# CONVERGENCE OF THE MASS-TRANSPORT STEEPEST DESCENT SCHEME FOR THE SUB-CRITICAL PATLAK-KELLER-SEGEL MODEL

ADRIEN BLANCHET, VINCENT CALVEZ AND JOSÉ A. CARRILLO

Abstract. Variational steepest descent approximation schemes for the modified Patlak-Keller-Segel equation with a logarithmic interaction kernel in any dimension are considered. We prove the convergence of the suitably interpolated in time implicit Euler scheme, defined in terms of the Euclidean Wasserstein distance, associated to this equation for sub-critical masses. As a consequence, we recover the recent result about the global in time existence of weak-solutions to the modified Patlak-Keller-Segel equation for the logarithmic interaction kernel in any dimension in the sub-critical case. Moreover, we show how this method performs numerically in one dimension. In this particular case, this numerical scheme corresponds to a standard implicit Euler method for the pseudo-inverse of the cumulative distribution function. We demonstrate its capabilities to reproduce easily without the need of mesh-refinement the blow-up of solutions for super-critical masses.

## 1. Introduction

The Patlak-Keller-Segel (PKS) equation is widely used in mathematical biology to model the collective motion of cells which are attracted by a self-emitted chemical substance, being the slime mold amoebae *Dictyostelium discoideum* a prototype organism for this behaviour. Moreover, the PKS equation has become a paradigmatic mathematical problem since it shows a concentration-collapse dichotomy: for masses larger than a critical value solutions aggregate their mass, as Delta Diracs, in finite time while solutions exist globally and disperse collapsing down to zero below this critical mass threshold. This coexistence of phenomena in this simple-looking mathematical model makes appealing and difficult to develop numerical schemes capable of dealing with both situations.

*Key words and phrases.* Patlak-Keller-Segel model, Steepest Descent, Wasserstein Distance.

**AMS subject classifications:** 65M99, 35K55, 35Q99 .

Historically, the first mathematical models in chemotaxis were introduced in 1953 by C. S. Patlak in [36] and E. F. Keller and L. A. Segel in [30] in 1970. Here, we focus on the modified Patlak-Keller-Segel system for the log interaction kernel introduced by B. Perthame, the second author and M. Sharifi tabar in [14]

$$\begin{cases} \dfrac{\partial n}{\partial t}(t,x) = \Delta n(t,x) - \chi \nabla \cdot [n(t,x) \nabla c(t,x)] & t>0\,,\ x\in\mathbb{R}^d\,, \\ c(t,x) = -\dfrac{1}{d\pi}\displaystyle\int_{\mathbb{R}^d} \log|x-y|\, n(t,y)\,\mathrm{d}y\,, & t>0\,,\ x\in\mathbb{R}^d\,, \\ n(0,x) = n_0 \geq 0 & x\in\mathbb{R}^d\,. \end{cases} \tag{1.1}$$

Here $(t,x)\mapsto n(t,x)$ represents the cell density, and $(t,x)\mapsto c(t,x)$ is the concentration of chemo-attractant. The constant $\chi>0$ is the *sensitivity* of the bacteria to the chemo-attractant. Mathematically, it measures the interaction force between cells, and hence, the strength of the non-linear coupling. Initial data are assumed to verify

$$(1+|x|^2)\, n_0 \in L^1_+(\mathbb{R}^d) \quad \text{and} \quad n_0 \log n_0 \in L^1(\mathbb{R}^d). \tag{1.2}$$

The solutions satisfy the formal conservation of the total mass of the system

$$\int_{\mathbb{R}^d} n_0(x)\, dx = \int_{\mathbb{R}^d} n(t,x)\, dx.$$

Without loss of generality we assume that the total mass is 1, such that all the parameters of the system are contained in the reduced parameter $\chi$. The center of mass is also conserved as time evolves, and thus, we fix it to be zero for the sake of simplicity,

$$\int_{\mathbb{R}^d} x\, n(t,x)\, dx = \int_{\mathbb{R}^d} x\, n_0(t,x)\, dx = 0.$$

We first remind that a notion of weak solution $n$ in the space $C^0\big([0,T);L^1_{\rm weak}(\mathbb{R}^d)\big)$, with fixed $T>0$, using the symmetry in $x$, $y$ for the concentration gradient, was introduced in [37] able to handle measure solutions. We shall say that $n$ is a weak solution to the system (1.1) if for all test functions $\zeta\in\mathcal{D}(\mathbb{R}^d)$,

$$\begin{aligned} \frac{d}{dt}\int_{\mathbb{R}^d}\zeta(x)\, n(t,x)\,\mathrm{d}x &= \int_{\mathbb{R}^d}\Delta\zeta(x)\, n(t,x)\,\mathrm{d}x \\ &\quad - \frac{\chi}{2\,d\,\pi}\iint_{\mathbb{R}^d\times\mathbb{R}^d}[\nabla\zeta(x)-\nabla\zeta(y)]\cdot\frac{x-y}{|x-y|^2}\, n(t,s)\, n(t,y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{1.3}$$

together with $n(t=0)=n_0$ in the distributional sense in $(0,T)$.

As proved in [10, 8, 14], this problem presents the following dichotomy: either solutions blow-up in finite time for the super-critical case $\chi > 2d^2\pi$ or rather solutions exist globally in time and spread in space decaying towards a stationary solution in rescaled variables as $t \to \infty$ in the sub-critical case $\chi < 2d^2\pi$.

Global improved weak solutions have been constructed for the system (1.1) in the sub-critical case, $\chi < 2d^2\pi$, for $d = 2$ [24, 8] and $d \neq 2$ in [14]; and in the critical case for $d = 2$ in [7]. These improved weak solutions satisfy the decreasing character of a free energy functional for the PKS equation given by:

$$t \mapsto \mathcal{F}[n](t) := \mathcal{S}[n](t) + \mathcal{W}[n](t) \tag{1.4}$$

where $\mathcal{S}[n]$ is the standard Boltzmann's entropy and $\mathcal{W}[n]$ is the interaction energy defined by

$$\mathcal{S}[n](t) := \int_{\mathbb{R}^d} n(t,x) \log n(t,x)\, \mathrm{d}x$$
$$\text{and} \quad \mathcal{W}[n](t) := \frac{\chi}{2d\pi} \iint_{\mathbb{R}^d} n(t,x)\, n(t,y)\, \log|x-y|\, \mathrm{d}x\, \mathrm{d}y.$$

The free energy $\mathcal{F}[n]$ is related to its time derivative, the corresponding Fisher information, in the following way: consider a non-negative solution $n \in \mathcal{C}^0([0,T), L^1(\mathbb{R}^d))$ of the Patlak-Keller-Segel system (1.1) such that $n(1+|x|^2)$, $n \log n$ are bounded in $L^\infty((0,T), L^1(\mathbb{R}^d))$, $\nabla\sqrt{n} \in L^1((0,T), L^2(\mathbb{R}^d))$ and $\nabla c \in L^\infty((0,T)\times\mathbb{R}^d)$, then

$$\frac{d}{dt}\mathcal{F}[n](t) = -\int_{\mathbb{R}^d} n(t,x)\, |\nabla \log n(t,x) - \chi \nabla c(t,x)|^2 \, \mathrm{d}x. \tag{1.5}$$

The functional $\mathcal{F}$ structurally belongs to the general class of free energies for interacting particles introduced in [33, 18, 19] and further analysed in [22, 2]. The functionals treated in those references are of the general form:

$$\mathcal{E}[n] := \int_{\mathbb{R}^d} U[n(x)]\, \mathrm{d}x + \int_{\mathbb{R}^d} n(x)\, V(x)\, \mathrm{d}x + \frac{1}{2} \iint_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\, n(x)\, n(y)\, \mathrm{d}x\, \mathrm{d}y \tag{1.6}$$

under the basic assumptions $U : \mathbb{R}^+ \to \mathbb{R}$ is a density of internal energy, $V : \mathbb{R}^d \to \mathbb{R}$ is a convex confinement potential and $W : \mathbb{R}^d \to \mathbb{R}$ is a symmetric convex interaction potential. The internal energy $U$ should satisfy the following dilation condition, introduced in McCann [33]

$$\lambda \longmapsto \lambda^d U(\lambda^{-d}) \qquad \text{is convex non-increasing on } \mathbb{R}^+. \tag{1.7}$$

The most important case of application, as it is for our case, is $U(s) = s \log s$, which identifies the internal energy with Boltzmann's entropy.

Continuity equations where the velocity field is formally derived from the variational derivative of free energy functionals of the type (1.6), given by

$$\frac{\partial \rho}{\partial t} = \operatorname{div}\left(\rho \nabla \frac{\delta \mathcal{E}}{\delta \rho}\right), \qquad \text{in} \qquad (0,+\infty)\times\mathbb{R}^d\,, \tag{1.8}$$

appear in various contexts: the interest for a convex interaction potential energy arose from its use in the modelling of granular flows: see the works of D. Benedetto, E. Caglioti, the last author, M. Pulvirenti, G. Toscani and C. Villani [5, 6, 39, 44] and the references therein for the physical background and related mathematical analysis. Nice mathematical and physical reviews are provided in [43, Chapter 5] and [44].

A very powerful theory has been developed in the past decade starting from the seminal paper by R. McCann [33] where the notion of displacement convexity for a functional acting on probability measures was introduced. This notion provides functionals of the form (1.6) with a natural convexity structure. However, the interacting kernel $W$ is itself required to be convex. Later, F. Otto [35] introduced a formal Riemaniann structure giving sense to this family of equations (1.8) as the gradient flow of the convex free-energy functional (1.8) with respect to a metric that induces the euclidean Wasserstein distance for measures. Geodesics in Otto's interpretation correspond to optimal transportation pathways (or *displacement interpolation*),

$$\rho^t = \Big((1-t)\mathrm{Id} + t\nabla\varphi\Big)\#\rho^0,$$

where $T = \nabla\varphi$ is the optimal static transport map between the endpoints $\rho^0$ and $\rho^1$.

On the other hand, a steepest descent scheme based on optimal transport of measures was introduced in [29] for the linear Fokker-Planck equation, exhibiting very nice properties. This scheme is now well understood and has been formalised for a large class of degenerate parabolic equations in [1], and in a more abstract setting, by L. Ambrosio, N. Gigli and G. Savaré [3] with the name of 'minimising movement scheme'. The idea corresponds to a discrete version of the gradient flow or steepest descent of the free energy under the Wasserstein metric structure, see Section 2 below for precise definitions.

In our case, the free-energy functional shows a non convex interaction potential, characteristic also of other models in mathematical biology [9, 10] and swarming [38]. To weaken the convexity assumption on the interaction kernel and to find under which conditions stationary states continue to be global attractors of the dynamics are issues of great interest for applications in mathematical biology.

The main results of this work, Theorem 3.4 and Proposition 4.1, show the convergence of the Jordan-Kinderlehrer-Otto steepest descent discrete method [29] using Otto's interpretation of the PKS equation (1.1) as the gradient flow of the free-energy functional for the sub-critical case and the exponential convergence towards a unique stationary profile in scaled variables for the sub-critical one-dimensional case. The first result recovers the available global existence results in the sub-critical cases for the PKS equation in [10, 8, 14]. Moreover, we solve numerically this scheme in the one dimensional case showing its abilities on capturing the blow-up for super-critical cases without the need of mesh-refinement.

The plan of this paper is the following: we first recall in Section 2.1 some recent results on free energies and rescaled variables which allows to obtain *a priori* estimates. We remind in Section 2.2 notions on optimal transport and on the Wasserstein distance that we will use in Section 3 to prove the convergence of the scheme (3.1). The exponential convergence towards a unique equilibrium is shown in the scaled one-dimensional setting in Subsection 4.1. Finally, one-dimensional numerical simulations are given in Subsection 4.2.

## 2. Preliminaries

2.1. *A priori* **estimates in the sub-critical case.** Here, we review some aspects of the PKS model that were already used in [10, 8, 14, 7] as the main tools for the proof of global existence of weak solutions in the sub-critical and critical cases, respectively.

We will make a fundamental use of the Logarithmic Hardy-Littlewood-Sobolev inequality [4, 15]: let $f$ be a non-negative function in $L^1(\mathbb{R}^d)$ such that $f\log f$ and $f\log(1+|x|^2)$ belong to $L^1(\mathbb{R}^d)$. If

$$\int_{\mathbb{R}^d} f\,dx = 1,$$

then

$$\int_{\mathbb{R}^d} f(x)\log f(x)\,\mathrm{d}x + d\iint_{\mathbb{R}^d\times\mathbb{R}^d} f(x)f(y)\log|x-y|\,\mathrm{d}x\,\mathrm{d}y \geq -\,C(d) \quad (2.1)$$

with $C(d) := (1/2)\log\pi + (1/d)\log[\Gamma(d/2)/\Gamma(d)] + (1/2)[\psi(d) - \psi(d/2)]$ where $\psi$ is the logarithmic derivative of the $\Gamma$-function. The Logarithmic Hardy-Littlewood-Sobolev inequality (2.1) implies that the functional energy (1.4) is bounded from below if $\chi = \chi_c := 2\,d^2\pi$.

