fokker-Planck equation.* Let us consider the porous medium equation with quadratic confinement potential, i.e., the porous medium equation with $V(x) = \frac{x^2}{2}$. In this case, regardless of the initial profile and since the total mass is one, the steady state is

$$\rho_\infty(x) := A(R^2 - x^2)_+^{\frac{1}{m-1}} \quad \text{for all } x \in \mathbb{R},$$

where $A = (\frac{m-1}{2m})^{1/(m-1)}$ and $R = (AB(1; \frac{1}{2}, \frac{m}{m-1}))^{(1-m)/(m+1)}$. Figure 4b was obtained from the continuum initial profile ρ_0^{por}, see (3.12), with $I_{\text{init}} = [-k_0, k_0]$, where again $k_0 := t_0^\alpha \sqrt{K/\kappa}$.



(a) Comparison at $T = 4$ for $N = 50$.



(b) Error with the final time T for different N's with respect to the discrete steady state.

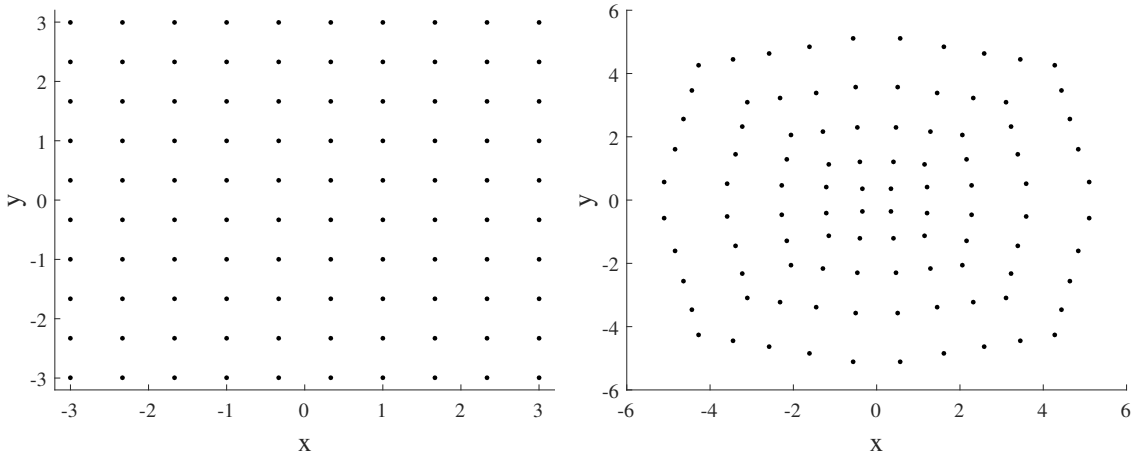
FIGURE 4. The nonlinear Fokker-Planck equation with $m = 2$ and $\Delta t = \frac{0.1}{N^2}$ – Stabilisation in time of the scheme (rate of convergence to the discrete steady state).

As already noted from Figure 3b, for the heat equation the numerical error can be written $e_{d_2}^* = \mathcal{O}\left(e^{-2T}\right)$, as it is actually expected from the theory; for the porous medium equation the

theory says that we should expect $e^*_{d_2} = \mathcal{O}\left(e^{-(m+1)T}\right)$, see [14] and references therein, which is nicely confirmed numerically by Figure 4b with $m = 2$.

3.4.5. *A two-dimensional test: the heat equation.* It is straightforward to generalise the scheme (3.7), derived in Section 3.1, to higher dimensions whenever the expression (3.2) for the Wasserstein distance can be used, which is the case when the approximation μ^{n+1} at time step $n+1$ is sufficiently close to μ^n, i.e., when the time step Δt is small enough. Indeed, let us emphasise that the Wasserstein distance is computed between empirical measures on points, and not between their approximations on non-overlapping balls (which are only used to define the diffusion part of the regularised discrete energy functional, see (2.2)). When the time step is small enough, the Wasserstein distance approximation (3.2) is exact, since then the splitting of mass between empirical measures possibly happening in higher-dimensional optimal transport actually does not occur.

We test our scheme for the heat equation in two dimensions. The initial continuum density is $\rho_0(x) = \frac{1}{4\pi t_0} e^{-|x|^2/(4t_0)}$ with $t_0 = 0.125$, and the particle positions at $T = 1$ are shown in Figure 5b. The data were initialised by fixing the particles on a regular grid as in Figure 5a, with weights being the integrals of the continuum density ρ_0 on the Voronoi cells generated from the particles.



(a) Initial time.

(b) Final time $T = 1$.

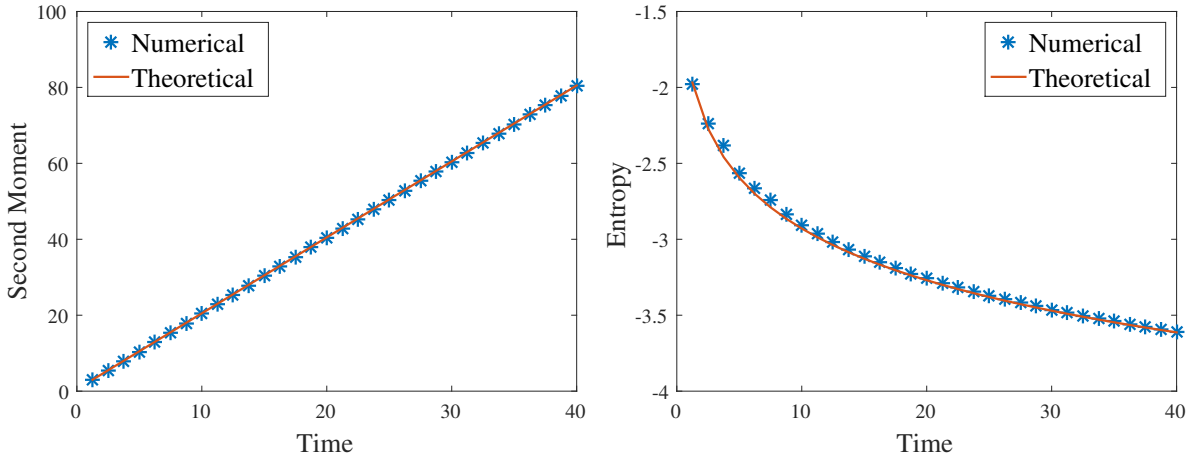
FIGURE 5. The particles' positions for the two-dimensional heat equation for $N = 100$ with $\Delta t = 10^{-4}$.

The averaged quantities along the time evolution, like the second moments $\sum_{i=1}^N w_i |x_i|^2$ and the entropy $\sum_{i=1}^N w_i \log \frac{w_i}{|B_i|}$, seem to be very accurate, as shown in Figure 6, and this accuracy does not seem to degenerate with time. However, representing the numerical solution in this two-dimensional test is a delicate issue since gaps between discretisation balls are significant; this is an issue in itself which we do not deal with here.

