MEASURE SOLUTIONS FOR NONLOCAL INTERACTION PDES WITH TWO SPECIES

MARCO DI FRANCESCO AND SIMONE FAGIOLI

ABSTRACT. This paper presents a systematic existence and uniqueness theory of weak measure solutions for systems of nonlocal interaction PDEs with two species, which are the PDE counterpart of systems of deterministic interacting particles with two species. The main motivations behind those models arise in cell biology, pedestrian movements, and opinion formation. In case of symmetrizable systems (i. e. with *cross-interaction* potentials one multiple of the other), we provide a complete existence and uniqueness theory within (a suitable generalization of) the Wasserstein gradient flow theory in [3, 20], which allows to consider interaction potentials with discontinuous gradient at the origin. In the general case of non symmetrizable systems, we provide an existence result for measure solutions which uses a semi-implicit version of the JKO scheme [43], which holds in a reasonable *non-smooth* setting for the interaction potentials. Uniqueness in the non symmetrizable case is proven for C^2 potentials using a variant of the method of characteristics.

1. INTRODUCTION

Several phenomena in particle physics, cell and population biology, and social sciences, can be modelled by a discrete set of N *interacting agents*, or *particles*, with *positions* $X_1(t),\ldots,X_N(t)\in\mathbb{R}^d$ depending on time, and with given *masses* $m_1,\ldots,m_N>0$. In a classical dynamic framework, and by neglecting both inertial effects (or *persistence* effects, in the language of cell biology) and the interaction of a particle with itself, the movement of the particles can be described through the Cauchy problem on $\mathbb{R}^{dN}$

$$\begin{cases} \dot{X}_j(t) = -\sum_{k\neq j} m_k \nabla G(X_j(t)-X_k(t)) \\ X_j(0)=X_j^0. \end{cases} \tag{1}$$

Thinking in terms of the empirical measure $\mu(t)$ of the particles, one has in the transport PDE

$$\frac{\partial\mu}{\partial t} = \mathrm{div}(\mu\nabla G*\mu). \tag{2}$$

the continuum counterpart of (1). The equation (2) is coupled with an initial condition $\mu_0\in\mathscr{P}(\mathbb{R}^d)$, where $\mathscr{P}(\mathbb{R}^d)$ denotes the space of probability measures on $\mathbb{R}^d$. Here G plays the role of an *interaction potential*, which is typically assumed to be (at least) continuous on $\mathbb{R}^d$ (with the possible exception of a singularity at $x=0$), and *even*, a property which ensures conservation of the center of mass

$$x_c = \int_{\mathbb{R}^d} x\,d\mu(x).$$

A more precise choice of G depends on the phenomenon under study. In population dynamics, μ measures the spatial distribution of individuals, and the interaction potential is often depending on the distance between them. In this case, G is *radial*,

2010 *Mathematics Subject Classification.* 35B40; 35B44; 45K05; 49K20; 92D25.
Key words and phrases. Nonlocal interaction equations; systems with many species; gradient flows; Wasserstein distance; semi-implicit Euler scheme; total and partial collapse.

and we use the notation $G(x) = g(|x|)$. We say that G is *attractive* if g is non-decreasing on $[0, +\infty)$, G is *repulsive* if g is non-increasing on $[0, +\infty)$. We use the term *attractive–repulsive* to denote the case in which G is repulsive on a sphere $|x| \leq R$ and attractive outside the sphere.

The setting posed in (1)–(2) can be cast in a very classical context of *mean-field limits* of large particle systems arising in statistical mechanics. We refer to [34] and the references therein for a very exhaustive description in that direction. Let us just emphasize here that this problem has several similarities with very longstanding problems in classical particle physics, cf. for instance the classical reference [50], [28] for Vlasov equations, [53] for vortex dynamics. In those contexts, the potential G features a singularity at the origin, which renders the rigorous analytical framework of the model a challenging issue. A different situation arises in more recently developed models in kinetic theory, cf. [8, 61, 21], in which (2) has been used to describe the large time dynamics of granular media. In those cases, G has the shape of a convex attractive potential, typically $G(x) = |x|^\alpha$ with $\alpha > 1$.

More recently, (1) and (2) have been recovered to provide a biologically meaningful description of aggregative phenomena in population dynamics, in particular for *swarming* phenomena, see [48, 15, 52, 60]. In those works, the nonlocal interacting forces in (1) are coupled with stochastic effects, which produce linear and nonlinear diffusions in the large particle limit, see [35, 63, 51]. Typical forms for the interaction potential in these cases are the attractive *Morse* potential $G(x) = -e^{-|x|}$, attractive-repulsive Morse potentials $G(x) = -C_a e^{-|x|/l_a} + C_r e^{-|x|/l_r}$ (where l_a and l_r are scales for the 'attractive range' and the 'repulsive range' respectively), combination of Gaussian potentials $G(x) = -C_a e^{-|x|^2/l_a} + C_r e^{-|x|^2/l_r}$, or characteristic functions of a set $G(x) = \alpha\chi_A(x)$. A mathematical property of these models which gained the attention of several researchers lately is the finite time explosion, or *blow-up*, of solutions. Without diffusion in the model, such property has been addressed in [46, 17, 22, 9, 11, 10, 12]. In [22], an optimal transport based theoretical approach led to a global well–posedness for (2) in the Wasserstein space of probability measures, with minor assumptions on G allowing to include the most relevant examples producing finite time blow-up, such as the attractive Morse potential. Part of the results in [22] generalize the theory previously developed in [3], which can be adopted for (2) with smooth potentials.

A modelling framework strictly related to (2) is that of cell movement by chemotaxis. More precisely, the two dimensional Patlak–Keller–Segel system [55, 45] in its parabolic–elliptic version corresponds to (2) with $G(x) = \frac{1}{2\pi}\log|x|$. The literature on this topic is extremely dense, we refer for instance to [42, 56, 13, 14]. The results in [12] cover a class of singular potentials which also include the Newtonian potential in all dimensions. However, most of the results in the literature do not provide a theory for measures as initial data (with the exception of [56, 29], in which however uniqueness is lacking), which would allow to describe concentrated solutions in a rigorous form. As far as the $2d$ Patlak–Keller–Segel is concerned, such hard problem has been recently successfully tackled in [47].