Since we will work mainly in the sub-critical case $\chi < \chi_c$, it is clearer, although not necessary, to solve the equation in rescaled variables. Let us

define the rescaled functions $\rho$ and $v$ by:

$$n(t,x)=\frac{1}{R^d(t)}\,\rho\left(\tau(t),\frac{x}{R(t)}\right)\quad\text{and}\quad c(x,t)=v\left(\tau(t),\frac{x}{R(t)}\right)\tag{2.2}$$

with

$$R(t)=\sqrt{1+2t}\quad\text{and}\quad\tau(t)=\log R(t)\,.$$

The rescaled system is

$$\begin{cases}\dfrac{\partial\rho}{\partial t}(t,x)=\Delta\rho(t,x)+\nabla\cdot\{\rho(t,x)\,[x-\chi\nabla v(t,x)]\} & t>0\,,\ x\in\mathbb{R}^d\,,\\ v(t,x)=-\dfrac{1}{d\,\pi}\,\log|\cdot|*\rho(t,x)-\dfrac{1}{d\,\pi}\,\tau(t) & t>0\,,\ x\in\mathbb{R}^d\,,\\ \rho(0,x)=\rho^0=n_0\ge0 & x\in\mathbb{R}^d\,.\end{cases}\tag{2.3}$$

In the rescaled variables, the confinement potential $V(x)=\frac12|x|^2$ is added and the free energy becomes

$$\begin{aligned}\mathcal{G}[\rho]=&\int_{\mathbb{R}^d}\rho(x)\log\rho(x)\,\mathrm{d}x+\frac12\int_{\mathbb{R}^d}|x|^2\,\rho(x)\,\mathrm{d}x\\&+\frac{\chi}{2\,d\,\pi}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\log|x-y|\,\rho(x)\,\rho(y)\,\mathrm{d}x\,\mathrm{d}y\end{aligned}\tag{2.4}$$

With the definition (1.3) we shall say that $\rho$ is a weak solution to the system (2.3) if for all test functions $\zeta\in\mathcal{D}(\mathbb{R}^d)$,

$$\begin{aligned}\frac{d}{dt}\int_{\mathbb{R}^d}\zeta(x)\,\rho(t,x)\,\mathrm{d}x=&\int_{\mathbb{R}^d}\Delta\zeta(x)\,\rho(t,x)\,\mathrm{d}x-\int_{\mathbb{R}^d}\nabla\zeta(x)\cdot x\,\rho(t,x)\,\mathrm{d}x\\&-\frac{\chi}{2\,d\,\pi}\iint_{\mathbb{R}^d\times\mathbb{R}^d}[\nabla\zeta(x)-\nabla\zeta(y)]\cdot\frac{x-y}{|x-y|^2}\,\rho(t,x)\,\rho(t,y)\,\mathrm{d}x\,\mathrm{d}y\end{aligned}\tag{2.5}$$

together with $\rho(t=0)=\rho^0$ in the distributional sense in $(0,T)$. The following Lemma extracts enough information from this decreasing free energy to proceed.

**Lemma 2.1** (*A priori* estimates)**.** *The functional $\mathcal{G}$ is bounded from below on the set*

$$\mathcal{K}:=\Big\{\rho\in L^1_+(\mathbb{R}^d)\quad:\quad\int_{\mathbb{R}^d}\rho(t,x)=1,|x|^2\,\rho\in L^1(\mathbb{R}^d),\ \int_{\mathbb{R}^d}\rho(t,x)\,|\log\rho(t,x)|\,\mathrm{d}x<\infty\Big\}$$

*if and only if $\chi\le\chi_c$. In addition, if $\chi<\chi_c$ we have on every subset $\{\mathcal{G}\le C\}$,*

i) no concentration: $\displaystyle\int_{\mathbb{R}^d} \rho|\log\rho| \le C$,

ii) mass confinement: $\displaystyle\int_{\mathbb{R}^d} |x|^2\rho \le C$,

*As a consequence, every level subset* $\{\mathcal{G} \le C\}$ *is equi-integrable.*

*Proof.* The first use of the Logarithmic Hardy-Littlewood-Sobolev inequality (2.1) to bound from below the free energy $\mathcal{G}$ is due to [24]. Rewrite

$$\mathcal{G}[\rho](t) = (1-\theta)\int_{\mathbb{R}^d} \rho(t,x)\,\log\rho(t,x)\,\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}^d}|x|^2\,\rho(t,x)\,\mathrm{d}x \tag{2.6}$$
$$+\theta d\left[\frac{1}{d}\int_{\mathbb{R}^d}\rho(t,x)\log\rho(t,x)\,\mathrm{d}x + \frac{\chi}{2d^2\pi\theta}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\rho(t,x)\rho(t,y)\log|x-y|\,\mathrm{d}x\,\mathrm{d}y\right].$$

The Logarithmic Hardy-Littlewood-Sobolev inequality (2.1) controls the third term if we choose $\theta = \chi/\chi_c$. Because the function $\rho\log\rho$ is negative for small $\rho$, we need to control somehow the density for large $x$. We use in fact the confinement potential, i.e., the second momentum of $\rho$.

**Lemma 2.2** (Carleman's estimates)**.** *For any probability density* $u \in L^1_+(\mathbb{R}^d)$*, if the second moment* $\int_{\mathbb{R}^d}|x|^2\,u(x)\,\mathrm{d}x$ *and the entropy* $\int_{\mathbb{R}^d} u\log u\,\mathrm{d}x$ *are bounded from above, then* $u\log u$ *is uniformly bounded in* $L^1(\mathbb{R}^d)$ *and we have*

$$\int_{\mathbb{R}^d} u(x)\,|\log u(x)|\,\mathrm{d}x \le \int_{\mathbb{R}^d} u(x)\left(\log u(x) + \frac{1}{2}|x|^2\right)\mathrm{d}x + d\log(4\pi) + \frac{2}{e}\,.$$

*Proof.* The proof goes as follows. Let $\bar u := u\mathbb{1}_{\{u\le 1\}}$ and $m = \int_{\mathbb{R}^d}\bar u(x)\,\mathrm{d}x \le \int_{\mathbb{R}^d} u(x)\,\mathrm{d}x = 1$. Then

$$\int_{\mathbb{R}^d}\bar u(x)\,\left(\log\bar u(x) + \frac{1}{4}|x|^2\right)\,\mathrm{d}x = \int_{\mathbb{R}^d} U(x)\log U(x)\,\mathrm{d}\mu - m\frac{d}{2}\log(4\pi)$$

where $U := \bar u/\mu$, $d\mu(x) = (4\pi)^{-d/2}e^{-|x|^2/4}\,\mathrm{d}x$. The Jensen inequality yields

$$\int_{\mathbb{R}^d} U(x)\,\log U(x)\,\mathrm{d}\mu \ge \left(\int_{\mathbb{R}^d} U(x)\,\mathrm{d}\mu\right)\,\log\left(\int_{\mathbb{R}^d} U(x)\,\mathrm{d}\mu\right) = m\,\log m$$

and

$$\int_{\mathbb{R}^d}\bar u(x)\,\log\bar u(x)\,\mathrm{d}x \ge m\log m - m\frac{d}{2}\log 4\pi - \frac{1}{4}\int_{\mathbb{R}^d}|x|^2\,\bar u(x)\,\mathrm{d}x$$
$$\ge -\frac{1}{e} - \frac{d}{2}\log(4\pi) - \frac{1}{4}\int_{\mathbb{R}^d}|x|^2\,u(x)\,\mathrm{d}x\,.$$

Using

$$\int_{\mathbb{R}^d} u(x)\,|\log u(x)|\,\mathrm{d}x = \int_{\mathbb{R}^d} u(x)\,\log u(x)\,\mathrm{d}x - 2\int_{\mathbb{R}^d} \bar{u}(x)\,\log \bar{u}(x)\,\mathrm{d}x\,,$$

this completes the proof. □

We apply this lemma to obtain the first part of the result from (2.6):

$$\mathcal{G}[\rho](t) \geq (1-\theta)\int_{\mathbb{R}^d} \rho(t,x)\,|\log \rho(t,x)|\,\mathrm{d}x + \frac{\theta}{2}\int_{\mathbb{R}^d} |x|^2\,\rho(t,x)\,\mathrm{d}x + C.$$

On the other hand, the functional $\mathcal{G}[\rho]$ has an interesting scaling property. For a given $\rho$, let $\rho_\lambda(x) = \lambda^d \rho(\lambda x)$. It is straightforward to check that $\|\rho_\lambda\|_{L^1(\mathbb{R}^d)} = 1$ and

$$\mathcal{G}[\rho_\lambda] = \mathcal{G}[\rho] + d\left(1 - \frac{\chi}{\chi_c}\right)\log\lambda + \frac{\lambda^{-2}-1}{2}\int_{\mathbb{R}^d} |x|^2\,\rho\,\mathrm{d}x\,.$$

Since $\lambda \mapsto \mathcal{G}[\rho_\lambda]$ is clearly not bounded from below if $\chi > \chi_c$, the proof concludes. □

We shall also state another technical Lemma, which will plays a major role when passing to the limit in the quadratic interaction contribution.

**Lemma 2.3** (Doubling of variables)**.** *Assume $f_i \rightharpoonup f$ in $L^1$ and the family $\{f_i\}$ is equi-integrable in the sense of Lemma* 2.1*, then $f_i \otimes f_i \rightharpoonup f \otimes f$ in $L^1(\mathbb{R}^d \times \mathbb{R}^d)$.*

*Proof.* Let $\psi(x,y)$ denote any test function in $L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$. For any $t \in (0,T)$ and for almost every $x \in \mathbb{R}^d$ define

$$\lim \Psi^i(x) := \lim \int_{\mathbb{R}^d} f_i(y)\,\psi(x,y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f(y)\,\psi(x,y)\,\mathrm{d}y =: \Psi(x)$$

Note that for any $t$ and $x$, $|\Psi^i(x)|$ and $|\Psi(x)|$ are bounded by $\|\psi\|_{L^\infty}$.

By Egorov's theorem for any $R > 0$ and $\delta > 0$, there exists $X_\delta$ such that $|X_\delta| < \delta$ and $\Psi^i$ uniformly converges to $\Psi$ in $B_R \setminus X_\delta$. We have

$$\begin{aligned}
&\left| \int_{\mathbb{R}^d} \left[ \Psi^i(x)\, f_i(x) - \Psi(x)\, f(x) \right] \, \mathrm{d}x \right| \\
&\quad \leq \int_{B_R \setminus X_\delta} \left| \Psi^i(x) f_i(x) - \Psi(x) f(x) \right| \, \mathrm{d}x + \int_{X_\delta} \left| \Psi^i(x) f_i(x) - \Psi(x) f(x) \right| \, \mathrm{d}x \\
&\qquad + \int_{B_R^c} \left| \Psi^i(x)\, f_i(x) - \Psi(x)\, f(x) \right| \, \mathrm{d}x \\
&\quad \leq \int_{B_R \setminus X_\delta} \left| \Psi^i(x)\, f_i(x) - \Psi(x)\, f(x) \right| \, \mathrm{d}x + \|\psi\| \int_{X_\delta} f_i(x) + f(x) \, \mathrm{d}x \\
&\qquad + \|\psi\| \, \frac{1}{R^2} \int_{B_R^c} |x|^2 \left[ f_i(x) + f(x) \right] \, \mathrm{d}x \, .
\end{aligned}$$

Egorov's theorem and the weak-$L^1$ convergence of $f_i$ towards $f$ ensures that the first term is as small as desired by choosing $i$ large enough. By the *a priori* estimates in Lemma 2.1, $\int_{X_\delta} f_i(x) \, \mathrm{d}x$ and $\int_{X_\delta} f(x) \, \mathrm{d}x$ can me made as small as desired by choosing $\delta$ small enough, as well as the third term can be made as small as desired by choosing $R$ large enough. □

2.2. **Optimal transport and the Wasserstein distance.** We recall some standard results related to optimal transportation and Wasserstein distance that we will use in the sequel of this paper. The interested reader can refer to the books of C. Villani [43, 45] and the book of L. Ambrosio, N. Gigli and G. Savaré [3]. A short summary of properties of the Wasserstein distance can be seen in [21].

Let $\mu$ and $\nu$ be in $\mathcal{P}(\mathbb{R}^d)$ the space of probability measure in $\mathbb{R}^d$, $\mathcal{P}_2(\mathbb{R}^d)$ the subset of probability measures with finite second-momentum, $\mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ its subset formed by the absolutely continuous measures with respect to Lebesgue and $T$ be a measurable map $\mathbb{R}^d \to \mathbb{R}^d$. We say that $T$ *transports* $\mu$ onto $\nu$ and we note $\nu = T\#\mu$ if for any measurable set $B \subset \mathbb{R}^d$, $\nu(B) = \mu \circ T^{-1}(B)$. We also say $\nu$ is the *push-forward* or *the image measure* of $\mu$ by $T$ *i.e.*

$$\int_{\mathbb{R}^d} \zeta \circ T(x) \, \mathrm{d}\mu(x) = \int_{\mathbb{R}^d} \zeta(y) \, \mathrm{d}\nu(y) \quad \forall \zeta \in \mathcal{C}_b^0(\mathbb{R}^d) \, . \tag{2.7}$$

The Wasserstein distance between $\mu$ and $\nu$, $d_W$ can be defined by

$$d_W^2(\mu, \nu) := \inf_{T:\ \nu = T\#\mu} \int_{\mathbb{R}^d} |x - T(x)|^2 \, \mathrm{d}\mu(x) \, .$$

By Brenier's theorem [11, 32, 34], see [43, Theorem 2.32, p.85] for a review, if $\mu$ is absolutely continuous with respect to Lebesgue measure, then there

is one measurable plan $T$ such that $\nu = T\#\mu$ and $T = \nabla\varphi$ for some convex function $\varphi$. As a consequence,

$$d_W^2(\mu,\nu) = \int_{\mathbb{R}^d} |x - \nabla\varphi(x)|^2 \, \mathrm{d}\mu(x) \, . \tag{2.8}$$

The variational problem leading to the definition of the Wasserstein distance can be relaxed to the linear program

$$d_W^2(\mu,\nu) = \inf_{\Pi\in\Gamma} \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \, d\Pi(x,y) \right\},$$

where $\Pi$ runs over the set of transference plans $\Gamma$, that is, the set of joint probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mu$ and $\nu$. In fact, the infimum above is a minimum by Kantorovich duality theorems [43, Chapter 1]. The optimal transference plan, in case Brenier's theorem applies, is given by $\Pi_o = (id_{\mathbb{R}^d} \otimes \nabla\varphi)\#\mu$.