4. Aggregation-Diffusion Equations

4.1. **The modified Keller-Segel equation.** We start by considering a modified one-dimensional Keller-Segel equation, that is the continuum gradient flow (1.2) with $H(\rho) = \rho \log \rho$ and $W(x) = 2\chi \log |x|$ (and $W(0) := 0$), where $\chi > 0$ is a parameter quantifying the attraction (see [11, 8] for well-posedness and qualitative properties, and [10] for the approximation of this equation by a different particle method).

This model shows a critical behaviour depending only on the chemosensitivity strength χ as the classical Keller-Segel model in two dimensions, that is, there is a dichotomy of blow-up in finite



(a) Evolution of the second moment, $\int |x|^2 \rho$.

(b) Evolution of the entropy, $\int \rho \log \rho$.

FIGURE 6. Accuracy for the two-dimensional heat equation with $\Delta t = 10^{-4}$.

time or global existence which is only determined by χ being larger or less than 1, see [11, 8]. In case $\chi < 1$ solutions spread in time behaving like self-similar solutions. To get the leading order profile given by the self-similar solution, a time-space scaling is done for $\chi < 1$, which is equivalent to impose a quadratic confinement potential on particles, see [8]. The long-time behaviour in this subcritical case is given by the profile of the self-similar solution.

4.1.1. *Theoretical properties.* We show that our particle approximation keeps approximately the criticality of the original Keller-Segel model at the discrete level. We show that there exist two positive constants χ_1 and $\chi_2(N)$ such that the following holds: if an appropriate confinement potential V is considered, the discrete and the p-approximated discrete gradient flows (2.7) and (2.10) for the modified Keller-Segel equation have steady states for $\chi < \chi_1$; while if $V = 0$, the p-approximated discrete gradient flow shows finite-time blow-up for $\chi > \chi_2(N)$. Quite surprisingly, χ_1 happens to be exactly the critical parameter at the continuum level, i.e., $\chi_1 = 1$, and $\chi_2(N)$ tends to 1 as $N \to \infty$ and does not depend on p (but only on N and the chosen weights).

By the term "blow-up" at the discrete level we mean the event of two, or more, particles colliding. Also note that in the following the term "maximal time of existence" indicates either the first time when two or more particles of a solution collide, i.e., the first blow-up time, or the first time when the norm of a solution equals $+\infty$.

First, let us prove that the discrete and p-approximated discrete confined Keller-Segel equations show no collisions of particles if $\chi < \chi_1$.

Proposition 4.1. *Consider the discrete gradient flow corresponding to the confined Keller-Segel equation with V coercive and such that the function $x \mapsto \inf_{y \in \mathbb{R}}(w_1 V(y) + w_N V(x+y)) - \log|x|$ is coercive. Suppose there exists a solution $\boldsymbol{x}$ to such gradient flow, emanating from an initial condition $\boldsymbol{x^0} \in \mathbb{R}^N_w$, up to some maximal time of existence, say $T^* > 0$. If*

$$\chi < \chi_1 := 1,$$

then no particles of $\boldsymbol{x}$ can collide in $[0, T^*)$; furthermore, the minimal inter-particle distance is uniformly bounded from below in time by a positive constant.

Proof. The energy $\widetilde{E}_N$ is a Lyapunov functional, i.e.,

$$\widetilde{E}_N(\boldsymbol{x}(t)) \leq \widetilde{E}_N(\boldsymbol{x^0}) := E_0 < +\infty \quad \text{for all } t \in [0, T^*), \tag{4.1}$$

see [4, Theorem 1 of Section 3.4]. Fix $t \in [0, T^*)$ and get, by (2.2),

$$\widetilde{E}_N(\boldsymbol{x}(t)) = \sum_{i=1}^N w_i \log w_i - \sum_{i=1}^N w_i \log r_i(t) + \sum_{i=1}^N w_i V(x_i(t)) + \chi \sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j \log|x_i(t) - x_j(t)|, \tag{4.2}$$

where $r_i(t) = \min(\Delta x_i(t), \Delta x_{i+1}(t))$. Writing $\log_- x := \log x$ if $0 < x < 1$ and $\log_- x := 0$ if $x \geq 1$, and using that log is increasing,

$$\begin{aligned}\sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j \log|x_i(t) - x_j(t)| &\geq \sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j \log \min_{\substack{k\in\{1,\dots,N\} \\ k\neq i}} |x_i(t) - x_k(t)| \\ &\geq \sum_{i=1}^N w_i(1-w_i) \log_- r_i(t) \geq \sum_{i=1}^N w_i \log_- r_i(t).\end{aligned}$$

Hence, writing $\log_+ x := \log x$ if $x \geq 1$ and $\log_+ x := 0$ if $0 \leq x < 1$, and by (4.2),

$$\begin{aligned}\widetilde{E}_N(\boldsymbol{x}(t)) &\geq \sum_{i=1}^N w_i \log w_i + \sum_{i=1}^N w_i V(x_i(t)) + (\chi - 1) \sum_{i=1}^N w_i \log_- r_i(t) - \sum_{i=1}^N w_i \log_+ r_i(t) \\ &\geq \sum_{i=1}^N w_i \log w_i + \sum_{i=2}^{N-1} w_i V(x_i(t)) + w_1 V(x_1(t)) + w_N V(x_N(t)) \\ &\quad + (\chi - 1) \sum_{i=1}^N w_i \log_- r_i(t) - \log_+(x_N(t) - x_1(t)),\end{aligned}$$

using that $-\log_+ r_i(t) \geq -\log_+(x_N(t) - x_1(t))$ for all $i \in \{1, \dots, N\}$. From the assumptions on V we know that V is bounded from below and also that $w_1 V(x_1(t)) + w_N V(x_N(t)) - \log_+(x_N(t) - x_1(t))$ is bounded from below uniformly in time. Therefore there exists a constant $C \in \mathbb{R}$, independent of t, such that

$$E_0 \geq \widetilde{E}_N(\boldsymbol{x}(t)) \geq (\chi - 1) \sum_{i=1}^N w_i \log_- r_i(t) + C, \tag{4.3}$$

using (4.1). We see that if $\chi < \chi_1 = 1$, the minimal inter-particle distances $r_i(t)$ cannot get arbitrarily small, or the energy $\widetilde{E}_N(\boldsymbol{x}(t))$ gets larger than its initial value E_0, which violates the fact that the system is a gradient flow. Hence the result. ■

Remark 4.2. The proof above is given only for the discrete case; however, note that it can be easily adapted to the p-approximated one, if $p \geq 1$, by Proposition 2.7(5) with $s = 2$, and without the need to change the constant χ_1.

We can now show the global existence in time and the existence of steady states for the discrete and p-approximated discrete confined Keller-Segel equations.