As pointed out in [27], time evolving measures resulting from binary interactions can be also applied to the modeling of pedestrian movements. Here, it was first modelled by Helbing that a 'social' force field biases the direction of the individuals according to the nearby distribution of neighbours, see [39, 36, 37, 38] and the references therein. In this framework, models with a nonlinear dependence on $\nabla G * \mu$ in (2) allow to describe over-crowding effects in a simpler way, see [23]. As a final interdisciplinary example, we mention opinion formation dynamics, in the was it was modelled in [58] and later studied in [62, 1]. In this case, the space variable should be rather regarded as a multi-dimensional string representing a set of opinions. A simple one-dimensional example is the political opinion, see [7] and the references therein.

Finite time concentration phenomena are sometimes considered as a very simple mathematical way to mimic *self-organization*, in the way they are opposed to a *diffusive* behavior, accounting for *spreading* of the individuals. The Patlak–Keller–Segel system for chemotaxis represents a very illustrative example in this sense, as it provides both finite time blow up and diffusive behavior (decay of solutions in the L^∞ norm for large times) for solutions to the same model with different initial data. Such a level of *complexity* in the large time behavior can be also obtained in more general systems including nonlinear diffusion, in which the interaction potential is smooth, see [18]. Here, the self-organizing behavior is represented by the emergence of spatial L^1 patterns for large times. Although the occurrence of a spatial pattern is more evocative of self–organization rather than a single particle delta measure (in particular because it incorporates more qualitative information in itself), the analysis of concentration phenomena can provide very meaningful insight to the phenomenon under study, see e. g. the recent [6]. Models allowing for solutions given by an absolutely continuous part and by a sum of particle deltas provide an easier way to detect multiple 'local' concentration phenomena, which should be distinguished by global concentration, or global *collapse*, which occurs when all particles stick together in one point. This fact is also pointed out in [22], in which absolutely solutions featuring 'multiple' concentrations before the total collapse are constructed.

A situation in which multiple collapse phenomena can provide an even deeper understanding of the qualitative behaviour is that of models with *more than one species*. For simplicity, in this paper we shall only consider the case of two species. Assume $X_1, \ldots, X_N$ are particles of the first species with masses $n_1, \ldots, n_N$, and $Y_1, \ldots, Y_M$ are particles of the second species with masses $m_1, \ldots, m_M$. A very natural way to generalize (1) is then the following:

$$\begin{cases} \dot{X}_i(t) = -\sum_{k\neq i} n_k \nabla H_1(X_i(t) - X_k(t)) - \sum_k m_k \nabla K_1(X_i(t) - Y_k(t)) \\ \dot{Y}_j(t) = -\sum_{k\neq j} m_k \nabla H_2(Y_j(t) - Y_k(t)) - \sum_k n_k \nabla K_2(Y_j(t) - X_k(t)) \end{cases} \tag{3}$$

with $i = 1, \ldots, N$ and $j = 1, \ldots, M$. Denoting with $\mu_1, \mu_2 \in \mathscr{P}(\mathbb{R}^d)$ the empirical measures of the sets X_j's and Y_j's respectively, one easily obtain the following system as continuum PDE counterpart of (3)

$$\begin{cases} \partial_t \mu_1 = \operatorname{div}\left(\mu_1 \nabla H_1 * \mu_1 + \mu_1 \nabla K_1 * \mu_2\right) \\ \partial_t \mu_2 = \operatorname{div}\left(\mu_2 \nabla H_2 * \mu_2 + \mu_2 \nabla K_2 * \mu_1\right). \end{cases} \tag{4}$$

In (3) and (4), H_1 and H_2 are called *self-interaction* potentials, whereas K_1 and K_2 are called *cross-interaction* potentials. By mass re-normalization, one can re-name the potentials in (4) in order to have μ_1 and μ_2 in the space of probability measures on $\mathbb{R}^d$. More details about this are provided in Remark 3.1.

A typical case in which multi species modeling improves significantly the understanding of the phenomena under study is the modelling of pedestrian movements. Here, models with two species are better designed in order to describe phenomena such as lane formation and segregation, see for instance the case of two-way multi lanes in a corridor in [5]. A mathematical theory with two species, though only allowing for 'smooth' interactions, has been performed in [24, 26]. In opinion formation, models with many species have been considered in [44, 30, 31].

In cells aggregation, transport models through diffusion and chemotaxis for two cells species have been considered in [40, 41, 25, 59, 65]. In [33], the question of simultaneous vs. non simultaneous blow-up in a Patlak–Keller–Segel type model with two species has been addressed. Segregation phenomena have been studied in [32] for a related multi-component model. See also [2, 57] for applications of multi-component chemotaxis flow in tumor growth.

In the present paper we aim at providing a systematic general theory for the system (4) under quite general assumptions on K_1, K_2, H_1, and H_2. Our first aim is to address the case in which there exists a constant $\alpha > 0$ such that

$$K_2 = \alpha K_1. \tag{5}$$

We shall call *symmetrizable systems* those which satisfy condition (5). Such condition is met in several applied cases such as chemotaxis modeling, see e. g. [33], and the explanation in Remark 3.1. In this case, there is a conserved quantity, which is the *joint center of mass* of the system

$$c_{M,\alpha} := \alpha \int x d\mu_1(x) + \int x d\mu_2(x).$$

Such information is useful as it provides a natural candidate for a point in which total collapse of particles of both species can occur, namely the initial joint center of mass. We want to stress here an essential difference between the two species particle system (3) and the model with one species (1) in the following example. Assume all potentials H_i, K_i, $i = 1, 2$ in (4) are all attractive and produce finite time collapse of particles in single species models, and assume that $H_1(x) \neq K_2(x)$ for all $x \in \mathbb{R}^d$. Assume two particles X_i and Y_j of separate species collide at some time t. Then, it is very unlikely that they will stick together after time t, and it is indeed very easy to produce examples in which this does not happen. Despite such a major structural difference to single species models, the symmetrizable case (5) features many similarities with the theory developed in [3, 22], in particular it can be cast in a variational Wasserstein gradient flow approach by means of the *interaction energy* functional

$$\mathcal{F}(\mu_1, \mu_2) = \frac{1}{2}\int_{\mathbb{R}^d} H_1 * \mu_1 d\mu_1 + \frac{1}{2}\int_{\mathbb{R}^d} H_2 * \mu_2 d\mu_2 + \int_{\mathbb{R}^d} H_1 * \mu_2 d\mu_1.$$

More precisely, in the spirit of [54, 3], under the symmetry assumption (5), and by suitably re-normalizing the potential H_2, the system (4) can be formally written as

$$\begin{cases} \partial_t \mu_1 = \mathrm{div}\left(\mu_1 \nabla \dfrac{\delta \mathcal{F}}{\delta \mu_1}\right) \\ \partial_t \mu_2 = \alpha \mathrm{div}\left(\mu_2 \nabla \dfrac{\delta \mathcal{F}}{\delta \mu_2}\right), \end{cases} \tag{6}$$

where the terms $\frac{\delta \mathcal{F}}{\delta \mu_1}$ and $\frac{\delta \mathcal{F}}{\delta \mu_2}$ can be interpreted at this stage as functional derivative in the spirit of Frechét derivative. As in the case of one species, see [20], the gradient flow structure will allow to stretch the regularity assumption on the interaction potentials in a way to allow for Lipschitz singularity at the origin.