Let us remind a simple consequence of the definition of the Wasserstein distance for controlling averages [21, Corollary 2.4].

**Lemma 2.4** (Convergence of averages with $d_W$)**.** *Given $\zeta$ a Lipschitz function with Lipschitz constant $L$ and $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$, then we have*

$$\left| \int_{\mathbb{R}^d} \zeta(x) \, \mathrm{d}\mu - \int_{\mathbb{R}^d} \zeta(x) \, \mathrm{d}\nu \right| \le L \, d_W(\mu,\nu).$$

*Proof.* Let $\Pi_o(x,y)$ the optimal plan between $\mu$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ for $d_W$. Then

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\Pi_o(x,y) = d_W^2(\mu,\nu),$$

and we can write

$$\int_{\mathbb{R}^d} \zeta(x) \, \mathrm{d}\mu - \int_{\mathbb{R}^d} \zeta(x) \, \mathrm{d}\nu = \int_{\mathbb{R}^d\times\mathbb{R}^d} (\zeta(x) - \zeta(y)) \, d\Pi_o(x,y).$$

Using that $\zeta$ is Lipschitz with constant $L$ and estimating by Hölder's inequality, we get

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \zeta(x) \, \mathrm{d}\mu - \int_{\mathbb{R}^d} \zeta(x) \, \mathrm{d}\nu \right| &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |\zeta(x) - \zeta(y)| \, d\Pi_o(x,y) \\ &\le L \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y| \, d\Pi_o(x,y) \le L \, d_W(\mu,\nu), \end{aligned}$$

giving the assertion. □

## 3. Time discretisation

We consider a time-step $\tau > 0$, an initial datum $\rho^0 \in \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$. We introduce the sequence $(\rho_\tau^n)_{n\in\mathbb{N}}$ recursively defined by $\rho_\tau^0 = \rho^0$ and

$$\rho_\tau^{n+1} \in \arg\inf_{\rho\in\mathcal{K}} \left\{ \mathcal{G}[\rho] + \frac{1}{2\,\tau} d_W^2(\rho_\tau^n, \rho) \right\} . \tag{3.1}$$

The JKO steepest descent scheme can be viewed formally as a time discretisation of the abstract gradient flow equation,

$$\frac{\partial \rho}{\partial t} = -\text{"}\nabla_W\text{"}\mathcal{G}[\rho] ,$$

where the space $\mathcal{K} \subset \mathcal{P}_2^{\mathrm{ac}}(\mathbb{R}^d)$ is endowed with a formal riemannian structure compatible with the Wasserstein $d_W$ distance [35]. We refer to [35, 43, 3, 19] for a deeper discussion and the rigorous sense of the "$\nabla_W$" definition. Next lemma ensures that this discrete scheme is well defined.

**Lemma 3.1** (Existence of minimisers)**.** *Let $\rho_0$ satisfies* (1.2) *and $\chi < \chi_c$, then there recursively exists a minimiser to* (3.1)*.*

*Proof.* Introduce the function

$$\mathcal{K} \ni \xi \mapsto \mathcal{G}[\xi] + \frac{1}{2\,\tau} d_W^2(\rho_\tau^n, \xi) . \tag{3.2}$$

By the *a priori* estimates in Lemma 2.1, this function is bounded from below. Consider $(\xi_k)_{k\in\mathbb{N}}$ a minimising sequence, without loss of generality, we can assume that it satisfies $\mathcal{G}[\xi_k] \le \mathcal{G}[\rho_\tau^n]$ for all $k \in \mathbb{N}$. Proceeding as in Lemma 2.1, we get

$$(1-\theta) \int_{\mathbb{R}^d} \xi_k(x)\, |\log \xi_k(x)|\, \mathrm{d}x + \frac{\theta}{2} \int |x|^2 \xi_k(x) \le \mathcal{G}[\rho_\tau^n] + \theta\, C(d) .$$

If $\chi < \chi_c$, then $\theta = \dfrac{\chi}{\chi_c} < 1$ and this shows that $\xi_k \log \xi_k$ is bounded in $L^1(\mathbb{R}^d)$.

The bound on the second momentum avoid vanishing, while the $L^1(\mathbb{R}^d)$-bound on $\xi_k \log \xi_k$ avoid concentration: indeed,

$$\int_{\{\xi_k \ge Q\}} \xi_k \, \mathrm{d}x \le \frac{1}{\log Q} \int_{\{\xi_k \ge Q\}} \xi_k \log \xi_k \, \mathrm{d}x \le \frac{1}{\log Q} \int_{\mathbb{R}^d} \xi_k \, |\log \xi_k| \, \mathrm{d}x ,$$

can be made as small as desired for $Q > 1$ large enough. Hence the family $\{\xi_k\}_{k\in\mathbb{N}}$ verifies the hypotheses in Dunford-Pettis theorem, and thus, there exists a sub-sequence still denoted $(\xi_k)_{k\in\mathbb{N}}$ which converges weakly $L^1$ to a density $\xi_*$.

It remains to prove that this candidate $\xi_*$ realizes in fact a minimum of (3.2). The weak-$L^1$ lower semi-continuity of the entropy $\mathcal{S}$, the second momentum and the Wasserstein distance are well known, see [29, 3] and references therein. We will prove that the quadratic interaction term is continuous for the weak-$L^1$ convergence in our situation. We split it into

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \xi_k(x)\,\xi_k(y)\,\log|x-y|\,\mathrm{d}x\,\mathrm{d}y = \mathrm{A}(t)+\mathrm{B}(t)+\mathrm{C}(t),$$

with

$$\mathrm{A}(t) := \iint_{|x-y|<\varepsilon} \xi_k(x)\,\xi_k(y)\,\log|x-y|\,\mathrm{d}x\,\mathrm{d}y,$$

$$\mathrm{B}(t) := \iint_{\varepsilon\le|x-y|\le R} \xi_k(x)\,\xi_k(y)\,\log|x-y|\,\mathrm{d}x\,\mathrm{d}y$$

$$\text{and}\quad \mathrm{C}(t) := \iint_{|x-y|>R} \xi_k(x)\,\xi_k(y)\,\log|x-y|\,\mathrm{d}x\,\mathrm{d}y\ ,$$

where $\varepsilon<1$ and $R>\sqrt{e}$.

**Control of** $\mathrm{A}(t)$**.** We use the duality inequality, $ab\le e^a+b\log b-b$.

$$\begin{aligned}
\big|\mathrm{A}(t)\big| \le\ & \\
\le\ & \iint_{|x-y|<\varepsilon} \xi_k(x)\xi_k(y)\log\frac{1}{|x-y|}\,\mathrm{d}x\,\mathrm{d}y \\
\le\ & \int\Big\{\xi_k(x)\int_{|x-y|<\varepsilon}\alpha^{-1}\xi_k(y)\log(\alpha^{-1}\xi_k(y))-\alpha^{-1}\xi_k(y)+\exp\Big(\alpha\log\frac{1}{|x-y|}\Big)\Big\} \\
\le\ & \alpha^{-1}\int\xi_k(y)\log(\alpha^{-1}\xi_k(y))\int_{|x-y|<\varepsilon}\xi_k(x)\,\mathrm{d}x\,\mathrm{d}y+\int\xi_k(x)\,\mathrm{d}x\int_{|z|<\varepsilon}\frac{1}{|z|^\alpha}\,\mathrm{d}z.
\end{aligned}$$

By the $L^1(\mathbb{R}^d)$-bound on $\xi_k\log\xi_k$, $\int_{X_\varepsilon}\xi_k$ is uniformly small on small sets $X_\varepsilon$, and therefore the last term can be made as small as desired uniformly in $k$ by choosing $\varepsilon$ small enough and $\alpha<d$.

**Control of** $\mathrm{B}(t)$**.** We shall use the Lemma 2.3 because $\log|x-y|$ is bounded on the set $\{\varepsilon<|x-y|<R\}$

**Control of** $\mathrm{C}(t)$**.** For $R>\sqrt{e}$, $R\mapsto R^2/\log R$ is an increasing function, so that

$$0\le\mathrm{C}(t)\le\frac{2\,\log R}{R^2}\int_{\mathbb{R}^2}|x|^2\,\xi_k(t,x)\,\mathrm{d}x\ .$$

Because we uniformly bound the second momentum, this can be made as small as desired while choosing $R$ large enough.

Finally, collecting terms we get that the difference

$$\left| \iint_{\mathbb{R}^d \times \mathbb{R}^d} \xi_k(t,x)\, \xi_k(t,y) \log|x-y| \, \mathrm{d}x \, \mathrm{d}y - \iint_{\mathbb{R}^d \times \mathbb{R}^d} \xi_*(t,x)\, \xi_*(t,y) \log|x-y| \, \mathrm{d}x \, \mathrm{d}y \right|$$

can be made as small as desired by choosing $\varepsilon$ and $\delta$ small enough, and $r$, $R$ and $k$ large enough. □

**Remark 3.2** (Uniqueness of Minimizers)**.** *Since the functional $\mathcal{G}[n]$ is not convex, we cannot conclude the uniqueness of minimizers for the discrete scheme, and thus, the scheme* (3.1) *is defined by choosing any element realizing the infimum as $\rho_\tau^{n+1}$. Each choice might in principle give rise to a solution in the limit $\tau \to 0$. It is an open problem to deal with the uniqueness of solutions in the subcritical case.*

Now, we define the time interpolation of the discrete scheme as a family of Lipschitz curves $(\rho_\tau)_{\tau>0}$ connecting every pair $\{\rho_\tau^n, \rho_\tau^{n+1}\}$ with a constant speed geodesic in the interval $[n\,\tau, (n+1)\,\tau)$. Accordingly for any $t \in [n\,\tau, (n+1)\,\tau)$ we have,

$$d_W\left(\rho_\tau^n, \rho_\tau(t)\right) = \frac{t - n\tau}{\tau}\, d_W\left(\rho_\tau^n, \rho_\tau^{n+1}\right) \ .$$

Obviously $\rho_\tau(n\tau) = \rho_\tau^n$. This is possible due to Brenier theorem by defining the displacement interpolant

$$\rho_\tau(t) = \left( \frac{(n+1)\tau - t}{\tau} \mathrm{Id} + \frac{t - n\tau}{\tau} \nabla\varphi^n \right) \# \rho_\tau^n$$

with $\nabla\varphi^n$ being the optimal map transporting $\rho_\tau^n$ onto $\rho_\tau^{n+1}$.

**Remark 3.3** (Comparison to Literature)**.** *Let us point out that once the free energy $\mathcal{G}[n]$ is bounded from below (Lemma* 2.1*) and the approximation scheme is well-defined (Lemma* 3.1*), then* [3, Theorem 11.1.6, pp. 288-289] *applies. For the convenience of the reader we give a shorter proof adapted to our problem. Our proof is based on the founding idea of* [29] *proving the convergence of the ad-hoc scheme for the linear Fokker-Planck equation. A nice sketch of the proof of* [29] *can be found in* [43, Section 8.4,pp. 256-262].

**Theorem 3.4** (Convergence of the scheme as $\tau \to 0$)**.** *Under assumption* (1.2)*, if $\chi < \chi_c$ then the family $(\rho_\tau)_{\tau>0}$ admits a sub-sequence converging weakly in $L^1(\mathbb{R}^d)$ to a weak solution of* (2.5).

*Proof.* We proceed in three steps:
**Step 1.- *A priori* estimates in space and time**: Since $\rho_\tau^{n+1}$ minimises (3.1) we obviously have

$$\mathcal{G}[\rho_\tau^{n+1}] + \frac{1}{2\,\tau} d_W^2(\rho_\tau^n, \rho_\tau^{n+1}) \le \mathcal{G}[\rho_\tau^n]\,.$$

As a consequence we obtain an *energy estimate*

$$\sup_{n\in\mathbb{N}} \mathcal{G}[\rho_\tau^n] \le \mathcal{G}[\rho_\tau^0], \tag{3.3}$$

and a *total square estimate*

$$\frac{1}{2\,\tau} \sum_{n\in\mathbb{N}} d_W^2(\rho_\tau^n, \rho_\tau^{n+1}) \le \mathcal{G}[\rho_\tau^0] - \inf_{n\in\mathbb{N}} \mathcal{G}[\rho_\tau^n]\,. \tag{3.4}$$

The right-hand side is bounded thanks to Lemma 2.1.