Proposition 4.3. *Consider the discrete gradient flow corresponding to the confined Keller-Segel equation with V satisfying the same hypotheses as in Proposition 4.1. If $\chi < \chi_1$, then any solution to this gradient flow, if it exists, exists globally in time and the gradient flow has a steady state.*

Proof. Let $\chi < \chi_1 = 1$. Suppose there exists such a solution $\boldsymbol{x}$, emanating from an initial condition $\boldsymbol{x^0} \in \mathbb{R}^N_w$, defined on $[0, T^*)$, see Proposition 4.1, and assume that $x_N(t) - x_1(t) \to +\infty$ as $t \to T^*$. For all $t \in [0, T^*)$, the proof of Proposition 4.1 implies that there exists a t-independent constant $C_1 \in \mathbb{R}$ such that

$$E_0 \geq \widetilde{E}_N(\boldsymbol{x}(t)) \geq (\chi - 1) \sum_{i=1}^{N} w_i \log_- r_i(t) + f(x_N(t) - x_1(t)) + C_1 \geq f(x_N(t) - x_1(t)) + C_1,$$

since $(\chi - 1) \sum_{i=1}^N w_i \log_- r_i(t) \geq 0$, and where $f(x_N(t) - x_1(t)) := \inf_{y \in \mathbb{R}} (w_1 V(y) + w_N V(x_N(t) - x_1(t) + y)) - \log_+(x_N(t) - x_1(t))$. By the growth assumption at infinity on V, we have $f(x_N(t) - x_1(t)) \to +\infty$ as $t \to T^*$, which implies that the inequality above is violated for a time t close enough to T^*. Therefore $x_N(t) - x_1(t)$ cannot diverge as $t \to T^*$, and thus there exists $C_2 \in \mathbb{R}$, uniform in t, such that $\log_+(x_N(t) - x_1(t)) \leq C_2$ for all $t \in [0, T^*)$. Therefore,

$$E_0 \geq \widetilde{E}_N(\boldsymbol{x}(t)) \geq w_1 V(x_1(t)) + w_N V(x_N(t)) - C_2 + C_1,$$

which, by coercivity of V shows that $x_1(t)$ and $x_N(t)$ cannot diverge. Thus, there exists some constant $\ell > 0$, independent of time, such that $\{x_1(t), \dots, x_N(t)\} \subset B(0, \ell)$ for all $t \in [0, T^*)$. This, together with the no-collision result in Proposition 4.1, shows that the maximal time of existence of the discrete gradient flow solution is $T^* = \infty$.

Finally, the functional $\widetilde{E}_N$ is lower semi-continous and bounded from below on any sublevel set of $\widetilde{E}_N$ which are compact due to (4.3). Indeed, the previous argument shows that $x_i \in B(0, \ell)$ for all $\boldsymbol{x} \in \mathbb{R}^N_w$ such that $E_0 \geq \widetilde{E}_N(\boldsymbol{x})$. Moreover, the same argument leading to (4.3) implies that

$$E_0 \geq \widetilde{E}_N(\boldsymbol{x}) \geq C,$$

for all $\boldsymbol{x} \in \mathbb{R}^N_w$ such that $E_0 \geq \widetilde{E}_N(\boldsymbol{x})$, since $\chi < 1$. Therefore, by a direct method of calculus of variations, we get that $\widetilde{E}_N$ has a global minimiser, which ends the proof. ∎

Remark 4.4. Similarly to Proposition 4.3 the proof above is given only for the discrete case, but is easily adaptable to the p-approximated one if $p \geq 1$, by Proposition 2.7(5) with $s = 2$.

Remark 4.5. The assumptions on V of Propositions 4.1 and 4.3 are in particular satisfied by $V(x) = \frac{x^k}{k}$ for any $k \geq 1$. Propositions 4.1 and 4.3 are also true for any W bounded from below and satisfying Hypothesis 1, with no required constraint on χ; in particular this is the case for the linear Fokker-Planck equation, that is with $V(x) = \frac{x^2}{2}$ and $W = 0$.

Let us now turn to the supercritical case. In the unconfined continuum modified Keller-Segel equation, it is known that solutions blow up in finite time if $\chi > 1$. The proof of non-existence of global-in-time solutions is obtained by computing the evolution of the second moment $M_2(t)$ of solutions $\rho(t)$ at any time $t > 0$. Then, a formal computation leads to

$$\frac{\mathrm{d}M_2}{\mathrm{d}t}(t) = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}} x^2 \, \mathrm{d}\rho(t, x) = 2(1 - \chi). \tag{4.4}$$

Therefore the evolution of the second moment is linear in time with slope $2(1 - \chi)$. This slope is negative if $\chi > 1$, which implies that $M_2(t)$ becomes zero in finite time leading to concentration of mass in finite time and contradiction with the assumption of global existence. We want here to show that our p-approximated discrete gradient flow (2.10) preserves this finite-time blow-up property for some numerical critical parameter $\chi_2(N)$, at least when all particles have same weight. Recall that, at the discrete level, we define blow-up as being the event of two particles colliding.

Proposition 4.6. *Let $p > 0$ and consider the p-approximated discrete unconfined Keller-Segel gradient flow with $w_i = 1/N$ for all $i \in \{1, \dots, N\}$, on the whole time line $[0, \infty)$. All solutions blow up in finite time if χ is greater than*

$$\chi_2(N) := 1 + \frac{1}{N-1}.$$

Proof. Suppose that there exists $\boldsymbol{x}$, a p-approximated discrete Keller-Segel gradient flow solution emanating from an initial condition $\boldsymbol{x^0} \in \mathbb{R}^N_w$, defined on some maximal interval of existence $[0, T^*)$. Let us compute the evolution of the second moment of μ_N, the empirical measure associated to $\boldsymbol{x}$, at any $t \in [0, T^*)$.

$$\frac{\mathrm{d}M_2}{\mathrm{d}t}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} x^2 \,\mathrm{d}\mu_N(t,x) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{N}\sum_{i=1}^N x_i^2(t) = \frac{2}{N}\sum_{i=1}^N x_i(t)\frac{\mathrm{d}x_i}{\mathrm{d}t}(t). \tag{4.5}$$

In the following we drop the dependencies on time for the sake of simplicity.

Suppose $T^* = \infty$. We want to find a contradiction if $\chi > \chi_2(N)$ by computing explicitly the evolution of the second moment in (4.5). Write $\Delta_i^j := 1 + (\Delta x_j/\Delta x_i)^p$ for all $i, j \in \{1, \dots, N\}$ with $i \neq j$, recalling the convention $\Delta x_1 = \Delta x_{N+1} = +\infty$ and setting $\Delta_0^1 = \Delta_{N+2}^{N+1} = +\infty$. By (2.10),

$$\begin{aligned}\sum_{i=1}^N x_i \frac{\mathrm{d}x_i}{\mathrm{d}t} = &\underbrace{\sum_{i=1}^N \left(\left(\frac{x_i/\Delta x_i}{\Delta_{i-1}^i} + \frac{x_i/\Delta x_i}{\Delta_{i+1}^i}\right) - \left(\frac{x_i/\Delta x_{i+1}}{\Delta_i^{i+1}} + \frac{x_i/\Delta x_{i+1}}{\Delta_{i+2}^{i+1}}\right)\right)}_{=:A_1} \\ &\underbrace{- \frac{2\chi}{N}\sum_{i=1}^N x_i \sum_{\substack{j=1 \\ j\neq i}}^N \bar{W}'(x_i - x_j)}_{=:A_2},\end{aligned}$$