The symmetry condition (5) may become too restrictive in other applied contexts, such as opinion formation, see e. g. [30, 44]. We shall therefore devote the second part of this paper to the case in which (5) is not necessarily satisfied. In this case, the classical gradient flow approach of [3] does not provide a direct tool to be used, in particular because the model is not endowed with a reasonable Lyapunov functional accounting for the total interaction energy of the system. Another applied context in which the absence of symmetry is of interest is in particle systems of predator-prey form, when H_1 and H_2 are either zero or attractive, K_1 is attractive, and K_2 is repulsive. For the use of nonlocal interaction models for predation-prey modeling, see e. g. [49].

A key aspect concerns with the regularity of the potentials. Clearly, when the potentials in (4) are smooth enough (say C^2), a classical characteristic method in the spirit of [28, 19] can be used to prove existence and uniqueness of solutions in a measure sense. However, such regularity assumption does not allow to include models producing singular phenomena such as finite time blow-up, separation, and total collapse. Most of our results are proven in way to include mildly singular potentials of Morse-type, which indeed will allow to detect those phenomena.

The paper is organized as follows. In Section 2 we recall the basics on the Wasserstein spaces of probability measures, and their generalization to product spaces needed in our theory. We also set up the exact set of assumptions on the interaction potentials, which will be used during the paper. Section 3 is devoted to the symmetrizable case. Here we prove:

- Existence and uniqueness of gradient flow solutions for mildly singular and λ convex potentials, the main result being that in Theorem 3.1.
- Total collapse for Non-Osgood potentials and total separation under further assumptions in Theorem 3.2.

In Section 4, we address the case of non symmetric systems. More precisely:

- In Theorem 4.1 we prove existence of weak measure solutions for locally Lipschitz self-interaction potentials and C^1 cross-interaction potentials.
- In Theorem 4.2, we prove uniqueness of solutions with smooth potentials.

Whereas the techniques used in Section 3 are rather reasonable generalizations of the Wasserstein gradient flow theory in [3], the existence result in Section 4 uses a semi-implicit version of the JKO scheme which is, to our knowledge, totally new. Since no gradient flow structure can be used in this case, and the regularity of the potentials is too weak to use the characteristic method, Theorem 4.1 can be considered as the main result of this paper. The uniqueness result in Section 4 is based on a simple generalization of the strategy used in [19].

2. Preliminaries on probability measures

For a given integer $n \in \mathbb{N}$, we denote with $\mathscr{P}(\mathbb{R}^n)$ the space of all the probability measures on $\mathbb{R}^n$ and with

$$\mathscr{P}_2(\mathbb{R}^n) = \left\{ \mu \in \mathscr{P}(\mathbb{R}^n) : m_2(\mu) = \int_{\mathbb{R}^n} |x|^2 \, d\mu(x) < \infty \right\}.$$

Consider a measure $\mu \in \mathscr{P}(\mathbb{R}^n)$ and a Borel map $T : \mathbb{R}^n \to \mathbb{R}^k$. We denote with $T_\# \mu \in \mathscr{P}(\mathbb{R}^d)$ the push-forward of μ through T, defined by

$$\int_{\mathbb{R}^k} f(y) dT_\# \mu(y) = \int_{\mathbb{R}^d} f(T(x)) d\mu(x) \qquad \text{for all } f \text{ Borel functions on } \mathbb{R}^n.$$

We endow the space $\mathscr{P}_2(\mathbb{R}^d)$ with the Wasserstein distance, cf. for instance [3]

$$W_2^2(\mu,\nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \left\{ \int_{\mathbb{R}^{2d}} |x-y|^2 d\gamma(x,y) \right\} \tag{7}$$

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

$$\pi^i_\# \gamma = \mu_i \quad i = 1,2$$

is satisfied. By introducing $\Gamma_o(\mu,\nu)$ as the class of optimal plans, in which the minimum in (7) is achieved, we can rewrite the Wasserstein distance as

$$W_2^2(\mu,\nu) = \int_{\mathbb{R}^{2d}} |x-y|^2 d\gamma(x,y), \qquad \gamma \in \Gamma_o(\mu,\nu).$$

The metric structure of $(\mathscr{P}_2(\mathbb{R}^d), W_2)$ (see [64, Chapter 7] for more details) can be seen as a length-space structure, as the distance between two measures μ and ν can be computed as the 'length' of a geodesic curve connecting them, in a suitable Riemannian structure, see [54, 3]. We recall that a constant speed geodesic connecting μ_1 to μ_2 can be constructed by setting

$$\gamma_t = ((1-t)\,\pi_1 + t\pi_2)_{\#}\,\gamma,$$

with $\gamma \in \Gamma_o(\mu_1,\mu_2)$, see [3, Theorem 7.2.2]. This concept allows to introduce the notion of λ-convexity of a given functional along geodesics. Let $\phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty,\infty]$ be a proper functional on $\mathscr{P}_2(\mathbb{R}^d)$. We say that ϕ is λ-convex along geodesics if, for all $\mu_1,\mu_2 \in \mathscr{P}_2(\mathbb{R}^d)$ there exists $\gamma \in \Gamma_o(\mu_1,\mu_2)$ and $\lambda \in \mathbb{R}$ such that, for all $t\in[0,1]$,

$$\phi(\gamma_t) \le (1-t)\phi\mu_1 + t\phi(\mu_2) - \frac{\lambda}{2}t(1-t)W_2^2(\mu_1,\mu_2).$$

In order to match the 'multi-species' structure (6) of our modeling setting, we shall work in the product space $\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)$ endowed with a product structure which can be adapted to the structure of the system. In particular, a weighted structure can be used to cast the symmetric case in a gradient flow theory, see Appendix A.