From the total square estimate (3.4) we deduce a $\frac{1}{2}$-*Hölder-estimate* in time of $(\rho_\tau)_{\tau\ge0}$: indeed, for any $0 \le m \le n$,

$$\begin{aligned} d_W(\rho_\tau^m, \rho_\tau^n) &\le \sqrt{2\,\tau} \sum_{k=m}^{n-1} \frac{1}{\sqrt{2\,\tau}} d_W(\rho_\tau^k, \rho_\tau^{k+1}) \\ &\le \sqrt{2\,\tau}\,\sqrt{(n-m)}\sqrt{\mathcal{G}[\rho_\tau^0] - \inf_{n\in\mathbb{N}} \mathcal{G}[\rho_\tau^n]}\,. \end{aligned}$$

As a consequence for any $0 \le s \le t$

$$\begin{aligned} d_W(\rho_\tau(s), \rho_\tau(t)) &\le d_W\Big(\rho_\tau(s), \rho_\tau^{\left[\frac{s}{\tau}+1\right]\tau}\Big) + d_W\Big(\rho_\tau^{\left[\frac{s}{\tau}+1\right]\tau}, \rho_\tau^{\left[\frac{t}{\tau}\right]\tau}\Big) \\ &\quad + d_W\Big(\rho_\tau^{\left[\frac{t}{\tau}\right]\tau}, \rho_\tau(t)\Big) \\ &\le \left(\left[\frac{s}{\tau}+1\right] - \frac{s}{\tau}\right) d_W\Big(\rho_\tau^{\left[\frac{s}{\tau}\right]\tau}, \rho_\tau^{\left[\frac{s}{\tau}+1\right]\tau}\Big) \\ &\quad + \sqrt{2\,\tau\left(\left[\frac{t}{\tau}\right] - \left[\frac{s}{\tau}+1\right]\right)\Big(\mathcal{G}(\rho_\tau^0) - \inf_{n\in\mathbb{N}} \mathcal{G}(\rho_\tau^n)\Big)} \\ &\quad + \left(\frac{t}{\tau} - \left[\frac{t}{\tau}\right]\right) d_W\Big(\rho_\tau^{\left[\frac{t}{\tau}\right]\tau}, \rho_\tau^{\left[\frac{t}{\tau}+1\right]\tau}\Big) \\ &\le \sqrt{6\,\Big(\mathcal{G}(\rho^0) - \inf_{n\in\mathbb{N}} \mathcal{G}(\rho^n)\Big)}\,(t-s)^{\frac{1}{2}}\,. \end{aligned} \tag{3.5}$$

**Step 2.- Compactness**: By the $\frac{1}{2}$-Hölder-estimate (3.5), $\rho_\tau$ is bounded in $\mathcal{P}_2(\mathbb{R}^d)$ so the family $\{\rho_\tau\}_{\tau>0}$ is narrow. By the *a priori* estimates (Lemma 2.1), the family $\{\rho_\tau(t)\}_{\tau>0}$ can neither concentrate nor vanish and the family

$\{\rho_\tau(t)\}_{\tau>0}$ is tight. In the other hand by the estimate (3.5) the curves $\rho_\tau(t,\cdot)$ are $\frac{1}{2}$-Hölder continuous in time. Ascoli-Arzéla's theorems yield the relative compactness of the family $(\rho_\tau(t,\cdot))_{\tau>0}$.

Finally $\{\rho_\tau\}_{\tau>0}$ is relatively compact in $\mathcal{C}^0([0,T],L^1_{\rm weak}(\mathbb{R}^d))$ for any $T>0$. As a consequence, for any $T>0$, there exists a sub-sequence still denoted $(\rho_\tau)_{\tau>0}$, such that $(\rho_\tau)_{\tau>0}$ converges in $\mathcal{C}^0([0,T],L^1_{\rm weak}(\mathbb{R}^d))$ to a function $\rho$ when $\tau$ goes to 0.

**Step 3.- Approximate Euler-Lagrange equation in weak formulation**:
*Weak space derivative.-* Let $\zeta$ be a test function and $\nabla\zeta$ be a smooth vector field with compact support. Let us define $T_\varepsilon := \nabla\varphi_\varepsilon$ with $\varphi_\varepsilon(x) := \frac{|x|^2}{2} + \varepsilon\,\zeta$. For $\varepsilon$ small enough $T_\varepsilon$ is a $\mathcal{C}^1$-diffeomorphism and $\mathrm{Det}\,\nabla T_\varepsilon = \mathrm{Det}\,(\mathrm{Id} + \varepsilon\, D^2\zeta) > 0$. We define $\overline{\rho_\varepsilon}$ the push-forward perturbation of $\rho_\tau^{n+1}$ by $T_\varepsilon$:

$$\overline{\rho_\varepsilon} = T_\varepsilon \# \rho_\tau^{n+1}\ .$$

Changing variables and using (2.7), we have

$$\mathcal{G}[\overline{\rho_\varepsilon}](t) = \int_{\mathbb{R}^d} \log \frac{\rho_\tau^{n+1}(x)}{\mathrm{Det}(\mathrm{Id} + \varepsilon D^2\zeta)} \rho_\tau^{n+1}(x)\,\mathrm{d}x + \int_{\mathbb{R}^d} \left[\frac{1}{2}|x - \varepsilon\nabla\zeta(x)|^2\right] \rho_\tau^{n+1}(x)\,\mathrm{d}x$$
$$+ \int_{\mathbb{R}^d} \left[\frac{\chi}{2\,d\,\pi} \int_{\mathbb{R}^d} \log|x - y + \varepsilon[\nabla\zeta(x) - \nabla\zeta(y)]|\,\rho_\tau^{n+1}(y)\,\mathrm{d}y\right] \rho_\tau^{n+1}(x)\,\mathrm{d}x.$$

Alternatively we introduce the optimal map $\nabla\varphi^n$ which transports $\rho_\tau^n$ onto $\rho_\tau^{n+1}$. By (2.8)

$$d_W^2(\rho_\tau^n, \rho_\tau^{n+1}) = \int_{\mathbb{R}^d} |x - \nabla\varphi^n(x)|^2 \rho_\tau^n(x)\,\mathrm{d}x\ . \tag{3.6}$$

The map $[\mathrm{Id} + \varepsilon\,\nabla\zeta) \circ \nabla\varphi^n]$ transports $\rho_\tau^n$ on $\overline{\rho_\varepsilon}$. We do not know if this is the optimal map however by definition of the Wasserstein distance

$$d_W^2(\rho_\tau^n, \overline{\rho_\varepsilon}) \le \int_{\mathbb{R}^d} \left|x - (\mathrm{Id} + \varepsilon\,\nabla\zeta) \circ \nabla\varphi^n(x)\right|^2 \rho_\tau^n(x)\,\mathrm{d}x\ . \tag{3.7}$$

Using the minimising property of the scheme (3.1), and combining (3.6) and (3.7) we obtain

$$
\begin{aligned}
0 &\le \frac{1}{2\,\tau} d_W^2(\rho_\tau^n, \overline{\rho_\varepsilon}) + \mathcal{G}[\overline{\rho_\varepsilon}] - \frac{1}{2\,\tau} d_W^2(\rho_\tau^{n+1}, \rho_\tau^n) - \mathcal{G}[\rho_\tau^{n+1}] \\
&\le \frac{1}{2\,\tau} \int_{\mathbb{R}^d} \left( |x - \nabla\varphi^n(x) - \varepsilon\, \nabla\zeta \circ \nabla\varphi^n(x)|^2 - |x - \nabla\varphi^n(x)|^2 \right) \rho_\tau^n(x)\, \mathrm{d}x \\
&\quad + \frac{1}{2} \int_{\mathbb{R}^d} \left( \left( |x - \varepsilon \nabla\zeta(x)|^2 - |x|^2 \right) - \log\left[ \mathrm{Det}\left( \mathrm{Id} + \varepsilon\, D^2\zeta \right) \right] \right) \rho_\tau^{n+1}(x)\, \mathrm{d}x \\
&\quad + \frac{\chi}{2d\pi} \iint \left( \log\left| x - y + \varepsilon(\nabla\zeta(x) - \nabla\zeta(y)) \right| - \log|x-y| \right) \rho_\tau^{n+1}(x)\, \mathrm{d}x \rho_\tau^{n+1}(y)\, \mathrm{d}y.
\end{aligned}
$$

Dividing by $\varepsilon$ and letting $\varepsilon > 0$ going to zero we find

$$
\begin{aligned}
0 \le \frac{1}{\tau} &\int_{\mathbb{R}^d} \left\langle \nabla\varphi^n(x) - x, \nabla\zeta \circ \nabla\varphi^n(x) \right\rangle \rho_\tau^n(x)\, \mathrm{d}x \\
&+ \int_{\mathbb{R}^d} [-\Delta\zeta(x) - x \cdot \nabla\zeta(x) + \\
&+ \frac{\chi}{2d\pi} \int_{\mathbb{R}^d} \frac{[\nabla\zeta(x) - \nabla\zeta(y)] \cdot (x-y)}{|x-y|^2} \rho_\tau^{n+1}(y)\, \mathrm{d}y \Bigg] \rho_\tau^{n+1}(x)\, \mathrm{d}x.
\end{aligned}
$$

Since we can change $\varepsilon$ in $-\varepsilon$ we have in fact

$$
\begin{aligned}
&\frac{1}{\tau} \int_{\mathbb{R}^d} \left\langle \nabla\varphi^n(x) - x, \nabla\zeta(x) \circ \nabla\varphi^n(x) \right\rangle \rho_\tau^n(x)\, \mathrm{d}x \qquad (3.8) \\
&= \int_{\mathbb{R}^d} [\Delta\zeta(x) + x \cdot \nabla\zeta(x) - \\
&- \frac{\chi}{2\,d\,\pi} \int_{\mathbb{R}^d} \frac{[\nabla\zeta(x) - \nabla\zeta(y)] \cdot (x-y)}{|x-y|^2} \rho_\tau^{n+1}(y)\, \mathrm{d}y \Bigg] \rho_\tau^{n+1}(x)\, \mathrm{d}x.
\end{aligned}
$$

*Weak time derivative.-* Using the Taylor's expansion

$$
\zeta(\nabla\varphi^n(x)) - \zeta(x) = \left\langle \nabla\varphi^n(x) - x, \nabla\zeta \circ \nabla\varphi^n(x) \right\rangle + O\Big( |x - \nabla\varphi^n(x)|^2 \Big)
$$

we can recast the left-hand side of (3.8) as

$$
\begin{aligned}
&\frac{1}{\tau} \left( \int_{\mathbb{R}^d} \zeta \circ \nabla\varphi^n(x) \rho_\tau^n(x)\, \mathrm{d}x - \int_{\mathbb{R}^d} \zeta(x) \rho_\tau^n(x)\, \mathrm{d}x \right) + \\
&\qquad\qquad + O\left( \frac{1}{\tau} \int_{\mathbb{R}^d} |x - \nabla\varphi^n(x)|^2\, \rho_\tau^n(t)\, \mathrm{d}x \right) \\
&= \frac{1}{\tau} \left( \int_{\mathbb{R}^d} \zeta(x) \rho_\tau^{n+1}(x)\, \mathrm{d}x - \int_{\mathbb{R}^d} \zeta(x) \rho_\tau^n(x)\, \mathrm{d}x \right) + O\left( \frac{1}{\tau} d_W^2\left( \rho_\tau^n, \rho_\tau^{n+1} \right) \right).
\end{aligned}
$$

We multiply (3.8) by $\tau$ and eventually obtain

$$\int_{\mathbb{R}^d} \zeta(x)\left[\rho_\tau^{n+1}(x)-\rho_\tau^n(x)\right]\,\mathrm{d}x + O\big(d_W^2\left(\rho_\tau^n,\rho_\tau^{n+1}\right)\big) \tag{3.9}$$
$$=\tau\int_{\mathbb{R}^d}[\Delta\zeta(x)+x\cdot\nabla\zeta(x)-$$
$$-\frac{\chi}{2\,d\,\pi}\int_{\mathbb{R}^d}\frac{[\nabla\zeta(x)-\nabla\zeta(y)]\cdot(x-y)}{|x-y|^2}\rho_\tau^{n+1}(y)\,\mathrm{d}y\Big]\,\rho_\tau^{n+1}(x)\mathrm{d}x.$$

let $0\le t_1<t_2$ be fixed times, $m=[t_1/\tau]+1$ and $n=[t_2/\tau]$. By summing equation (3.9) we have thanks to the total square estimate (3.4),

$$\int_{\mathbb{R}^d}\zeta(x)[\rho_\tau^n(x)-\rho_\tau^m(x)]\,\mathrm{d}x+O(\tau)=\sum_{k=m}^{n-1}\tau\int_{\mathbb{R}^d}[\Delta\zeta(x)+x\cdot\nabla\zeta(x)]\,\rho_\tau^{k+1}(x)\,\mathrm{d}x$$
$$-\frac{\chi}{2\,d\,\pi}\sum_{k=m}^{n-1}\tau\iint_{\mathbb{R}^d\times\mathbb{R}^d}\frac{[\nabla\zeta(x)-\nabla\zeta(y)]\cdot(x-y)}{|x-y|^2}\rho_\tau^{k+1}(y)\,\mathrm{d}y\rho_\tau^{k+1}(x)\,\mathrm{d}x, \tag{3.10}$$

On the other hand, we can split

$$\int_{\mathbb{R}^d}\zeta(x)\left[\rho_\tau(t_2,x)-\rho_\tau(t_1,x)\right]\,\mathrm{d}x=\int_{\mathbb{R}^d}\zeta(x)\left[\rho_\tau(t_2,x)-\rho_\tau^n(x)\right]\,\mathrm{d}x$$
$$+\int_{\mathbb{R}^d}\zeta(x)\left[\rho_\tau^n(x)-\rho_\tau^m(x)\right]\,\mathrm{d}x+\int_{\mathbb{R}^d}\zeta(x)\left[\rho_\tau^m(x)-\rho_\tau(t_1,x)\right]\,\mathrm{d}x\,.$$