where $\bar{W} := \log|\cdot|$. First, compute A_1 by appropriately rearranging the sum terms,

$$\begin{aligned}A_1 = &\sum_{i=3}^{N-2}\left(\left(\frac{x_i/\Delta x_i}{\Delta_{i-1}^i} + \frac{x_i/\Delta x_i}{\Delta_{i+1}^i}\right) - \left(\frac{x_i/\Delta x_{i+1}}{\Delta_i^{i+1}} + \frac{x_i/\Delta x_{i+1}}{\Delta_{i+2}^{i+1}}\right)\right) + \left(1 + \frac{1}{\Delta_3^2}\right) \\ &+ \left(1 + \frac{1}{\Delta_{N-1}^N}\right) - \left(\frac{x_2/\Delta x_3}{\Delta_2^3} + \frac{x_2/\Delta x_3}{\Delta_4^3}\right) + \left(\frac{x_{N-1}/\Delta x_{N-1}}{\Delta_N^{N-1}} + \frac{x_{N-1}/\Delta x_{N-1}}{\Delta_{N-2}^{N-1}}\right).\end{aligned}$$

Also,

$$\begin{aligned}&\sum_{i=3}^{N-2}\left(\left(\frac{x_i/\Delta x_i}{\Delta_{i-1}^i} + \frac{x_i/\Delta x_i}{\Delta_{i+1}^i}\right) - \left(\frac{x_i/\Delta x_{i+1}}{\Delta_i^{i+1}} + \frac{x_i/\Delta x_{i+1}}{\Delta_{i+2}^{i+1}}\right)\right) \\ &\quad = \left(\frac{x_3/\Delta x_3}{\Delta_4^3} + \frac{x_3/\Delta x_3}{\Delta_2^3}\right) - \left(\frac{x_{N-1}/\Delta x_{N-1}}{\Delta_{N-2}^{N-1}} + \frac{x_{N-1}/\Delta x_{N-1}}{\Delta_N^{N-1}}\right) + \sum_{i=3}^{N-2}\left(\frac{1}{\Delta_i^{i+1}} + \frac{1}{\Delta_{i+2}^{i+1}}\right).\end{aligned}$$

Hence, by combining the last two computations, we get

$$\begin{aligned}A_1 &= \sum_{i=3}^{N-2}\left(\frac{1}{\Delta_i^{i+1}} + \frac{1}{\Delta_{i+2}^{i+1}}\right) + \left(\frac{1}{\Delta_4^3} + \frac{1}{\Delta_2^3}\right) + \left(1 + \frac{1}{\Delta_3^2}\right) + \left(1 + \frac{1}{\Delta_{N-1}^N}\right) \\ &= \sum_{i=2}^{N-1}\left(\frac{1}{\Delta_i^{i+1}} + \frac{1}{\Delta_{i+1}^i}\right) + 2 = \sum_{i=2}^{N-1} 1 + 2 = N.\end{aligned}$$

Then, compute A_2 by using the anti-symmetry of $\bar{W}'(x) = \frac{1}{x}$,

$$\begin{aligned} A_2 &= (x_1 - x_2)\bar{W}'(x_1 - x_2) + (x_1 - x_3)\bar{W}'(x_1 - x_3) + \cdots + (x_1 - x_N)\bar{W}'(x_1 - x_N) \\ &\quad + (x_2 - x_3)\bar{W}'(x_2 - x_3) + (x_2 - x_4)\bar{W}'(x_2 - x_4) + \cdots + (x_2 - x_N)\bar{W}'(x_2 - x_N) + \dots \\ &\quad + (x_{N-1} - x_N)\bar{W}'(x_{N-1} - x_N) = (N-1) + (N-2) + \cdots + 1 = \textstyle\sum_{i=1}^{N-1}(N-i) = \frac{(N-1)N}{2}. \end{aligned}$$

Therefore, for all $t \in [0, T^*) = [0, \infty)$,

$$\frac{\mathrm{d}M_2}{\mathrm{d}t}(t) = \frac{2}{N}\left(N - \frac{2\chi}{N}\frac{(N-1)N}{2}\right) = 2\left(1 - \chi\left(1 - \frac{1}{N}\right)\right) = 2\left(1 - \frac{1}{N}\right)(\chi_2(N) - \chi). \tag{4.6}$$

Hence the evolution of the second moment is linear with a negative slope, since by assumption $\chi > \chi_2(N)$, which clearly contradicts the fact that the maximal time of existence $T^* = \infty$, and therefore the solution $\boldsymbol{x}$ exists only up to a finite time: $T^* < \infty$. At exactly that time, only two things may happen: either the norm of the solution equals $+\infty$, i.e., $|\boldsymbol{x}|_w = +\infty$, or two or more particles collide. The first option is not plausible since trivially the second moment of an empirical measure is finite at all times. We are thus only left with the collision of particles, that is $\boldsymbol{x}$ has to blow up in finite time. ■

4.1.2. *Simulations.* We give here a few simulations for the modified Keller-Segel equation showing various blow-up characteristics when $V = 0$. As we want to capture the blow-up we used an adaptive time-step size as follows. For every time step $n \in \{0, \dots, M-1\}$, suppose we have a time-step size $\Delta_n t$, and compute the velocity v_i^n of each particle x_i^n, $1 \le i \le N$. Then fix a tolerance $\delta = 0.25$, and define

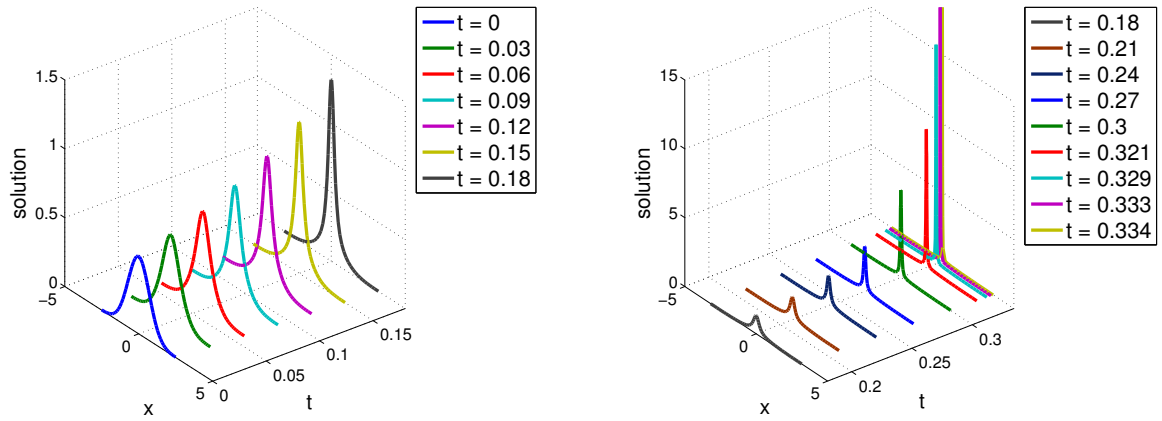
$$\delta_i = \begin{cases} \Delta_n t & \text{if } \Delta_n t \le \frac{\delta \min_p(\Delta x_i^n, \Delta x_{i+1}^n)}{|v_i^n|}, \\ \frac{\delta \min_p(\Delta x_i^n, \Delta x_{i+1}^n)}{|v_i^n|} & \text{otherwise.} \end{cases}$$

Finally, renew $\Delta_n t := \min_{i \in \{1,\dots,N\}} \delta_i$, compute the positions x_i^{n+1} with the new $\Delta_n t$, and start over until $\Delta_n t > 10^{-7}$; when $\Delta_n t \le 10^{-7}$ stop the simulation. We took $\Delta_0 t = 10^{-5}$ as the very initial time-step size. In Figures 7, 8 and 9 the simulations shown stopped due to this adaptive time-step procedure.