Remark on the notation: Throughout the whole paper we shall use bold symbols to denote elements in a product space. For instance, we use

$$\boldsymbol{\mu} = (\mu_1,\mu_2) \in \mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d),$$

or

$$\boldsymbol{x} = (x_1,x_2) \in \mathbb{R}^d\times\mathbb{R}^d.$$

Let $\alpha>0$ be fixed. For all $\boldsymbol{\mu},\boldsymbol{\nu} \in \mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)$, we define the *$\alpha$-product Wasserstein distance* as follows

$$\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu},\boldsymbol{\nu}) = W_2^2(\mu_1,\nu_1) + \frac{1}{\alpha}W_2^2(\mu_2,\nu_2).$$

The choice of the factor $1/\alpha$ above is justified in the toy model in the AppendixA. In the case $\alpha=1$ we adopt the notation $\mathcal{W}_{2,\alpha}=\mathcal{W}_2$. The notion of geodesics can be similarly generalized to the product space, and we get the following representation for a constant speed geodesic $\gamma_t \in \mathscr{P}_2(\mathbb{R}^{2d}\times\mathbb{R}^{2d})$ connecting $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$:

$$\begin{aligned} \boldsymbol{\gamma}_t &= \left(\gamma_t^1,\gamma_t^2\right), \\ \gamma_t^1 &= ((1-t)\,\pi_1 + t\pi_2)_{\#}\,\gamma_1 \qquad \gamma_t^2 = ((1-t)\,\pi_1 + t\pi_2)_{\#}\,\gamma_2, \end{aligned} \tag{8}$$

with $\gamma_1 \in \Gamma_o(\mu_1,\nu_1)$ and $\gamma_2\in\Gamma_o(\mu_2,\nu_2)$. We shall use the notation $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$ to denote the metric space $(\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d),\mathcal{W}_{2,\alpha})$, with the convention $\mathscr{P}_2(\mathbb{R}^d)^2$ when $\alpha=1$.

We can now generalize the notion of λ convexity for a functional on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$.

Definition 2.1 (λ convexity)**.** *Let $\mathcal{F} : \mathscr{P}_2(\mathbb{R}^d)^2_\alpha \to (-\infty,+\infty]$ be a proper functional. We say that $\mathcal{F}$ is* λ-convex along geodesics *on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$ if for all $\boldsymbol{\mu} = (\mu_1,\mu_2), \boldsymbol{\nu}=(\nu_1,\nu_2) \in \mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)$ there exists a geodesic curve $\boldsymbol{\gamma}_t$ as in (8) such that, for all $t\in[0,1]$,*

$$\mathcal{F}(\boldsymbol{\gamma}_t) \le (1-t)\mathcal{F}(\boldsymbol{\mu}) + t\mathcal{F}(\boldsymbol{\nu}) - \frac{\lambda}{2}t(1-t)\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu},\boldsymbol{\nu}).$$

Throughout the paper we shall work with different sets of assumptions on the interaction potentials. Let us try to collect the basic properties in the following definition.

Definition 2.2 (Interaction potentials)**.** *A function $K:\mathbb{R}^d\to\mathbb{R}$ is called an* admissible potential *if*

(Adm1) *$K\in C(\mathbb{R}^d)$ and $K(0)=0$,* (Adm2) $K(-x) = K(x)$.

*An admissible potential K is said to be λ-*convex *for some $\lambda \in \mathbb{R}$ if*

(Co) *the map $\mathbb{R}^d \ni x \mapsto K(x) - \frac{\lambda}{2}|x|^2 \in \mathbb{R}$ is convex.*

K is said to be smooth *if*

(Sm) $K \in C^1(\mathbb{R}^d)$.

K is said to be mildly singular *if*

(MS) $K \in C^1(\mathbb{R}^d \setminus \{0\})$.

K is said to be sub-quadratic at infinity *if there exists a constant $C > 0$ such that*

(SQ) $K(x) \leq C(1 + |x|^2)$ *for all $x \in \mathbb{R}^d$.*

K is said to be an attractive non-Osgood potential *if*

(Rad) *K is radial, i.e. there exists a function k such that $K(x) = k(|x|)$,*

(Mon) *k is increasing on $r > 0$, and the function $[0, +\infty) \ni r \mapsto k'(r)/r$ is non increasing,*

(N-Os) *the non-Osgood condition holds for some $\epsilon > 0$:*

$$\int_0^\epsilon \frac{dr}{k'(r)} < \infty. \tag{9}$$

Given K_{11}, K_{22}, K_{12} admissible potentials satisfying (SQ), we now introduce the *interaction energy* functional

$$\mathcal{F}(\mu_1, \mu_2) = \frac{1}{2}\int_{\mathbb{R}^d} K_{11} * \mu_1 d\mu_1 + \frac{1}{2}\int_{\mathbb{R}^d} K_{22} * \mu_2 d\mu_2 + \int_{\mathbb{R}^d} K_{12} * \mu_2 d\mu_1. \tag{10}$$

The functional $\mathcal{F}$ is well defined on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ in view of (Adm1) and (SQ). It is well known [3] that if K is λ-convex for some $\lambda < 0$, then the interaction energy functional

$$\mathcal{K} = \frac{1}{2}\int_{\mathbb{R}^d} K * \mu d\mu$$

is 2λ-geodesically convex in the usual 2-Wasserstein space. The following lemma extends such property to the two-species functional (10).

Lemma 2.1. *Let K_{11}, K_{22}, K_{12} be admissible potentials satisfying (SQ), and assume K_{ij} is λ_{ij} convex. Then, the interaction energy functional $\mathcal{F}$ on $\mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ is λ-convex along geodetics on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$, with*

$$\lambda = 2\min\{(\lambda_{11} + \lambda_{12}), \alpha(\lambda_{22} + \lambda_{12})\}.$$

Proof. Let us consider

$$\boldsymbol{\mu} = (\mu_1, \mu_2), \quad \boldsymbol{\nu} = (\nu_1, \nu_2)$$

with $\mu_i, \nu_i \in \mathscr{P}_2(\mathbb{R}^d)$, $i = 1, 2$. Let $\boldsymbol{\gamma}_t = (\gamma_t^1, \gamma_t^2)$ be a constant speed geodesic connecting $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ constructed as in (8). For the sake of convenience, we split the functional $\mathcal{F}$ as follows

$$\mathcal{F}(\boldsymbol{\mu}) = \mathcal{F}_1(\mu_1) + \mathcal{F}_2(\mu_2) + \mathcal{F}_{12}(\boldsymbol{\mu}), \tag{11}$$

with

$$\mathcal{F}_i(\mu_i) = \frac{1}{2}\int_{\mathbb{R}^d} K_{ii} * \mu_i d\mu_i, \quad i = 1, 2,$$

and

$$\mathcal{F}_{12}(\boldsymbol{\mu}) = \int_{\mathbb{R}^d} K_{12} * \mu_1 d\mu_2.$$