By Lemma 2.4, we control the bordering averages,

$$\begin{aligned}\int_{\mathbb{R}^d}\zeta(x)\left|\rho_\tau^m(x)-\rho_\tau(t_1,x)\right|\,\mathrm{d}x &\le C\,d_W(\rho_\tau^m,\rho_\tau(t_1,x))\\ &\le C\,d_W(\rho_\tau^m,\rho_\tau^{m-1})\le O(\tau^{1/2})\,.\end{aligned}$$

In addition, integrating in time Lemma 2.4 implies for all $k$ in $[[m,n-1]]$

$$\tau\int_{\mathbb{R}^d}\psi(x)\rho_\tau^{k+1}(x)dx=\int_{k\tau}^{(k+1)\tau}\int_{\mathbb{R}^d}\psi(x)\rho_\tau(s,x)\,\mathrm{d}s+O(\tau\;d_W(\rho_\tau^k,\rho_\tau^{k+1}))\,,$$

where $\psi$ denotes any bounded test function. Hence we can transform the discrete in time sum (3.10) into a continuous time integration. Finally, the test contributions are bounded in $L^\infty((0,T)\times\mathbb{R}^d)$ so that the bordering time integrands are negligible,

$$\int_{t_1}^{m\tau}\int_{\mathbb{R}^d}[\Delta\zeta(x)+x\cdot\nabla\zeta(x)]\,\rho_\tau(s,x)\mathrm{d}x\,\mathrm{d}s$$
$$-\frac{\chi}{2\,d\,\pi}\int_{t_1}^{m\tau}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\frac{[\nabla\zeta(x)-\nabla\zeta(y)]\cdot(x-y)}{|x-y|^2}\rho_\tau(s,y)\rho_\tau(s,x)\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}s=O(\tau)\,.$$

Collecting all the terms we end up with

$$\int_{\mathbb{R}^d}\zeta(x)\,[\rho_\tau(t_2,x)-\rho_\tau(t_1,x)]\;\mathrm{d}x=\int_{t_1}^{t_2}\int_{\mathbb{R}^d}[\Delta\zeta(x)+x\cdot\nabla\zeta(x)]\,\rho_\tau(s,x)\;\mathrm{d}x\;\mathrm{d}s \tag{3.11}$$

$$-\frac{\chi}{2\,d\,\pi}\int_{t_1}^{t_2}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\frac{[\nabla\zeta(x)-\nabla\zeta(y)]\cdot(x-y)}{|x-y|^2}\rho_\tau(s,y)\;\mathrm{d}y\rho_\tau(s,x)\;\mathrm{d}x\;\mathrm{d}s+O(\tau^{1/2}).$$

**Step 4.- Passing to the limit**: The relative compactness of the family of curves $\{\rho_\tau\}_{\tau>0}$ in $\mathcal{C}^0([0,T],L^1_{\rm weak}(\mathbb{R}^d))$ for any $T>0$, allows to pass to the limit in the linear parts of (3.11) because $\nabla\zeta$ and $\Delta\zeta$ are bounded in $L^\infty(\mathbb{R}^d)$. The quadratic last term of (3.11), coming from the concave interaction energy, is more difficult to handle with. Actually, we shall make another use of the doubling of variables' trick: $\rho_\tau\otimes\rho_\tau\rightharpoonup\rho\otimes\rho$ thanks to the equi-integrability, obtained in Lemma 2.3. In addition, recall that the convergence is uniform in time thanks to Ascoli-Arzéla's theorems, and that equi-integrability bounds are also uniform with respect to $\tau$. We can thus pass to the limit when $\tau$ goes to zero in (3.11) to obtain for any $t_1$, $t_2$

$$\int_{\mathbb{R}^d}\zeta(x)\,[\rho(t_2,x)-\rho(t_1,x)]\;\mathrm{d}x=\int_{t_1}^{t_2}\int_{\mathbb{R}^d}[\Delta\zeta(x)+x\cdot\nabla\zeta(x)]\,\rho(s,x)\;\mathrm{d}x\;\mathrm{d}s$$
$$-\frac{\chi}{2\,d\,\pi}\int_{t_1}^{t_2}\int_{\mathbb{R}^d\times\mathbb{R}^d}\frac{[\nabla\zeta(x)-\nabla\zeta(y)]\cdot(x-y)}{|x-y|^2}\rho(s,y)\rho(s,x)\;\mathrm{d}y\;\mathrm{d}x\;\mathrm{d}s\;,$$

which is a formulation of the weak solution as defined in (2.5). □

**Remark 3.5** (Original variables)**.** *This theorem is true in original variables* (1.1) *locally in time, when we consider the free energy* (1.4) *with minor changes, see also* [29]*. However, the long time asymptotics are better understood in scaled variables (see the numerical results, Section* 4.2.1*).*

**Remark 3.6** (Comparison to Literature)**.** *Previous results of convergence of numerical schemes have been presented in* [26]*. Theorem* 3.4 *shows the convergence of the JKO scheme for all values $\chi<\chi_c$, where $\chi_c=2d^2\pi$ is known to be optimal, whereas in* [26] *the threshold is non-optimal and depends upon the mesh grid regularity. Furthermore, because the free energy is not really used in* [26]*, there is currently no hope to get the optimal threshold in his method. However, the 'minimising movement' scheme is much more involved numerically in other dimension than one, at least as a direct implicit discretisation detailed in Section* 4.2*. Alternative ideas based*

*on minimisation procedures may be of application in more dimensions. Beyond the optimal threshold, there is another result arguing on this alternative scheme's behalf. Namely for a given $\tau$ the solution $\rho_\tau^n$ is shown to converge towards the unique stationary state, with explicit exponential rate in the one-dimensional case, see next section Proposition* 4.1.

**Remark 3.7** (Generalizations)**.** *The ongoing work extends to non-linear diffusions, under suitable assumptions, $\Delta f(n)$ without deeper difficulty. The main points is that the free energy should be bounded from below – this results from a balance between diffusion at high density level (diffusion must be super-linear at infinity above the optimal threshold) and at low density level (basically $f(u) \gtrsim u^\alpha$ where $\alpha > \max(1/2, 1-2/d)$ preventing mass to escape too fast towards infinity), see* [13] *and* [17].

## 4. One-dimensional Case

In the case of the real line, consider $\mu$ and $\nu$ two absolutely continuous measures with respect to the Lebesgue measure, of respective densities $f$ and $g$, and of cumulative distribution functions $F$ and $G$. As the cumulative distribution function is non-decreasing we can define the pseudo-inverse function by

$$V(z) = F^{-1}(z) := \inf\{x \,:\, F(x) \geq z\} \,.$$

By Brenier's theorem and the definition of the image measure (2.7), we have

$$F \circ \varphi'(x) = \int_0^{\varphi(x)} f(y)\,\mathrm{d}y = \int_0^x g(y)\,\mathrm{d}y = G(x) \,.$$

Hence, the transport map can be stated explicitly $\varphi' = F^{-1} \circ G$ and the Wasserstein distance can be expressed in the following more tractable way

$$d_W^2(\mu,\nu) = \int_0^1 |F^{-1}(w) - G^{-1}(w)|^2 \,\mathrm{d}w \,. \tag{4.1}$$

In fact, $F^{-1} \circ G$ is the optimal map for all convex costs in one dimension [43, Theorem 2.18].

This expression of the one dimensional Wasserstein distance has been used for non-linear diffusions and non-linear non-local friction equations in granular media [20, 31, 16] to analyse the long-time asymptotics and the contraction properties with respect to Wasserstein distances of those equations. Moreover, these ideas have been used in [27, 28] for numerical purposes. Explicit in time numerical schemes for the equations of the inverse distribution function are proposed keeping the contraction of the Wasserstein distance at the discrete level. Here, we prefer to solve it by

an implicit in time Euler scheme since it coincides with the JKO scheme through the representation (4.1) and moreover, the contraction property of the Wasserstein distance is not true due to the lack of convexity of the functional $\mathcal{G}$.

More precisely, let $F_n$ and $F_{n+1}$ be the cumulative distribution functions associated respectively to $\rho_\tau^n$ and $\rho_\tau^{n+1}$. By the expression (4.1) of the Wasserstein distance on the real line, the scheme (3.1) can be rewritten in terms of $V_n = F_n^{-1}$ and $V_{n+1} = F_{n+1}^{-1}$ as the gradient flow of the inverse distribution function subject to $L^2-$metric structure:

$$V_{n+1} = \inf_W \left[ \mathcal{G}[W] + \frac{1}{2\,\tau} \|W - V_n\|^2_{L^2(0,1)} \right] .$$

Here the metric is Euclidean, hence the Euler-Lagrange equation associated to this minimisation problem is

$$\frac{V_{n+1} - V_n}{\tau} = -\nabla \mathcal{G}[V_{n+1}] \,,$$

where $\nabla$ is the usual gradient operator in $L^2(\mathbb{R})$. This Euler-Lagrange equation can be rewritten

$$-\frac{V_{n+1}(w) - V_n(w)}{\tau} = \frac{\partial}{\partial w} \left[ \left( \frac{\partial V_{n+1}(w)}{\partial w} \right)^{-1} \right] + V_{n+1}(w) + \frac{\chi}{\pi} H[V_{n+1}] \quad (4.2)$$

where $H$ corresponds to the Hilbert transform $\mathcal{H}[\rho] = -\nabla v$, and is defined by

$$H[V](w) := \frac{1}{\pi} \lim_{\varepsilon \to 0} \int_{|V(w)-V(z)| \geq \varepsilon} \frac{1}{V(w) - V(z)} \, \mathrm{d}z \,.$$

For sake of simplicity we assume that the space step is constant, equal to $h$. If we set $V_n^i := V_n(ih)$, for any $i = 0 \cdots N$, and $Nh = 1$, the finite difference discretisation in space of (4.2) is the following implicit Euler scheme in rescaled variables,

$$-\frac{V_{n+1}^i - V_n^i}{\tau} = \frac{1}{V_{n+1}^{i+1} - V_{n+1}^i} - \frac{1}{V_{n+1}^i - V_{n+1}^{i-1}} + V_{n+1}^i + \frac{\chi}{\pi} \lim_{\varepsilon \to 0} \sum_{j : |V_{n+1}^i - V_{n+1}^j| \geq \varepsilon} \frac{1}{V_{n+1}^i - V_{n+1}^j} \,. \quad (4.3)$$

with initial condition $V_0 = \rho_0$. We impose Neumann boundary conditions in the $\rho-$problem (2.3), *i.e.* for any $n$, $\frac{1}{V_n^N - V_n^{N-1}} = 0$ and $\frac{1}{V_n^1 - V_n^0} = 0$, so that the 'centre of mass' is conserved:

$$\forall n \ \sum_{i=0}^{N} V_n^i = 0.$$

The solution at each time step of the non-linear system of equations is obtained by an iterative Newton-Raphson procedure.

4.1. **Exponential Convergence in 1-D.** Here comes the second main improvement and motivation for this numerical scheme. In addition to convergence as the time step goes to zero, we are able to show that for a fixed $\tau > 0$, the discrete solution converges to a unique steady state as time goes to infinity.

**Proposition 4.1** (Convergence in the sub-critical case)**.** *Assume $\chi < \chi_c$. Then the solution of the numerical scheme* (4.3) *converges to the (unique) steady-state of the problem with exponential rate.*

*Proof.* First we need the following two characterizations of the (unique) equilibrium state. The uniqueness will in fact follow from the convergence proof, as we shall see later (Remark 4.2). The discrete function $(U^i)$ is an equilibrium if and only if

$$\forall i \quad 0 = \frac{1}{U^{i+1} - U^i} - \frac{1}{U^i - U^{i-1}} + U^i + \frac{\chi}{\pi} \sum_{j \neq i} h \frac{1}{U^i - U^j}, \tag{4.4}$$

or equivalently

$$\forall k \quad (U^{k+1} - U^k) \left\{ \frac{\chi}{\pi} \sum_{j=0}^{k} \sum_{i=k+1}^{N} h \frac{1}{U^i - U^j} - \sum_{i=0}^{k} U^i \right\} = 1. \tag{4.5}$$

To see that (4.4) and (4.5) are equivalent, rewrite the latter as

$$\forall k \quad \frac{1}{U^{k+1} - U^k} = \frac{\chi}{\pi} \sum_{j=0}^{k} \sum_{i=k+1}^{N} h \frac{1}{U^i - U^j} - \sum_{i=0}^{k} U^i,$$

then 'derive' the ongoing expression in a discrete way,

$$\begin{aligned} &\frac{1}{U^{k+1} - U^k} - \frac{1}{U^k - U^{k-1}} = \\ &= \frac{\chi}{\pi} \sum_{j=0}^{k} \sum_{i=k+1}^{N} h \frac{1}{U^i - U^j} - \frac{\chi}{\pi} \sum_{j=0}^{k-1} \sum_{i=k}^{N} h \frac{1}{U^i - U^j} - U^k \\ &= \frac{\chi}{\pi} \sum_{i=k+1}^{N} h \frac{1}{U^i - U^k} - \frac{\chi}{\pi} \sum_{j=0}^{k-1} h \frac{1}{U^k - U^j} - U^k. \end{aligned}$$

We shall split the proof into an analytic part (existence of the stationary state) and an algebraic computation (convergence estimate). Because the

existence proof shares many similarities with previous estimates (*e.g.* existence of a minimiser to the corresponding free energy, see Lemma 3.1), we leave it to the reader.