Figures 7 and 8 show the results of simulations with initial continuum profile ρ_0^{heat}, with $I_{\text{init}} = [-2.5, 2.5]$. From Figure 7 one can see that the scheme we used captures nicely the formation of the blow-up for a supercritical parameter χ.

Figure 8a shows that the evolution of the second moment is linear for some range of time with a slope that deviates slightly from the theoretical one, as expected from Proposition 4.6 and its proof. Actually, by comparing (4.4) and (4.6), it can be seen that this slope deviation decreases as N increases. For larger times, as the blow-up is approached and the time step refined, the numerical slope deviates even more from the theoretical one.

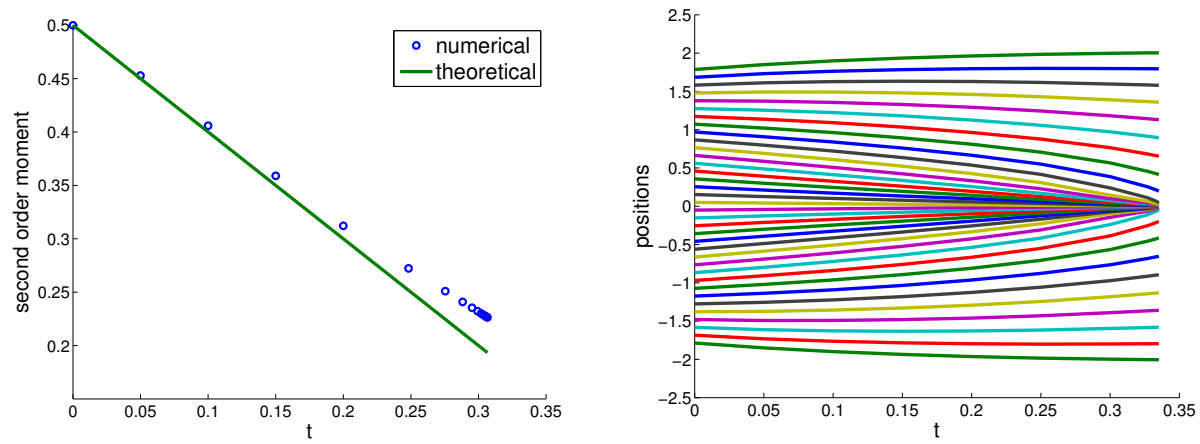
Figure 8b shows the trajectories of the particles up to the first blow-up time, defined numerically as the first time when the distance between two particles gets smaller than some chosen number $d_{\min}$, or equivalently when the adaptive time-step size $\Delta_n t$ gets larger than 10^{-7}, see above. A possible procedure to continue the evolution after the first blow-up time is merging particles into a new heavier one whenever the distance between these particles gets less than a certain threshold, say proportional to $d_{\min}$; the weight of the new particle is then chosen to be the sum of the merged particles, and the position the barycentre of the merged particles. This procedure might give an idea of how the particles behave after the first blow-up, but we found that it is not very accurate since the post-collision trajectories strongly depended on the choice of the threshold, which is arbitrary. We found that the analysis of the post-collision behaviour is very delicate without having clear



(a) Early evolution.

(b) Blow-up formation.

FIGURE 7. The modified Keller-Segel equation with $\chi = 1.5$ for $N = 50$.



(a) Evolution of the second moment for $N = 100$.

(b) Particle trajectories up to first numerical blow-up for $N = 50$ (not all particles are represented).

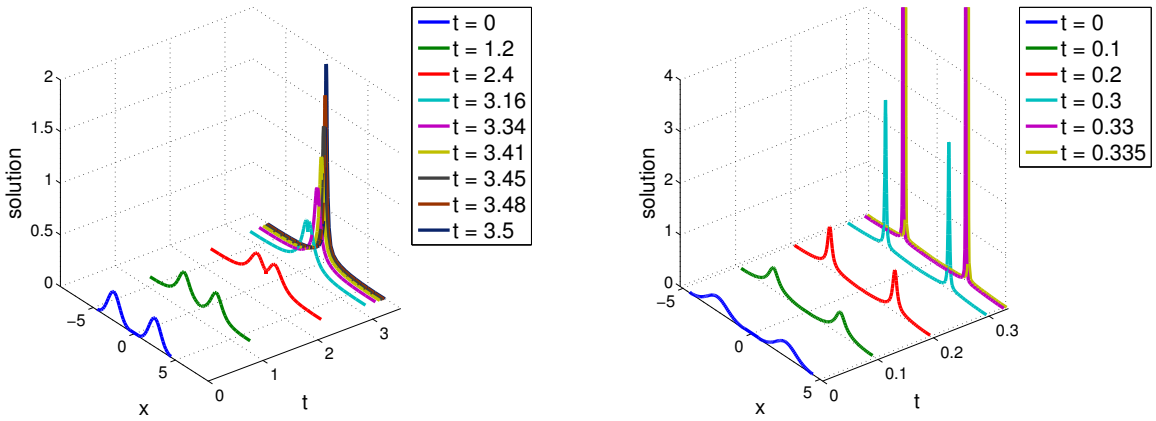
FIGURE 8. The blow-up of the modified Keller-Segel equation with $\chi = 1.5$.

criteria for deciding when to merge particles and how many simultaneously. We thus leave this issue for further analysis and future work.

Figure 9 shows the result with a continuum two-bump initial profile, with $I_{\text{init}} = [-4.5, 4.5]$:

$$\rho_0(x) = \frac{1}{2\sqrt{4\pi t_0}} e^{-\frac{(x+2)^2}{4t_0}} + \frac{1}{2\sqrt{4\pi t_0}} e^{-\frac{(x-2)^2}{4t_0}} \quad \text{with } t_0 = 0.25.$$

Figure 9 shows the possible formation of several Dirac masses, according to how much attraction is involved in the system and to how many bumps are present at the beginning. It seems like the more attraction, the more Dirac masses can form. Note that the two peaks in Figure 9b are actually of same height despite a displaying artifact; also, even if not clear from Figure 9b, the two peaks get slightly closer to each other with time, but then blow up before merging.