Similarly to [20], we compute

$$
\begin{aligned}
\mathcal{F}_1(\gamma_t^1) &= \frac{1}{2}\int_{\mathbb{R}^d} K_{11} * \gamma_t^1 d\gamma_t^1 \\
&= \frac{1}{2}\int_{\mathbb{R}^d} K_{11}\left((1-t)(x_1-x_2)+t(y_1-y_2)\right) d\gamma^1(x_1,y_1)d\gamma^1(x_2,y_2).
\end{aligned}
$$

Using the λ_{11}-convexity of K_{11}, we get

$$
\begin{aligned}
&\mathcal{F}_1(\gamma_t^1) \leq \frac{1}{2}(1-t)\int_{\mathbb{R}^d} K_{11}(x_1-x_2)d\gamma^1(x_1,y_1)d\gamma^1(x_2,y_2) \\
&+ t\int_{\mathbb{R}^d} K_{11}(y_1-y_2)d\gamma^1(x_1,y_1)d\gamma^1(x_2,y_2) \\
&- \frac{\lambda_{11}}{4}t(1-t)\int_{\mathbb{R}^d} \left|(x_1-x_2)-(y_1-y_2)\right|^2 d\gamma^1(x_1,y_1)d\gamma^1(x_2,y_2) \\
&= (1-t)\mathcal{F}_1(\mu_1)+t\mathcal{F}_1(\nu_1) - \frac{\lambda_{11}}{4}t(1-t)\int_{\mathbb{R}^d} \left|(x_1-y_1)-(x_2-y_2)\right|^2 d\gamma^1 d\gamma^1 \\
&\leq (1-t)\mathcal{F}_1(\mu_1)+t\mathcal{F}_1(\nu_1) - \lambda_{11}t(1-t)W_2^2(\mu_1,\nu_1),
\end{aligned}
$$

which means that $\mathcal{F}_{11}(\gamma_t^1)$ is $2\lambda_{11}$-convex w.r.t. t. Similarly, for $\mathcal{F}_{22}(\gamma_t^3)$ we obtain

$$
\begin{aligned}
&\mathcal{F}_2(\gamma_t^2) \leq (1-t)\mathcal{F}_2(\mu_2)+t\mathcal{F}_2(\nu_2) - \frac{\lambda_{22}}{4}t(1-t)\int_{\mathbb{R}^d} \left|(x_1-y_1)-(x_2-y_2)\right|^2 d\gamma^2 d\gamma^2 \\
&\leq (1-t)\mathcal{F}_2(\mu_2)+t\mathcal{F}_2(\nu_2) - \frac{\lambda_{22}}{\alpha}t(1-t)\alpha W_2^2(\mu_2,\nu_2).
\end{aligned}
$$

Finally, by using the definition of λ_{12}-convexity for K_{12} in the mixed term $\mathcal{F}_{12}(\boldsymbol{\gamma}_t)$, we obtain

$$
\begin{aligned}
\mathcal{F}_{12}(\boldsymbol{\gamma}_t) &= \int_{\mathbb{R}^d} K_{12} * \gamma_t^1(x)d\gamma_t^2(x) \\
&= \iint_{\mathbb{R}^d}\iint_{\mathbb{R}^d} K_{12}((1-t)(x_1-y_1)+t(x_2-y_2))d\gamma^1(y_1,y_2)d\gamma^2(x_1,x_2) \\
&\leq (1-t)\mathcal{F}_{12}(\boldsymbol{\mu})+t\mathcal{F}_{12}(\boldsymbol{\nu}) - \lambda_{12}t(1-t)\left[W_2^2(\mu_1,\nu_1)+\frac{1}{\alpha}\alpha W_2^2(\mu_2,\nu_2)\right].
\end{aligned}
$$

we can combine the above computations with (11) as follows

$$
\begin{aligned}
&\mathcal{F}(\boldsymbol{\gamma}_t) \leq (1-t)\mathcal{F}(\boldsymbol{\mu})+t\mathcal{F}(\boldsymbol{\nu}) - \lambda_{11}t(1-t)W_2^2(\mu_1,\nu_1) - \frac{\lambda_{22}}{\alpha}t(1-t)\alpha W_2^2(\mu_2,\nu_2) \\
&- \lambda_{12}t(1-t)\left[W_2^2(\mu_1,\nu_1)+\frac{1}{\alpha}\alpha W_2^2(\mu_2,\nu_2)\right]
\end{aligned}
$$

Since all λ_{ij}'s are negative

$$
\mathcal{F}(\boldsymbol{\gamma}_t) \leq (1-t)\mathcal{F}(\boldsymbol{\mu})+t\mathcal{F}(\boldsymbol{\nu}) - \frac{\lambda}{2}t(1-t)\mathcal{W}_2^{2,\alpha}(\boldsymbol{\mu},\boldsymbol{\nu})
$$

■

Remark 2.1. Following [20, Remark 1.1], assuming all K_{ij} satisfy assumptions (Co) and (SQ), we can easily prove that

$$
|\nabla K_{ij}(x)| \leq C(1+|x|) \tag{12}
$$

for some positive constant C independent of x.

3. Symmetric cross-interaction

In this section we analyse the symmetrizable case

$$\begin{cases} \partial_t \mu_1 = \mathrm{div}\left(\mu_1 \nabla K_{11} * \mu_1 + \mu_1 \nabla K_{12} * \mu_2\right) \\ \partial_t \mu_2 = \alpha \mathrm{div}\left(\mu_2 \nabla K_{22} * \mu_2 + \mu_2 \nabla K_{12} * \mu_1\right). \end{cases} \tag{13}$$

Throughout this section we shall assume that all the kernels K_{ij} in (13) are admissible and satisfy (Co), (MS), and (SQ).