We proceed as computing the time evolution of the $L^2-$distance between $V_n$ and the stationary state $U$.

$$\begin{aligned}
&\frac{1}{2\tau}\Big(\|V_{n+1}-U\|^2-\|V_n-U\|^2\Big)=\\
&\quad=\quad\frac{1}{2\tau}\sum_i h(V^i_{n+1}-V^i_n)(V^i_{n+1}+V^i_n-2U^i)\\
&\quad=\quad\sum_i h\frac{V^i_{n+1}-V^i_n}{\tau}(V^i_{n+1}-U^i)-\frac{1}{2\tau}\sum_i h(V^i_{n+1}-V^i_n)^2.
\end{aligned}$$

We then input the evolution equation for $V_{n+1}-V_n$, and obtain thanks to (4.4),

$$\begin{aligned}
&\frac{1}{2\tau}\Big(\|V_{n+1}-U\|^2-\|V_n-U\|^2\Big)\le\\
&\quad\le\quad-\sum_i h\Big(\frac{1}{V^{i+1}-V^i}-\frac{1}{V^i-V^{i-1}}-\frac{1}{U^{i+1}-U^i}\\
&\qquad\qquad+\frac{1}{U^i-U^{i-1}}+V^i-U^i+\frac{\chi}{\pi}\sum_{j\neq i}h\frac{1}{V^i-V^j}\\
&\qquad\qquad-\frac{\chi}{\pi}\sum_{j\neq i}h\frac{1}{U^i-U^j}\Big)(V^i-U^i)\\
&\quad=\quad A_n+B_n+C_n,
\end{aligned}$$

where $V$ stands for $V_{n+1}$ without any ambiguity. We integrate by part the first (diffusion) contribution,

$$\begin{aligned}
A=&\\
&=\quad-\sum_i h\Big(\frac{1}{V^{i+1}-V^i}-\frac{1}{V^i-V^{i-1}}-\frac{1}{U^{i+1}-U^i}+\frac{1}{U^i-U^{i-1}}\Big)(V^i-U^i)\\
&=\quad\sum_i h\Big(\frac{1}{V^{i+1}-V^i}-\frac{1}{U^{i+1}-U^i}\Big)(V^{i+1}-U^{i+1}-V^i+U^i).
\end{aligned}$$

We have carefully used the boundary conditions $\frac{1}{V^N-V^{N-1}}=0$ and $\frac{1}{V^1-V^0}=0$. We can rewrite $A$ using zero-homogeinity of the last expression, namely

$$A=\sum_i h\gamma\Big(\frac{V^{i+1}-V^i}{U^{i+1}-U^i}\Big),$$

where $\gamma(\lambda) = 2 - \lambda - \lambda^{-1}$ is concave and non-positive. The second contribution coming from variables rescaling is obvious but crucial, namely

$$B = -\sum_i h(V^i - U^i)^2 = -\|V_{n+1} - U\|^2.$$

The last (interaction) contribution is more tricky to handle with, and involves variables doubling, as it is known from granular media, see [20]. We have

$$\begin{aligned} C &= -\frac{\chi}{\pi}\sum_i h\Big(\sum_{j\neq i} h\frac{1}{V^i - V^j} - \sum_{j\neq i} h\frac{1}{U^i - U^j}\Big)(V^i - U^i) \\ &= -\frac{\chi}{2\pi}\sum\sum_{i,j,\, i\neq j} h^2\Big(\frac{1}{V^i - V^j} - \frac{1}{U^i - U^j}\Big)(V^i - V^j - U^i + U^j) \\ &= -\frac{\chi}{2\pi}\sum\sum_{i,j,\, i\neq j} h^2\gamma\Big(\frac{V^i - V^j}{U^i - U^j}\Big). \end{aligned}$$

Notice that the above expression is symmetric between the two possible choices $i < j$ and $j < i$. We shall also make use of the concavity property of $\gamma$ as following,

$$\begin{aligned} C &= -2\frac{\chi}{2\pi}\sum\sum_{j<i} h^2\gamma\Big(\sum_{j\leq k<i}\frac{V^{k+1} - V^k}{U^{k+1} - U^k}\cdot\frac{U^{k+1} - U^k}{U^i - U^j}\Big) \\ &\leq -\frac{\chi}{\pi}\sum\sum_{i<j} h^2\sum_{j\leq k<i}\gamma\Big(\frac{V^{k+1} - V^k}{U^{k+1} - U^k}\Big)\frac{U^{k+1} - U^k}{U^i - U^j} \\ &= -\frac{\chi}{\pi}\sum_k h\gamma\Big(\frac{V^{k+1} - V^k}{U^{k+1} - U^k}\Big)(U^{k+1} - U^k)\sum_{j=0}^{k}\sum_{i=k+1}^{N} h\frac{1}{U^i - U^j} \\ &\leq -\sum_k h\gamma\Big(\frac{V^{k+1} - V^k}{U^{k+1} - U^k}\Big)(U^{k+1} - U^k)\left\{\frac{\chi}{\pi}\sum_{j=0}^{k}\sum_{i=k+1}^{N} h\frac{1}{U^i - U^j} - \sum_{i=0}^{k} U^i\right\}, \end{aligned}$$

where we have used the fact that $D_k := \sum_{i=0}^{k} U^i$ is a non-positive quantity for all $k$.

Let us prove this last claim. Indeed, $D_{k+1} - D_{k-1} - 2\,D_k = U^{k+1} - U^k \geq 0$. Hence, $D_{k+1} - D_k \geq D_{k-1} - D_k$. Since $D_N = 0$ and $D_0 = U^0 \leq 0$, there exists $k_0$ such that $(D_k)_{k\in[[1,k_0]]}$ is non-decreasing and $(D_k)_{k\in[[k_0-1,N]]}$ is non-increasing. As a consequence $D_k \leq D_0 \leq 0$ for any $k \in [[1, k_0]]$ and $D_k \leq D_N = 0$ for any $k \in [[k_0 + 1, N]]$ which proves the assertion.

At this stage we bring in the alternative representation of the stationary solution (4.5), so that $A + C \le 0$. As a consequence we obtain

$$\frac{1}{2\tau}\Big(\|V_{n+1} - U\|^2 - \|V_n - U\|^2\Big) \le -\|V_{n+1} - U\|^2. \tag{4.6}$$

We finally get the exponential convergence rate,

$$\|V_n - U\|^2 \le \Big(\frac{1}{1+2\tau}\Big)^n \|V^0 - U\|^2.$$

If $\tau$ is small, we can thus approximate $\log(1+2\tau) \approx 2\tau$ and $\Big(\frac{1}{1+2\tau}\Big)^n \approx \exp(-2n\tau) \approx \exp(-2t)$. Thus, the bound on the rate, we find, does not depend on the parameter $\chi < \chi_c$. □

**Remark 4.2** (Uniqueness of Stationary Solution)**.** *We can deduce* a posteriori *the uniqueness of the equilibrium. As a matter of fact let consider another equilibrium state $\tilde U$ and set $V_{n+1} = V_n = \tilde U$ in the above computations. We eventually obtain $\|\tilde U - U\| \le 0$ from* (4.6)*, which proves the uniqueness.*

**Remark 4.3** (Original Variables)**.** *When coming back in the continuous setting to the original variables $n(t,x)$ by* (2.2)*, we are not able to show even that*

$$n(t,x) = n^\infty(t,x) + O_W(1),$$

*where $n^\infty$ is the dilatation of the stationary state,*

$$n^\infty(t,x) = \frac{1}{1+2t} U\left(\frac{x}{\sqrt{1+2t}}\right),$$

*and $O_W$ means the infinitesimal in the $d_W$ sense. The reason is that the found estimate on the speed of convergence does not depend on the reduced parameter $\chi$, but only on the variables rescaling, and the change of variables restores back exactly the factor $e^t$ due to the scaling properties of $d_W$* [43, 21]*. This result should be improved as seen from the numerical experiments below and it is an open problem how to get a faster speed of convergence in the scaled equation leading to a polynomial decay in original variables. In fact, we conjecture that if we fix the center of mass then the rate on convergence of the solution to the stationary solution in rescaled variables is of order $\mathrm{e}^{-2t}$. This fact coincides with other situations as in nonlinear diffusions in which fixing certain invariants of the equation improves the rate of convergence* [23]*. Certainly the situation is close to the heat equation for small mass solutions* [25].

In the next subsections, we will show some numerical experiments for the PKS equation using the scheme (4.3). We begin with the sub-critical case $\chi = \pi$ (remind that in one dimension the critical parameter is $\chi_c = 2\,\pi$) for the not rescaled (Section 4.2.1) as opposed to the rescaled system (Section 4.2.2). We next approach the critical parameter plugging $\chi = 1.8\pi$, and initialise the scheme with a two-peaks density $n_0$ (*resp.* Sections 4.2.3 and 4.2.4). We then investigate the upper-critical case starting with respectively a single peak (Section 4.3.1), two symmetric peaks attracting each other (Section 4.3.2) and two asymmetric peaks (Section 4.4).

In the following, we assume a uniform in space discretisation $w_i = ih$, $i = 0 \dots N$, and $Nh = 1$.

## 4.2. **Sub-critical case.**

4.2.1. *Not rescaled case.* Starting with the centred initial data,

$$V_0^i = 2\,\frac{w_i - 0.5}{\left[(w_i + 0.01)\,(1.01 - w_i)\right]^{1/4}}\,,$$

corresponding to a compact supported density $n_0$, we numerically solve the PKS equation on the time interval $[0, 400]$ with $\chi = \pi$. Figures 4.1 and 4.2 show the evolution of the solution both for the density (Fig. 4.2) and its inverse distribution function (Fig. 4.1).

Observe in Figure 4.1 that the branches of the inverse cumulative function $V$ goes eventually to $\pm\infty$. This is expected because the support of the cell density spreads as time goes on, and therefore the distribution tails are wider. Remind that in the sub-critical regime, the diffusion process dominates. The scheme captures well the collapse down to zero of the cell density and the spreading of the solution. Interestingly, this scheme handles easily with moving density's support (note that finite speed of propagation is a numerical artifact) whereas the reference domain $[0, 1]$ is fixed because we deal with probability densities (mass is conserved).

Moreover, the spreading towards zero seems to be polynomial from Figure 4.3 showing the evolution of the $L^2$-norm of the cell density in log-log scale. The entropy decay is plotted in Figure 4.4.

4.2.2. *Rescaled variables:* $\chi = \pi$. Given the compactly supported initial data,

$$V_0^i = 2\,\frac{w_i - 0.5}{\left[(w_i + 0.01)\,(1.01 - w_i)\right]^{1/4}}\,,$$

we numerically solve the PKS equation in rescaled variables on the time interval $[0, 5]$ with $\chi = \pi$ (corresponding results are shown in figures 4.5 and 4.6, 4.7 and 4.8).

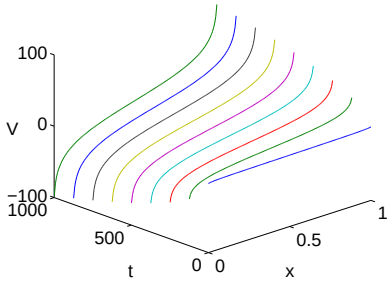


FIGURE 4.1. Inverse cumulative distribution function for $\chi = \pi$. Note that the initial data seems to be flat relatively to the very large scale on the $V-$axis, as opposed to Figure 4.5.

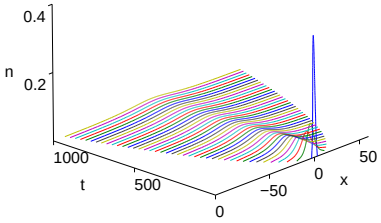


FIGURE 4.2. Cell density $n$ as time evolves, obtained from its inverse cumulative distribution function. Accordingly to Figure 4.1, the space scale is also very large, and therefore the density seems highly concentrated at $t = 0$.

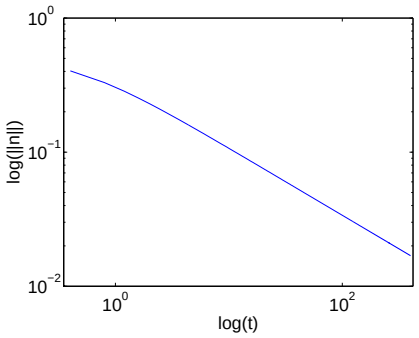


FIGURE 4.3. $L^2$-norm's evolution for the cell density $n$, in a $\log - \log$ scale. The decay appears to be polynomial.

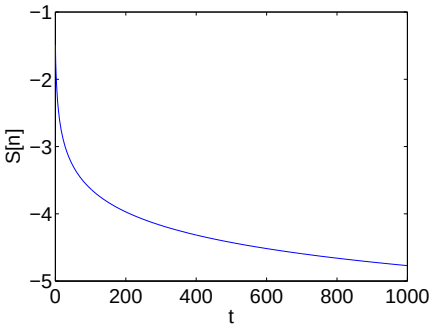


FIGURE 4.4. Evolution of the entropy $\mathcal{S}[n]$ showing slow decay.