(a) Evolution with $\chi = 1.8$ for $N = 50$.

(b) Evolution with $\chi = 3$ for $N = 100$.

FIGURE 9. Blow-up formation for the modified Keller-Segel equation with two initial Gaussian bumps.

4.2. **The modified Keller-Segel equation with nonlinear diffusion.** Let us now consider the modified one-dimensional Keller-Segel equation with nonlinear diffusion, i.e., the continuum gradient flow (1.2) with $H(\rho) = \frac{\rho^m}{m-1}$, $m > 1$, $V = 0$ and $W(x) = 2\chi \log|x|$ (and $W(0) := 0$). The initial continuum profile we used here is ρ_0^{heat}, with $I_{\text{init}} = [-2.5, 2.5]$.

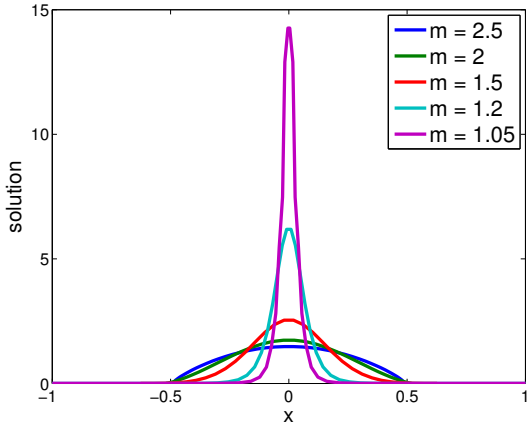
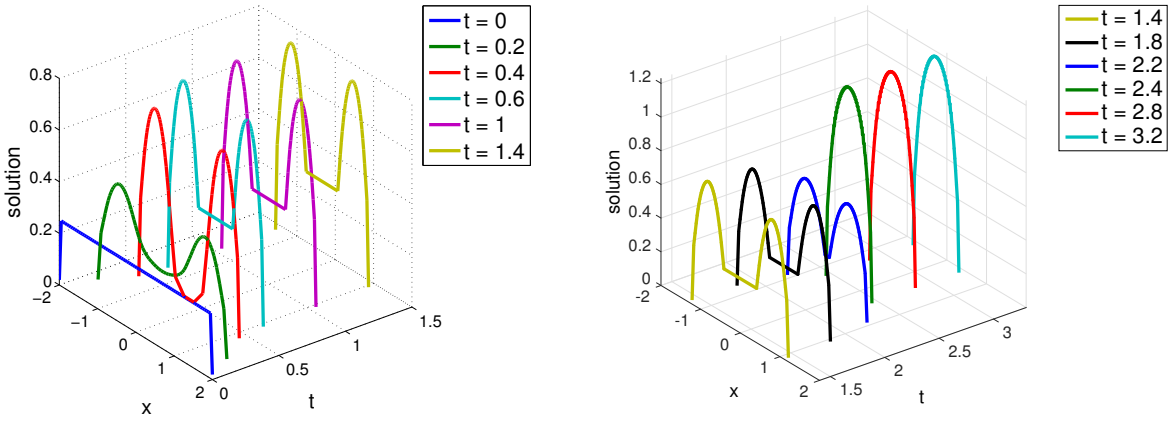


FIGURE 10. The modified nonlinear Keller-Segel equation with $\chi = 1.4$ for different choices of m, for $N = 50$ at $T = 4$ with $\Delta t = 10^{-5}$.

Each curve in Figure 10 is a good approximation of a steady state for the modified Keller-Segel equation with nonlinear diffusion. For each m, the steady state is different; as m tends to 1 the steady state "squeezes" and looks as if it is approaching a Dirac mass, which is the "steady state" of the modified Keller-Segel equation with linear diffusion studied in Section 4.1.

4.3. **A compactly supported potential with nonlinear diffusion.** We consider here the continuum gradient flow when $H(\rho) = \frac{\rho^m}{m-1}$, $m > 1$, $V = 0$ and $W(x) = -c\max(1-|x|, 0) + c$, $c > 0$, is a compactly supported interaction potential. In Figure 11 the considered continuum initial profile

is a uniform distribution on the interval $[-2, 2]$. Here $I_{\text{init}} = [-2, 2]$ and the end particles were set to have weights equal to 0.001.



(a) Formation of metastable state.

(b) Formation of steady state.

FIGURE 11. Compactly supported potential $W(x) = -c\max(1 - |x|, 0) + c$ with nonlinear diffusion with $c = 8$ and $m = 3$, for $N = 80$ with $\Delta t = 10^{-5}$.

Figure 11a shows the formation of a metastable state made of two bumps, while Figure 11b shows how this metastable state breaks into a single-bumped steady state. This behaviour, for exactly this interaction potential (up to a multiplicative constant, c), was already noted in [12, Example 3] when using a finite-volume scheme on gradient flows. The bumps are actually supposed to be disconnected since no particles were found numerically in-between, although this does not seem to be the case in the plots. This is because the particle of each bump which is closest to the origin is not located at the actual boundary of the exact solution's corresponding bump, and so the line connecting the two bumps does not show 0 density but more. Visually, a better separation of the bumps can be obtained by increasing the number of particles in the simulation.

APPENDIX A. CONVERGENCE OF THE p-APPROXIMATED GRADIENT FLOW IN ONE DIMENSION

In this appendix we give the proof of the convergence of the p-approximated gradient flow (2.10) to the discrete one (2.7) in one dimension, see Section 2.3. Before doing so we need to recall some notions and results from monotone operator theory.

If $A\colon \{y \in X \mid Ay \neq \emptyset\} \subset X \to X^*$, where X is a Banach space and X^* its dual, is a maximal monotone operator, then the *graph* of A is given by $\operatorname{graph}(A) = \{(x, Ax) \mid x \in \{y \in X \mid Ay \neq \emptyset\}\}$.

Definition A.1 (Graph convergence)**.** Let X be a reflexive Banach space, $(A_k)_{k\in\mathbb{N}}$ a sequence of maximal monotone operators from X to X^*, and A a maximal monotone operator from X to X^*. We say that $(A_k)_{k\in\mathbb{N}}$ *graph-converges* to A if, for every $(x, y) \in \operatorname{graph}(A)$, there exists a sequence $((x_k, y_k))_{k\in\mathbb{N}}$ with $(x_k, y_k) \in \operatorname{graph}(A_k)$ for all $k \in \mathbb{N}$ such that $x_k \to x$ strongly in X and $y_k \to y$ strongly in X^* as $k \to \infty$.

Remark A.2. One can check that if ϕ is a lower semi-continuous, convex function from a reflexive Banach space to $\mathbb{R}$, then $\partial\phi$ is a maximal monotone operator, see [3, Section 3.7.1].

For the sake of completeness we give the defintion of Γ-convergence.