Remark 3.1 (Motivation)**.** In order to justify the structure (13) let us consider the more general system

$$\begin{cases} \partial_\tau \widetilde{\mu}_1 = \mathrm{div}\left(\widetilde{\mu}_1 \nabla G_{11} * \widetilde{\mu}_1 + \widetilde{\mu}_1 \nabla G_{12} * \widetilde{\mu}_2\right) \\ \partial_\tau \widetilde{\mu}_2 = \mathrm{div}\left(\widetilde{\mu}_2 \nabla G_{22} * \widetilde{\mu}_2 + \widetilde{\mu}_2 \nabla G_{21} * \widetilde{\mu}_1\right), \end{cases} \tag{14}$$

endowed with the condition

$$G_{21} = \beta G_{12}$$

for some $\beta > 0$, with

$$\int_{\mathbb{R}^d} d\mu_1(x) = M_1, \qquad \int_{\mathbb{R}^d} d\mu_2(x) = M_2.$$

Such situation occurs e. g. in the case of two aggregating species of cells ρ_1 and ρ_2, with motion driven up the gradient of a chemical substance c by two chemical coefficients $\chi_1, \chi_2 > 0$, namely

$$\begin{cases} \frac{\partial \rho_1}{\partial t} = \chi_1 \mathrm{div}(\rho_1 \nabla c) \\ \frac{\partial \rho_2}{\partial t} = \chi_2 \mathrm{div}(\rho_1 \nabla c) \end{cases}$$

and the concentration of the chemical being nonlocally regulated by the two species as

$$c = B * (a\rho_1 + b\rho_2), \qquad a, b > 0$$

with B being an admissible attractive kernel. In chemotaxis [33, 40], B is the Newtonian potential. Let us go back to system (14). By re-normalizing the masses

$$\mu_i := \frac{1}{M_i} \widetilde{\mu}_i, \quad i = 1, 2,$$

one obtains

$$\begin{cases} \frac{1}{M_2} \partial_\tau \mu_1 = \mathrm{div}\left(\mu_1 \nabla \frac{M_1}{M_2} G_{11} * \mu_1 + \mu_1 \nabla G_{12} * \mu_2\right) \\ \frac{1}{\beta M_1} \partial_\tau \mu_2 = \mathrm{div}\left(\mu_2 \nabla \frac{M_2}{\beta M_1} G_{22} * \mu_2 + \mu_2 \nabla G_{12} * \mu_1\right). \end{cases}$$

By time-normalization

$$t = M_2 \tau$$

we obtain (13) with $\alpha = \frac{\beta M_1}{M_2}$, $K_{11} = \frac{M_1}{M_2} G_{11}$, $K_{22} = \frac{M_2}{\beta M_1} G_{22}$, $K_{12} = G_{12}$.

3.1. **Gradient flow structure of the system.** System (13) can be studied by generalizing the theory of gradient flows on probability spaces developed in [3], [4] and [20] for nonlocal interaction equations with one single species. We shall adopt the 'product' metric structure on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$ introduced in Section 2, with α being the constant in the second equation of (13). The system (13) will be recovered as the gradient flow on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$ of the functional $\mathcal{F}$ defined in (10). For simplicity, from now on we shall use the notation

$$L^2(\boldsymbol{\mu}) = L^2(d\mu_1) \times L^2(d\mu_2),$$

and for $\boldsymbol{v} = (v_1, v_2) \in L^2(\boldsymbol{\mu})$,

$$\|\boldsymbol{v}\|^2_{L^2_\alpha(\boldsymbol{\mu})} = \int_{\mathbb{R}^d} v_1^2(x) d\mu_1(x) + \frac{1}{\alpha} \int_{\mathbb{R}^d} v_2^2(x) d\mu_2(x).$$

We recall that, given an admissible potential K that satisfies (Co), we can define the sub-differential of K as:

$$\partial K(x) = \left\{ \xi \in \mathbb{R}^d : K(y) - K(x) \geq \xi \cdot (y - x) + o(|y - x|), \forall y \in \mathbb{R}^d \right\}.$$

For the a potential K being admissible, λ-convex and mildly singular, we can easily prove that the element of minimal L^2-norm in $\partial K(x)$, called $\partial^0 K(x)$, is given by:

$$\partial^0 K(x) = \begin{cases} \nabla K(x) & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

Let $\mathcal{F}$ be the functional (10) defined on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$. Following the strategy in [3, 20], we want to introduce the notion of sub-differential $\partial\mathcal{F}[\boldsymbol{\mu}]$ for the functional $\mathcal{F}$ on an element $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathscr{P}_2(\mathbb{R}^d)^2_\alpha$, as well as the concept of *minimal sub-differential* $\partial^0\mathcal{F}[\boldsymbol{\mu}]$, i. e. the set of elements in $\partial\mathcal{F}[\boldsymbol{\mu}]$ with minimal $L^2_\alpha(\boldsymbol{\mu})$ norm. The sub-differential of the functional $\mathcal{F} : \mathscr{P}_2(\mathbb{R}^d)^2_\alpha \to \mathbb{R}$.

Definition 3.1. *Let $\mathcal{F}$ be the functional* (10) *on $\mathscr{P}_2(\mathbb{R}^d)^2_\alpha$ and let $\boldsymbol{\mu} \in \mathscr{P}_2(\mathbb{R}^d)^2_\alpha$. A vector field $\boldsymbol{k} = (k_1, k_2) \in L^2(\boldsymbol{\mu})$ is an element of the sub-differential of $\mathcal{F}$ in $\boldsymbol{\mu}$, in formulas $\boldsymbol{k} \in \partial\mathcal{F}[\boldsymbol{\mu}]$, if*

$$\begin{aligned}\mathcal{F}(\boldsymbol{\nu}) - \mathcal{F}(\boldsymbol{\mu}) \geq & \inf_{\gamma_i \in \Gamma_o(\mu_i, \nu_i), i=1,2} \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} [k_1(x_1) \cdot (y_1 - x_1) \\ & + \alpha k_2(x_2) \cdot (y_2 - x_2)] \, d\gamma_1(x_1, y_1) d\gamma_2(x_2, y_2) + \\ & + o\left(\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu})\right) \qquad \forall \boldsymbol{\nu} \in \mathscr{P}(\mathbb{R}^d) \times \mathscr{P}(\mathbb{R}^d).\end{aligned}$$

We denote with $\partial^0\mathcal{F}[\boldsymbol{\mu}]$ the set of elements of minimal $L^2(\boldsymbol{\mu})$ norm in $\partial\mathcal{F}[\boldsymbol{\mu}]$.

Since the functional $\mathcal{F}$ is λ-convex in view of Lemma 2.1, the condition in the above definition can be easily proven to be equivalent (we omit the details, see [20]) to:

$$\begin{aligned}&\liminf_{t \to 0} \frac{\mathcal{F}(\boldsymbol{\gamma}_t) - \mathcal{F}(\boldsymbol{\mu})}{t} \\ &\geq \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} [k_1(x_1) \cdot (y_1 - x_1) + \alpha k_2(x_2) \cdot (y_2 - x_2)] \, d\gamma_1(x_1, y_1) d\gamma_2(x_2, y_2) \quad (15)\end{aligned}$$

for all $\gamma_i \in \Gamma_o(\mu_i, \nu_i)$. Now we can state our definition of solution for the system (13).