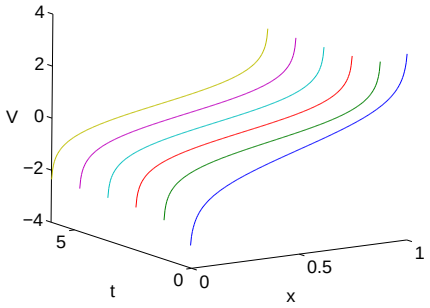


FIGURE 4.5. Fast convergence towards the stationary solution for $\chi = \pi$ and rescaled variables.

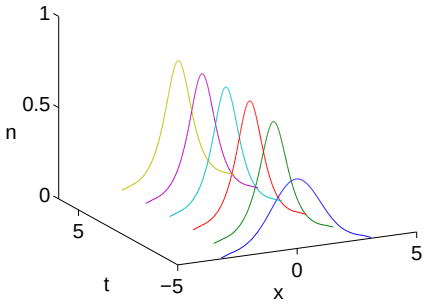


FIGURE 4.6. Evolution of the corresponding cell density $n$.

Contrary to previous Section 4.2.1, we observe an exponential convergence towards the stationary solution (see Figures 4.5 and 4.7). When computing the Wasserstein distance between the density at time $t$ and the expected stationary solution (last computed time), we find out that the convergence is faster than $e^{-t}$ (Figure 4.7) obtained in Proposition 4.1. This confirms the open problem of trying to find a better decay rate in scaled variables that will eventually lead to a polynomial decay rate towards self-similarity in original variables for subcritical masses.

4.2.3. *Rescaled variables:* $\chi = 1.8\pi$. Given the compactly supported initial data,

$$V_0^i = 2\,\frac{w_i - 0.5}{[(w_i + 0.01)\,(1.01 - w_i)]^{1/4}}\,,$$

we numerically solve the PKS equation in rescaled variables on the time interval $[0, 5.5]$ with $\chi = 1.8\pi$. Figures 4.9 and 4.10 show the evolution of the solution.

The initial data is the same as in Section 4.2.2 but $\chi$ is closer to the critical parameter $\chi_c$. The solution again converges exponentially to the stationary solution (see Figures 4.9 and 4.10). According to Proposition 4.1, the rate of convergence is at least the same as in Section 4.2.2, compare Fig. 4.11 to Fig. 4.7 and the slope are the same (of order $\mathrm{e}^{-2t}$). On the other hand, the equilibrium state is more concentrated (Fig. 4.10), corresponding to a flat plateau in Fig. 4.9, as we expect it converges to a Dirac mass.

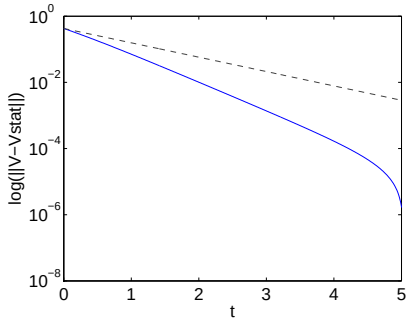


FIGURE 4.7. Wasserstein distance between the density at time $t$ and the final computed density assumed to be almost the stationary solution. Dash-line: decay rate proven in Proposition 4.1.

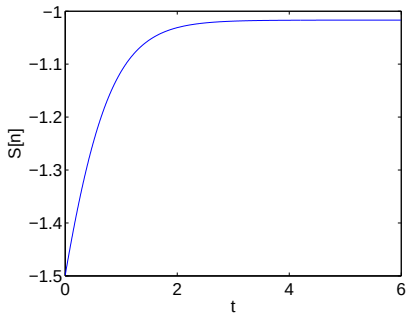


FIGURE 4.8. Evolution of the entropy.

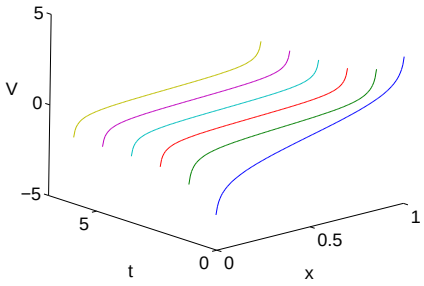


FIGURE 4.9. Cumulative distribution function $V$ for the sub-critical case $\chi = 1.8\pi < \chi_c$.

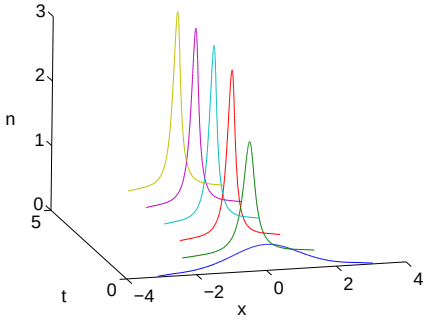


FIGURE 4.10. Evolution of the cell density $n$.

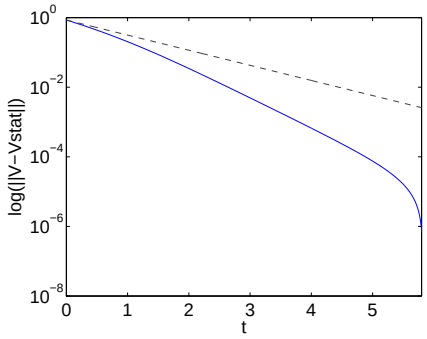


FIGURE 4.11. Wasserstein distance between the density at time $t$ and the final computed density assumed to be almost the stationary solution. Dash-line: decay rate proven in Proposition 4.1.

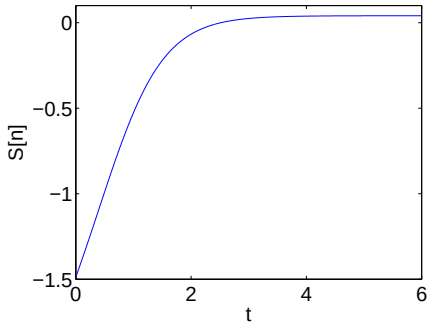


FIGURE 4.12. Evolution of the entropy.

4.2.4. *Two peaks initial data.* Initialise with the centred cumulative distribution function,

$$V_0^i = \frac{\exp\left[10\ (w_i - 0.5)\right] - 1}{\left[(w_i + 0.01)\,(1.01 - w_i)\right]^{1/4}}\ ,$$

corresponding to a two-peaks like density with compact support. We numerically solve the PKS equation in rescaled variables on the time interval $[0, 5]$ with $\chi = \pi$.

Whereas the parameter $\chi$ is the same as in Section 4.2.2, the initial data is qualitatively different. The two peaks diffuse, eventually merging and finally converging to the stationary solution with exponential speed (see Figures 4.13 and 4.14). Let us finally mention that the numerical scheme does not preserve the exact value of the critical mass. The scheme does not preserve the law of evolution of the second moment. However, with the same initial data as in the first two subsections, the numerical critical mass is situated between $1.973\pi$ and $1.974\pi$.

## 4.3. **Super-critical case.**

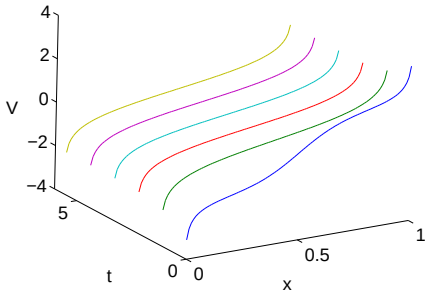


FIGURE 4.13. Cumulative distribution function $V$ for $\chi = \pi$ and a two-peaks initial condition.

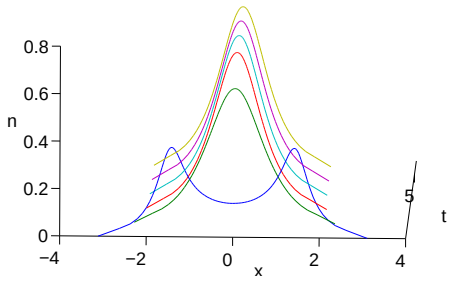


FIGURE 4.14. Cell density $n$.

4.3.1. *Single peak initial data.* Given the compactly supported initial data,

$$V_0^i = 2\,\frac{w_i - 0.5}{[(w_i + 0.01)\,(1.01 - w_i)]^{1/4}}\ ,$$

we numerically solve the PKS equation in original variables on the time interval $[0, 0.32]$ with $\chi = (5/2)\,\pi$. Note that in the upper-critical case, the variables' rescaling seems to play no role.

The solution blows-up in finite time (either a flat portion or a highly concentrated region appears, *resp.* Figure 4.15 and Figure 4.16). Visualized in Wasserstein distance (namely the square root of the second momentum), the convergence to the Dirac mass located at zero seems to be linear in time (see Figure 4.17) as it should be from the theoretical viewpoint. However the computed distance does not reach zero in finite time. This is not surprising, because when blow-up occurs, part of the mass is still away from the blow-up point (here, zero). In order to see some vanishing distance, one can renormalize the process in the following away: localize the Wasserstein distance in the transport variable ($L^2-$distance for the cumulative distribution function), to capture only the final plateau. This plateau is *a priorily* known from the beginning because it is entirely determined by the ratio $\chi_c/\chi$. However this does not provide any new insight of what happens after blow-up, and it is known from theoretical works that the behavior highly depends upon the regularization procedure [41, 42].

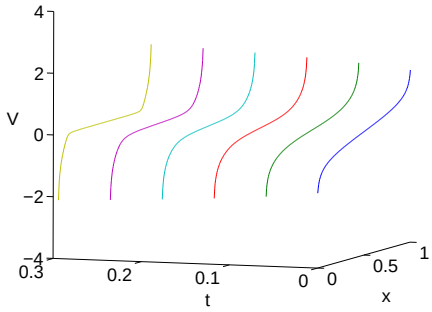


FIGURE 4.15. Cumulative distribution function $V$ for $\chi > \chi_c$. The solution blows-up exhibiting a plateau in finite time.

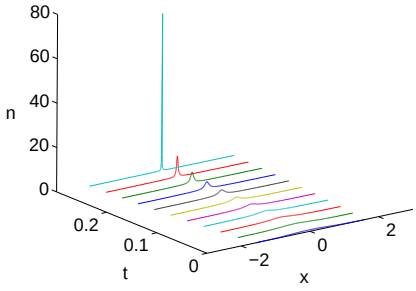


FIGURE 4.16. Cell density $n$. We observe blow-up in finite time.

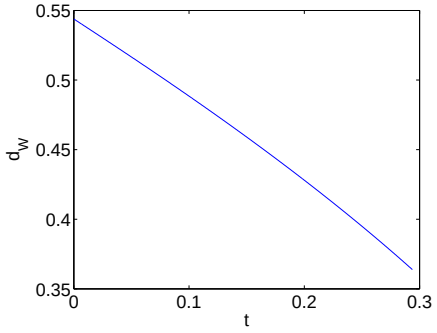


FIGURE 4.17. Wasserstein distance to the Dirac mass at zero $\delta_0$. Blow-up occurs previously, and part of the mass has not yet reached zero at this time.

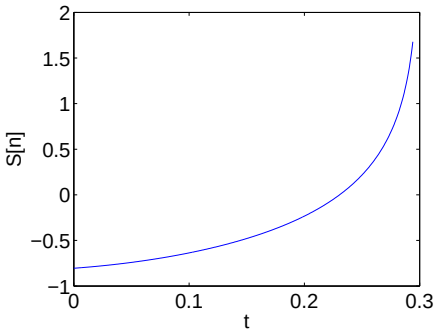


FIGURE 4.18. The variation of the entropy $\mathcal{S}[n]$ seems to blow-up.

Interestingly, numerics are able to track the blow-up phenomenon quite precisely, without mesh refinement. Indeed, if the space step is even uniform, the number of space points at the density level adapt to the highly

concentrated (blow-up) regions, corresponding to plateaus (compare Figure 4.15 and Figure 4.16). This is the counterpart of the 'moving support' observed in Section 4.2.1.

4.3.2. *Two symmetric peaks: case* $\chi = 3\,\pi$. Given the compactly supported initial data,

$$V_0^i = \frac{\exp\left[10\ (w_i - 0.5)\right] - 1}{\left[(w_i + 0.01)\,(1.01 - w_i)\right]^{1/4}}\ ,$$

we numerically solve the PKS equation in original variables on the time interval $[0, 1.3]$ with $\chi = 3\,\pi$.

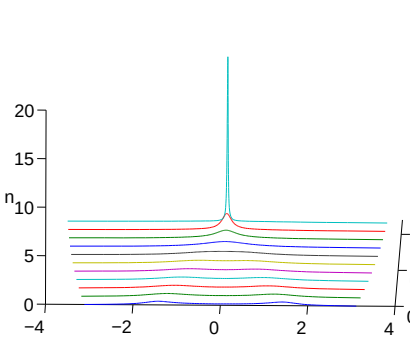


FIGURE 4.19. Cell density $n$ for $\chi = 3\pi$ and two initial peaks.

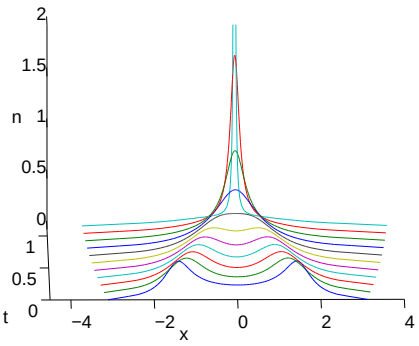


FIGURE 4.20. Zoom of Figure 4.19. Because both two peaks do not contain enough mass to blow-up far from each other, they first merge, then the solution blows-up.