Definition A.3 (Γ-convergence)**.** Let $(X_k)_{k\in\mathbb{N}}$ be a sequence of metric spaces endowed with a distance d and $(\phi_k)_{k\in\mathbb{N}}$ be a sequence of functionals $\phi_k\colon X_k \to \mathbb{R}$ for all $k \in \mathbb{N}$. We say that $(\phi_k)_{k\in\mathbb{N}}$ Γ-*converges* to ϕ if the following two conditions are met for all $u \in X$:

(i) ("liminf" condition) $\phi(u) \le \liminf_{k\to\infty} \phi_k(u_k)$ for all sequences $(u_k)_{k\in\mathbb{N}}$ with $u_k \in X_k$ for all $k \in \mathbb{N}$ and $d(u_k, u) \to 0$ as $k \to \infty$,

(ii) ("limsup" condition) $\limsup_{k\to\infty} \phi_k(u_k) \le \phi(u)$ for some sequence $(u_k)_{k\in\mathbb{N}}$ with $u_k \in X_k$ for all $k \in \mathbb{N}$ and $d(u_k, u) \to 0$ as $k \to \infty$.

Theorem A.4 connects the notion of Γ-convergence to that of graph-convergence in finite dimension. It is a consequence of [3, Theorem 3.66] and the fact that Γ-convergence is equivalent to Mosco convergence in finite dimension. Indeed, in a general dimensional setting, Mosco convergence means that the "liminf" condition of Γ-convergence holds for the weak topology and the "limsup" condition holds for the strong topology, see [27, Definition 2.2] and [3, Definition 3.17].

Theorem A.4. *Let X be a finite-dimensional Banach space. Let $(\phi_k)_{k\in\mathbb{N}}$ be a sequence of lower semi-continuous, convex functions with $\phi_k\colon X \to \mathbb{R}$ for all $k \in \mathbb{N}$, and $\phi\colon X \to \mathbb{R}$ a lower semi-continuous, convex function. If $(\phi_k)_{k\in\mathbb{N}}$ Γ-converges to ϕ, then $(\partial\phi_k)_{k\in\mathbb{N}}$ graph-converges to $\partial\phi$.*

We can now give the general convergence and regularity result, whose proof can be deduced by [3, Theorem 3.74], [9, Theorem 3.1] and [4, Theorem 1, Section 3.2].

Theorem A.5. *Let X be a Hilbert space. Let $(\phi_k)_{k\in\mathbb{N}}$ be a sequence of lower semi-continuous, convex functions with $\phi_k\colon X \to \mathbb{R}$ for all $k \in \mathbb{N}$ and $\phi\colon X \to \mathbb{R}$ be a lower semi-continuous, convex function. Suppose that $(\partial\phi_k)_{k\in\mathbb{N}}$ graph-converges to $\partial\phi$. Consider the following differential inclusions, for all $k \in \mathbb{N}$.*

$$u_k'(t) \in -\partial\phi_k(u_k(t)), \quad u_k(0) = u_k^0 \quad \text{for almost every } t \in (0,T],$$

and

$$u'(t) \in -\partial\phi(u(t)), \quad u(0) = u_0 \quad \text{for almost every } t \in (0,T],$$

where $u, u_k\colon [0,T] \to X$ are the unknown curves. Unique solutions u and u_k exist and

(1) u_k and u are continuous on $[0,T]$,

(2) u_k' and u' are right-continuous on $[0,T]$.

Moreover, assume that $u_k^0 \to u_0$ strongly and that in this case, $\phi_k(u_k^0) \to \phi(u_0)$ as $k \to \infty$. Then

(3) $u_k \to u$ uniformly on $[0,T]$ as $k \to \infty$,

(4) $\int_0^T t\,|u_k'(t) - u'(t)|^2 \,\mathrm{d}t \to 0$ as $k \to \infty$,

(5) $u_k' \to u'$ strongly in $L^2([0,T], X)$ as $k \to \infty$ (so $u_k'(t) \to u'(t)$ for almost every $t \in [0,T]$),

(6) $\phi_k(u_k) \to \phi(u)$ uniformly on $[0,T]$ as $k \to \infty$.

It is not hard to see that $\widetilde{E}_N^p$, see (2.11), Γ-converges to $\widetilde{E}_N$, see (2.4), as $p \to \infty$. Furthermore, it is easily verified that $\widetilde{E}_N^p(\boldsymbol{x}^{\mathbf{0}}) \to \widetilde{E}_N(\boldsymbol{x}^{\mathbf{0}})$ as $p \to \infty$, where $\boldsymbol{x}^{\mathbf{0}} \in \mathbb{R}_w^N$ is taken here to be the initial condition for both the discrete and p-approximated discrete gradient flows (2.7) and (2.10). Therefore, in order to use Theorems A.4 and A.5 combined, we are only left with checking that $\widetilde{E}_N^p$ and $\widetilde{E}_N$ are lower semi-continuous and convex. The first condition is trivial to verify based on the assumptions on H, V and W, whereas the convexity condition is shown below whenever V and W are assumed to be convex. Actually the following proposition shows the convexity of $\widetilde{E}_N$ only, but also holds for $\widetilde{E}_N^p$ since the proof is easily adapted from the classical minimum function to the p-approximated one.

Proposition A.6. *Let $d = 1$ and the confinement and interaction potentials V and W be convex. Then the discrete energy $\widetilde{E}_N$ defined in (2.4) is convex.*

Proof. Let $\lambda \in [0,1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N_w$. Then, by the facts that min is concave on $\mathbb{R}^2$ and h is non-increasing and convex on $(0,\infty)$, we know that $h \circ \min$ is convex on $(0,\infty)^2$, where $\circ$ is the composition operator. Define, for all $a, b \in \mathbb{R}$, $[a,b]_\lambda = \lambda a + (1-\lambda)b$, and $r_i(\boldsymbol{x}) = \min(\Delta x_i, \Delta x_{i+1})$ and $r_i(\boldsymbol{y}) = \min(\Delta y_i, \Delta y_{i+1})$. Therefore, since V and W are convex,

$$\begin{aligned}
\widetilde{E}_N([\boldsymbol{x},\boldsymbol{y}]_\lambda) &= \sum_{i=1}^N w_i h\left(\frac{1}{w_i}\min([\Delta x_i, \Delta y_i]_\lambda, [\Delta x_{i+1}, \Delta y_{i+1}]_\lambda)\right) \\
&\quad + \sum_{i=1}^N w_i V([x_i, y_i]_\lambda) + \frac{1}{2}\sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j W([x_i - x_j, y_i - y_j]_\lambda) \\
&\leq \lambda \sum_{i=1}^N w_i h\left(\frac{r_i(\boldsymbol{x})}{w_i}\right) + (1-\lambda)\sum_{i=1}^N w_i h\left(\frac{r_i(\boldsymbol{y})}{w_i}\right) + \lambda \sum_{i=1}^N w_i V(x_i) + (1-\lambda)\sum_{i=1}^N w_i V(y_i) \\
&\quad + \frac{\lambda}{2}\sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j W(x_i - x_j) + \frac{1-\lambda}{2}\sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N w_i w_j W(y_i - y_j) \\
&= \lambda \widetilde{E}_N(\boldsymbol{x}) + (1-\lambda)\widetilde{E}_N(\boldsymbol{y}).
\end{aligned}$$

Hence convexity of $\widetilde{E}_N$. ∎

Theorem A.4 now tells us that $\partial_w \widetilde{E}_N^p$ graph-converges to $\partial_w \widetilde{E}_N$ as $p \to \infty$, and Theorem A.5 tells us in which sense the p-approximated discrete gradient flow converges to the discrete one and also gives us some regularity on the discrete and p-approximated discrete gradient flow solutions. The use of the p-approximated gradient flow (2.10) is therefore justified to approximate (2.7) (at least in the case when V and W are convex and $d = 1$).