Definition 3.2. *We say that an absolutely continuous curve $\boldsymbol{\mu}_t = (\mu_{1,t}, \mu_{2,t}) : [0, T] \to \mathscr{P}_2(\mathbb{R}^d) \times \mathscr{P}_2(\mathbb{R}^d)$ is a gradient flow for $\mathcal{F}$ if $\mu_{1,t}$ and $\mu_{2,t}$ solve the system of two continuity equations:*

$$\begin{aligned}\partial_t \mu_{1,t} &= div\,(\mu_1 v_{1,t}) \\ \partial_t \mu_{2,t} &= div\,(\mu_2 v_{2,t})\,, \end{aligned} \quad (16)$$

with $v_{i,t} \in \left(\partial^0\mathcal{F}[\boldsymbol{\mu}]\right)_i$ for $i = 1, 2$.

Proposition 3.1. *Let K_{ij} be admissible potentials satisfying (Co), (SQ), and (MS). Then the vector field*

$$\partial^0\mathcal{F}[\boldsymbol{\mu}](\boldsymbol{x}) = \begin{pmatrix} \partial^0 K_{11} * \mu_1 + \partial^0 K_{12} * \mu_2 \\ \frac{1}{\alpha}\partial^0 K_{22} * \mu_2 + \frac{1}{\alpha}\partial^0 K_{12} * \mu_1 \end{pmatrix} \quad (17)$$

with

$$\partial^0 K_{ij} * \mu_j(x_2) = \int_{x_1 \neq x_2} \nabla K_{ij}(x_1 - x_2) d\mu_j(x_1). \quad (18)$$

is the unique element of minimal subdifferential of $\mathcal{F}$.

Proof. Decompose K_{ij} as

$$K_{ij}(x) = \tilde{K}_{ij}(x) + \frac{\lambda_{ij}}{2}|x|^2,$$

with $\tilde{K}_{ij}$ convex and such that $0 \in \partial^0 \tilde{K}_{ij}$. Define

$$\tilde{\mathcal{F}}_{ij}(\mu_i, \mu_j) = \frac{\epsilon_{ij}}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{K}_{ij}(x-y) d\mu_j(x) d\mu_i(y)$$
$$\mathcal{Q}(\mu_i, \mu_j) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\mu_j(x) d\mu_i(y),$$

with $\epsilon_{ij} = 1$ if $i = j$ and $\epsilon_{ij} = 2$ if $i \neq j$. Let us consider $\gamma_t^i = ((1-t)\pi_1 + t\pi_2)_\sharp \gamma_i$, $i = 1, 2$, and $\gamma_i \in \Gamma_o(\mu_i, \nu_i)$. We get

$$\begin{aligned}&\frac{\tilde{\mathcal{F}}_{ij}(\gamma_t^i, \gamma_t^j) - \tilde{\mathcal{F}}_{ij}(\mu_i, \mu_j)}{t} \\ &= \frac{\epsilon_{ij}}{2t} \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} \Big(\tilde{K}_{ij}((1-t)(x_1 - y_1) + t(x_2 - y_2)) \\ &-\tilde{K}_{ij}(x_1 - y_1) \Big) d\gamma_1(x_1, x_2) d\gamma_1(y_1, y_2).\end{aligned}$$

Let us first consider the diagonal term $i = j = 1$. We obtain

$$\begin{aligned}&\frac{\tilde{\mathcal{F}}_{11}(\gamma_t^1, \gamma_t^1) - \tilde{\mathcal{F}}_{11}(\mu_1, \mu_1)}{t} \\ &= \frac{1}{2t} \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} \Big(\tilde{K}_{11}((1-t)(x_1 - y_1) + t(x_2 - y_2)) \\ &-\tilde{K}_{11}(x_1 - y_1) \Big) d\gamma_1(x_1, x_2) d\gamma_1(y_1, y_2).\end{aligned}$$

In the limit for $t \to 0$ (see [20, Proposition 2.2]), as $\tilde{K}_{11} \geq 0$, $\nabla \tilde{K}$ is even, and $\tilde{K}_{11}$ satisfies (SQ), we obtain

$$\begin{aligned}&\frac{\tilde{\mathcal{F}}_{11}(\gamma_t^1, \gamma_t^1) - \tilde{\mathcal{F}}_{11}(\mu_1, \mu_1)}{t} \\ &\geq \int_{x_1 \neq x_2} \nabla \tilde{K}_{11}(x_1 - y_1) \cdot (x_2 - x_1) d\gamma_1(x_1, x_2) d\gamma_1(y_1, y_2).\end{aligned}$$

Similarly, we get

$$\begin{aligned}&\frac{\tilde{\mathcal{F}}_{22}(\gamma_t^2, \gamma_t^2) - \tilde{\mathcal{F}}_{22}(\mu_2, \mu_2)}{t} \\ &\geq \int_{x_1 \neq x_2} \nabla \tilde{K}_{22}(x_1 - y_1) \cdot (x_2 - x_1) d\gamma_2(x_1, x_2) d\gamma_2(y_1, y_2).\end{aligned}$$

Let us consider now the mixed term

$$\begin{aligned}&\frac{\tilde{\mathcal{F}}_{12}(\gamma_t^1, \gamma_t^2) - \tilde{\mathcal{F}}_{12}(\mu_1, \mu_2)}{t} \\ &= \frac{1}{t} \int_{\mathbb{R}^{2d} \times \mathbb{R}^{2d}} \Big(\tilde{K}_{12}((1-t)(x_1 - y_1) + t(x_2 - y_2)) \\ &-\tilde{K}_{12}(x_1 - y_1) \Big) d\gamma_1(x_1, x_2) d\gamma_2(y_1, y_2).\end{aligned}$$

As $\tilde{K}_{12} \geq 0$ and $\nabla\tilde{K}$ is even, we get as above

$$\begin{aligned}&\frac{\tilde{\mathcal{F}}_{12}(\gamma_t^1,\gamma_t^2)-\tilde{\mathcal{F}}_{12}(\mu_1,\mu_2)}{t}\\&\geq \frac{1}{t}\int_{x_1\neq x_2}\Big(\tilde{K}_{12}((1-t)(x_1-y_1)+t(x_2-y_2))\\&-\tilde{K}_{12}(x_1-y_1)\Big)\,d\gamma_1(x_1,x_2)d\gamma_2(y_1,y_2),\end{aligned}$$

which converges as $t \searrow 0$ to

$$\begin{aligned}&\int_{x_1\neq x_2}\nabla\tilde{K}_{12}(x_1-y_1)\cdot(x_2-x_1)d\gamma_2(y_1,y_2)d\gamma_1(x_1,x_2)\\&+\int_{x_1\neq x_2}\nabla\tilde{K}_{12}(y_1-x_1)\cdot(y_2-y_1)d\gamma_1(x_1,x_2)d\gamma_2(y_1,y_2).\end{aligned}$$