The factor $\chi$ is super-critical but is less than $2\,\chi_c$. Then, according to the conjectures in [40, 41, 42] there should be only one blow-up point. The density first diffuses (see Figure 4.20) and then concentrates in a delta dirac (see Figures 4.19 and 4.21).

4.3.3. *Two symmetric peaks: case* $\chi = 5\,\pi$. Starting with the centered cumulative distribution function,

$$V_0^i = \frac{\exp\left[10\ (w_i - 0.5)\right] - 1}{\left[(w_i + 0.01)\,(1.01 - w_i)\right]^{1/4}}\ ,$$

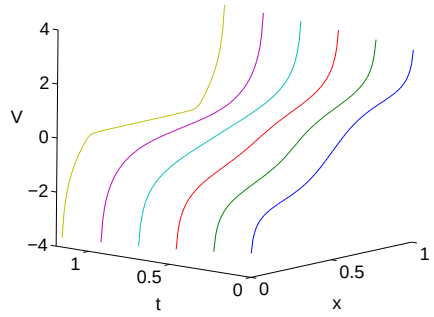


FIGURE 4.21. Cumulative function distribution function $V$ for $\chi = 3\pi$ and two initial plateaus (that is, density peaks). The solution flattens into a single plateau.

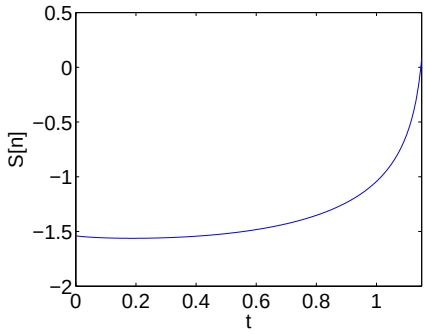


FIGURE 4.22. Evolution of the entropy.

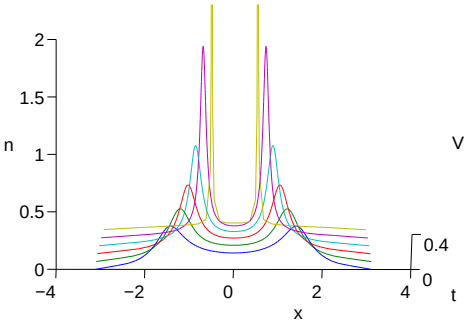


FIGURE 4.23. Cell density $n$ for two initial peaks and $\chi = 5\pi$. As opposed to the previous Section 4.3.2, each peak contains enough mass to blow-up itself.

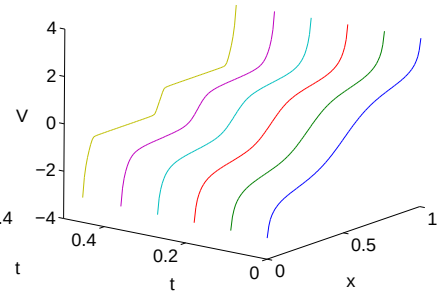


FIGURE 4.24. Cumulative distribution function $V$. Two distinct plateaus may appear when $\chi$ is above twice the critical paramater $\chi_c$.

corresponding to a two-peaks like initial density, we numerically solve the PKS equation in original variables on the time interval $[0, 0.45]$ with $\chi = 5\,\pi$.

The initial condition is the same as in Section 4.3.2 but $\chi$ is now bigger than $2\,\chi_c$. The blow-up occurs in two different points (see Figures 4.23 and 4.24).

4.4. **Two asymmetric peaks.** Given the compactly supported initial data,

$$V_0^i = \frac{\exp\left[10\ (w_i - 0.45)\right] - 1}{\left[(w_i + 0.01)\,(1.01 - w_i)\right]^{1/4}}\ ,$$

we numerically solve the PKS equation in original variables on the time interval $[0, 1.1]$ with $\chi = 3\,\pi$. Note that the initial density is not centered, but it has no effect because proposition 4.1 does not hold in this case.

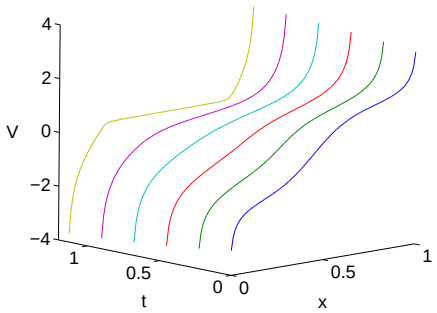


FIGURE 4.25. Cumulative distribution function $V$ when $\chi \in (\chi_c, 2\,\chi_c)$, and initial data is a two-peaks like density.

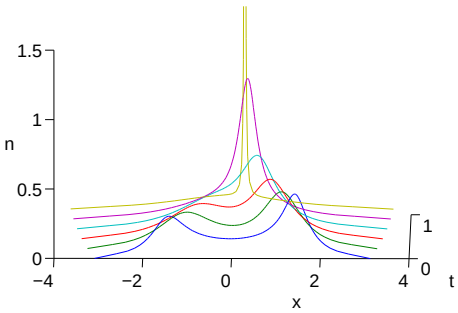


FIGURE 4.26. Evolution of the cell density $n$.

When the parameter is between the critical parameter $\chi_c$ and twice the critical parameter $2\,\chi_c$, if the peaks are asymmetric the blowup occurs at the centre of mass which is closer to the highest peak. The peaks diffuses and then the density blows-up at the centre of mass (see Figures 4.25 and 4.26).

**Acknowledgements.-** AB acknowledges the support of ESF Network "Global" and bourse Lavoisier. JAC acknowledges the support from DGI-MEC (Spain) project MTM2005-08024. AB and JAC acknowledge partial support of the Acc. Integ./Picasso program HF2006-0198. We thank the

Centre de Recerca Matemàtica (Barcelona) for providing an excellent atmosphere for research.



## References


[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations., 10, No 3 (2005) pp. 309-360.

[2] M. Agueh, N. Ghoussoub and X. Kang, *Geometric inequalities via a general comparison principle for interacting gases*, Geom. Funct. Anal., 14, (2004), pp. 215-244.

[3] L.A. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics, Birkhäuser, 2005.

[4] W. Beckner, *Sharp Sobolev inequalities on the sphere and the Moser-Trudinger inequality*, Ann. of Math., 2, 138 (1993), pp. 213–242.

[5] D. Benedetto, E. Caglioti, M. Pulvirenti, *A kinetic equation for granular media*, RAIRO Modél. Math. Anal. Numér., 31, (1997), 615–641.

[6] D. Benedetto, E. Caglioti, J.A. Carrillo, M. Pulvirenti, *A non-maxwellia ansteady distribution for one-dimensional granular media*, J. Stat. Phys., 91, (1998), pp. 979-990.

[7] A. Blanchet, J. A. Carrillo, and N. Masmoudi, *Infinite Time Aggregation for the Critical Patlak-Keller-Segel model in* $\mathbb{R}^2$, preprint.

[8] A. Blanchet, J. Dolbeault, and B. Perthame, *Two-dimensional Keller-Segel model: optimal critical mass and qualitative properties of the solutions*, Electron. J. Differential Equations, (2006), pp. No. 44, 32 pp. (electronic).

[9] M. Bodnar, J.J.L. Velazquez, *An integro-differential equation arising as a limit of individual cell-based models*, J. Diff. Eqs. 222 (2006), pp. 341-380.

[10] M. Burger and M. Di Francesco, *Large time behavior of nonlocal aggregation models with non-linear diffusion*, preprint.

[11] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. 44, (1991), pp. 375–417.

[12] E. Caglioti, P.-L. Lions, C. Marchioro and M. A. Pulvirenti, *A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description*, Comm. Math. Phys. 143 (1992), no. 3, pp. 501–525.

[13] V. Calvez and J. A. Carrillo, *Volume effects in the Keller-Segel model: energy estimates preventing blow-up*, Journal Mathématiques Pures et Appliquées, 86 (2006), pp. 155-175.

[14] V. Calvez, B. Perthame, and M. Sharifi tabar, *Modified Keller-Segel system and critical mass for the log interaction kernel*, preprint.

[15] E. Carlen and M. Loss, *Competing symmetries, the logarithmic HLS inequality and Onofri's inequality on* $S^n$, Geom. Funct. Anal., 2 (1992), pp. 90–104.

[16] J. A. Carrillo , M. P. Gualdani and G. Toscani, *Finite speed of propagation for the porous medium equation by mass transportation methods*, C. R. Math. Acad. Sci. Paris, Ser. I 338, (2004), pp. 815-818.

[17] J. A. Carrillo, A. Jüngel, P. A. Markowich, G. Toscani and A. Unterreiter, *Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities*, Monatsh. Math., 133 (2001), pp. 1–82.

[18] 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. Matemática Iberoamericana, 19 (2003), pp. 1-48.

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

[20] J. A. Carrillo and G. Toscani, *Wasserstein metric and large–time asymptotics of non-linear diffusion equations*, New Trends in Mathematical Physics, (In Honour of the Salvatore Rionero 70th Birthday), World Scientific, 2005.

[21] J. A. Carrillo and G. Toscani, *Contractive Probability Metrics and Asymptotic Behavior of Dissipative Kinetic Equations*, Porto Ercole 2006 Summer School Notes, to appear in Riv. Mat. Parma.

[22] D. Cordero-Erausquin, W. Gangbo and C. Houdre, *Inequalities for generalized entropy and optimal transportation*, Recent advances in the theory and applications of mass transport, pp. 73–94, Contemp. Math., 353, Amer. Math. Soc., Providence, RI, 2004.

[23] J. Denzler and R. McCann, *Fast diffusion to Self-Similarity: Complete Spectrum, Long-time Asymtotics and Numerology*, Arch. Rational Mech. Anal., 175 (2004), pp. 301–342.

[24] J. Dolbeault and B. Perthame, *Optimal critical mass in the two-dimensional Keller-Segel model in* $\mathbb{R}^2$, C. R. Math. Acad. Sci. Paris, 339 (2004), pp. 611–616.

[25] J. Duoandikoetxea and E. Zuazua, *Moments, masses de Dirac et décomposition de fonctions*, C. R. Acad. Sci. Paris Sér. I Math., 315 (1992), pp. 693-698.

[26] F. Filbet, *A finite volume scheme for the Patlak-Keller-Segel chemotaxis model*, Numerisch. Math., 104 (2006), pp. 457–488.

[27] L. Gosse and G. Toscani, *Identification of asymptotic decay to self-similarity for one-dimensional filtration equations*, SIAM J. Numer. Anal., 43 (2006), pp. 2590–2606.

[28] L. Gosse and G. Toscani, *Lagrangian numerical approximations to one-dimensional convolution-diffusion equations*, SIAM J. Sci. Comput., 28 (2006), pp. 1203–1227.

[29] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17 (electronic).

[30] E.F. Keller and L.A. Segel, *Initiation of slide mold aggregation viewed as an instability*, J. Theor. Biol., 26 (1970).

[31] H. Li and G. Toscani, *Long–time asymptotics of kinetic models of granular flows*, Arch. Rat. Mech. Anal., 172, (2004), pp. 407–428.

[32] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. 80, (1995), pp. 309–323.

[33] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128, 1 (1997), pp. 153–179.

[34] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geom. Funct. Anal., 11 (2001) pp. 589–608.

[35] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001) pp. 101–174.

[36] C. S. Patlak, *Random walk with persistence and external bias*, Bull. Math. Biophys., 15 (1953), pp. 311–338.

[37] T. Senba and T. Suzuki, *Weak solutions to a parabolic-elliptic system of chemotaxis*, J. Funct. Anal., 191 (2002), pp. 17–51.

[38] C.M. Topaz, A.L. Bertozzi, and M.A. Lewis, *A nonlocal continuum model for biological aggregation. Bulletin of Mathematical Biology*, 68(7), pp. 1601-1623, 2006.
[39] G. Toscani, *One-dimensional kinetic models of granular flows*, RAIRO Modél. Math. Anal. Numér., 34, 6 (2000), pp. 1277–1291.
[40] J. J. L. Velázquez, *Stability of some mechanisms of chemotactic aggregation*, SIAM J. Appl. Math., 62 (2002), pp. 1581–1633 (electronic).
[41] ———, *Point dynamics in a singular limit of the Keller-Segel model. I. Motion of the concentration regions*, SIAM J. Appl. Math., 64 (2004), pp. 1198–1223 (electronic).
[42] ———, *Point dynamics in a singular limit of the Keller-Segel model. II. Formation of the concentration regions*, SIAM J. Appl. Math., 64 (2004), pp. 1224–1248 (electronic).
[43] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics Vol. 58, Amer. Math. Soc, Providence, 2003.
[44] C. Villani, *Mathematics of granular materials*, J. Statist. Phys., 124 (2006), pp. 781-822.
[45] C. Villani, *Optimal transport, old and new*, Lecture Notes for the 2005 Saint-Flour summer school, to appear in Springer 2007.

Centre de Recerca Matemàtica, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain; http://www.ceremade.dauphine.fr/∼blanchet/.
*E-mail address*: blanchet@ceremade.dauphine.fr

Département de Mathématiques et Applications, ENS, 45 rue d'Ulm, 75005 Paris, France; http://www.dma.ens.fr/∼vcalvez/.
*E-mail address*: vincent.calvez@ens.fr

ICREA and Departament de Matemàtiques, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain; http://kinetic.mat.uab.es/∼carrillo/.
*E-mail address*: carrillo@mat.uab.es