Acknowledgements. JAC, YH and FSP are supported by Engineering and Physical Sciences Research Council grant EP/K008404/1. JAC is also supported by the Royal Society through a Wolfson Research Merit Award. GW is supported by ISF grant 998/5.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser Basel, 2005.

[2] L. Ambrosio and G. Savaré. Gradient flows of probability measures. In *Handbook of Differential Equations: Evolutionary Equations*, volume 3, pages 1–136. North-Holland, 2007.

[3] H. Attouch. *Variational Convergence for Functions and Operators*. Applicable Mathematics. Pitman Advanced Publishing Program, 1984.

[4] J.-P. Aubin and A. Cellina. *Differential Inclusions*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1984.

[5] J.-D. Benamou, G. Carlier, Q. Mérigot, and E. Oudet. Discretization of functionals involving the Monge–Ampère operator. *Numer. Math.*, 134(3):611–636, 2016.

[6] D. Benedetto, E. Caglioti, J. A. Carrillo, and M. Pulvirenti. A non-Maxwellian steady distribution for one-dimensional granular media. *J. Statist. Phys.*, 91(5-6):979–990, 1998.

[7] M. Bessemoulin-Chatard and F. Filbet. A finite volume scheme for nonlinear degenerate parabolic equations. *SIAM J. Sci. Comput.*, 34(5):B559–B583, 2012.

[8] A. Blanchet, V. Calvez, and J. A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model. *SIAM J. Numer. Anal.*, 46(2):691–721, 2008.

[9] H. Brézis. *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*. North-Holland Mathematics Studies. Elsevier Science, 1973.

[10] V. Calvez and T. Gallouët. Particle approximation of the one dimensional Keller–Segel equation, stability and rigidity of the blow-up. *Discrete Contin. Dyn. Syst.*, 36(3):1175–1208, 2016.

[11] V. Calvez, B. Perthame, and M. Sharifi-tabar. Modified Keller–Segel system and critical mass for the log interaction kernel. *Contemp. Math.*, 429:45–62, 2007.
[12] J. A. Carrillo, A. Chertock, and Y. Huang. A finite-volume method for nonlinear nonlocal equations with a gradient flow structure. *Commun. Comput. Phys.*, 17(1):233–258, 2015.
[13] J. A. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: mean-field limit and Wasserstein distances. In *Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation*, volume 553, pages 1–46. Springer Vienna, 2014.
[14] J. A. Carrillo, M. Di Francesco, and G. Toscani. Strict contractivity of the 2-Wasserstein distance for the porous medium equation by mass-centering. *Proc. Amer. Math. Soc.*, 135(2):353–363, 2007.
[15] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.*, 19(3):971–1018, 2003.
[16] J. A. Carrillo and J. S. Moll. Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms. *SIAM J. Sci. Comput.*, 31(6):4305–4329, 2009.
[17] J. A. Carrillo, F. S. Patacchini, P. Sternberg, and G. Wolansky. Convergence of a particle method for diffusive gradient flows in one dimension. *SIAM J. Math. Anal.*, 48(6):3708–3741, 2016.
[18] P. Degond and F. J. Mustieles. A deterministic approximation of diffusion equations using particles. *SIAM J. Sci. Statist. Comput.*, 11(2):293–310, 1990.
[19] L. Gosse and G. Toscani. Lagrangian numerical approximations to one-dimensional convolution-diffusion equations. *SIAM J. Sci. Comput.*, 28(4):1203–1227, 2006.
[20] S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2000.
[21] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
[22] O. Junge, H. Osberger, and D. Matthes. A fully discrete variational scheme for solving nonlinear Fokker–Planck equations in higher space dimensions. Preprint, *arXiv:1509.07721*.
[23] P.-L. Lions and S. Mas-Gallic. Une méthode particulaire déterministe pour des équations diffusives non linéaires. *C. R. Acad. Sci. Paris Sér. I Math.*, 332(4):369–376, 2001.
[24] S. Mas-Gallic. The diffusion velocity method: a deterministic way of moving the nodes for solving diffusion equations. *Transp. Theory Stat. Phys.*, 31(4-6):595–605, 2002.
[25] R. J. McCann. *A Convexity Theory for Interacting Gases and Equilibrium Crystals*. PhD thesis, Princeton University, 1995.
[26] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
[27] A. Mielke. On evolutionary Gamma-convergence for gradient systems. In *Macroscopic and Large Scale Phenomena: Coarse Graining, Mean Field Limits and Ergodicity*, pages 187–249. Springer, 2016.
[28] H. Osberger and D. Matthes. Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation. *ESAIM Math. Model. Numer. Anal.*, 48:697–726, 2014.
[29] H. Osberger and D. Matthes. Convergence of a fully discrete variational scheme for a thin-film equation. Accepted in *Radon Ser. Comput. Appl. Math.*, 2015.
[30] H. Osberger and D. Matthes. A convergent Lagrangian discretization for a nonlinear fourth order equation. *Found. Comput. Math.*, pages 1–54, 2015.
[31] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
[32] G. Russo. Deterministic diffusion of particles. *Comm. Pure Appl. Math.*, 43(6):697–733, 1990.
[33] G. Russo. A particle method for collisional kinetic equations. I. Basic theory and one-dimensional results. *J. Comput. Phys.*, 87(2):270–300, 1990.
[34] E. Sandier and S. Serfaty. Gamma-convergence of gradient flows with applications to Ginzburg–Landau. *Comm. Pure Appl. Math.*, 57(12):1627–1672, 2004.
[35] S. Serfaty. Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *Discrete Contin. Dyn. Syst.*, 31(4):1427–1451, 2011.
[36] C. M. Topaz, A. L. Bertozzi, and M. A. Lewis. A nonlocal continuum model for biological aggregation. *Bull. Math. Biol.*, 68(7):1601–1623, 2006.
[37] J. L. Vázquez. *The Porous Medium Equation*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, 2007.
[38] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, Providence (R.I.), 2003.

Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
E-mail address: carrillo@imperial.ac.uk

School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK.
E-mail address: yanghong.huang@manchester.ac.uk

Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
E-mail address: f.patacchini13@imperial.ac.uk

Mathematics Dept., Technion–Israel Institute of Technology, Haifa 32000, Israel.
E-mail address: gershonw@math.technion.ac.il