The quadratic term $\mathcal{Q}$ satisfies the following limit as $t \searrow 0$

$$\begin{aligned}&\frac{\mathcal{Q}(\gamma_t^i,\gamma_t^j)-\mathcal{Q}(\mu_i,\mu_j)}{t}=\\&=\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}}\left(|(1-t)(x_1-y_1)+t(x_2-y_2)|^2-|x_1-y_1|^2\right)d\gamma_i(x_1,x_2)d\gamma_j(y_1,y_2)\to\\&2\int_{x_1\neq x_2}(x_1-y_1)\cdot(x_2-x_1-y_2+y_1)d\gamma_j(x_2,y_2)d\gamma_i(x_1,y_1).\end{aligned}$$

Summing all the contributions, and using the definition (18) of ∂^0, we obtain

$$\begin{aligned}\liminf_{t\to0}\frac{\mathcal{F}(\boldsymbol{\gamma}_t)-\mathcal{F}(\boldsymbol{\mu})}{t}&\geq\int\partial^0K_{11}*\mu_1(x_1)\cdot(x_2-x_1)d\gamma_1(x_1,x_2)+\\&+\int\partial^0K_{22}*\mu_2(y_1)\cdot(y_2-x_2)d\gamma_2(y_1,y_2)+\\&+\int\partial^0K_{12}*\mu_2(x_1)\cdot(x_2-x_1)d\gamma_1(x_1,x_2)+\\&+\int\partial^0K_{12}*\mu_1(y_1)\cdot(y_2-y_1)d\gamma_2(y_1,y_2),\end{aligned}$$

and this proves the assertion with simple manipulations. In order to show that $\partial^0\mathcal{F}$ is the element of minimal L^2_α-norm, one can retrace the argument in [20, Proposition 2.2].

■

3.2. **Existence and uniqueness of solution.** We are now ready to prove that we can construct solutions to (13) as the gradient flow of the functional $\mathcal{F}$ in the sense of Definition 3.2. As suggested by the preliminaries in subsection 3.1, we shall retrace the strategies in [3, 20]. Some steps in our construction are very similar to those in [3, 20], and they will therefore be skipped. In order to prove the existence of the solution of the equation in the gradient flow framework, we shall first state the existence of a curve of maximal slope for the functional $\mathcal{F}$. Let us recall the definitions for

- slope of a functional $\mathcal{F}$:
$$|\partial\mathcal{F}|[\boldsymbol{\mu}]=\limsup_{\boldsymbol{\nu}\to\boldsymbol{\mu}}\frac{(\mathcal{F}(\boldsymbol{\mu})-\mathcal{F}(\boldsymbol{\nu}))^+}{\mathcal{W}_{2,\alpha}(\boldsymbol{\mu},\boldsymbol{\nu})};$$
- metric derivative of an absolutely continuous curve $\boldsymbol{\mu}_t:[0,T]\to\mathscr{P}(\mathbb{R}^d)^2_\alpha$:
$$|\boldsymbol{\mu}'|(t)=\limsup_{s\to t}\frac{\mathcal{W}_{2,\alpha}(\boldsymbol{\mu}_s,\boldsymbol{\mu}_t)}{|s-t|}.$$

We recall the definition of a curve of maximal slope for the functional $\mathcal{F}$, that is a curve of maximal slope with respect to $|\partial\mathcal{F}|[\boldsymbol{\mu}]$.

Definition 3.3. *An absolutely continuous curve $\boldsymbol{\mu}_t : [0,T] \to \mathscr{P}(\mathbb{R}^d)\times\mathscr{P}(\mathbb{R}^d)$ is a curve of maximal slope for the functional $\mathcal{F}$ if $t \to \mathcal{F}[\boldsymbol{\mu}_t]$ is an absolutely continuous function, and the following inequality holds for all $0 \le s \le t \le T$:*

$$\mathcal{F}[\boldsymbol{\mu}_s] - \mathcal{F}[\boldsymbol{\mu}_t] \geq \frac{1}{2}\int_s^t \left[|\boldsymbol{\mu}'|^2(\tau) + |\partial\mathcal{F}|[\boldsymbol{\mu}_\tau]^2\right] d\tau$$

We obtain the existence of a curve of maximal slope by means of the Jordan-Kinderlehrer-Otto (JKO) scheme [43], which we shall recall here for the reader's convenience: given an initial product measure $\boldsymbol{\mu}_0 \in \mathscr{P}(\mathbb{R}^d)\times\mathscr{P}(\mathbb{R}^d)$ and a time step $\tau > 0$, we define the recursive sequence $\boldsymbol{\mu}_k^\tau$ via $\boldsymbol{\mu}_0^\tau = \boldsymbol{\mu}_0$ and

$$\boldsymbol{\mu}_\tau^{k+1} \in \operatorname{argmin}_{\boldsymbol{\mu}\in\mathscr{P}_2(\mathbb{R}^d)^2_\alpha}\left\{\frac{1}{2\tau}\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu}_\tau^k,\boldsymbol{\mu}) + \mathcal{F}[\boldsymbol{\mu}]\right\}.$$

The convergence of the scheme can be proven by retracing the proof in [20, Proposition 2.5-2.6 and Theorem 2.8], without adding any element of relevance. We shall therefore omit it. A key step in the proof is the energy inequality

$$\mathcal{F}[\boldsymbol{\mu}(0)] - \mathcal{F}[\boldsymbol{\mu}^n(T)] \geq \frac{1}{2}\int_0^T \|\boldsymbol{v}^n\|^2_{L^2_\alpha(\boldsymbol{\mu}^n)}(t) + |\partial\mathcal{F}|[\tilde{\boldsymbol{\mu}}(t)]^2 dt,$$

where $\tilde{\boldsymbol{\mu}}(t)$ is the De Giorgi variational interpolation [3], obtained thanks to the lower semi-continuity of the slope, which can be proven similarly to [20, Lemma 2.7]. The (possibly multiple) curve of maximal slope obtained as a limit of the JKO scheme can be proven to be actually a gradient flow solution to (13) in the sense of Definition 3.2 by following the same technique in [3, Theorem 11.1.3]. We shall omit this step. Please notice that in Section 4 we shall rigorously prove (in detail) the existence of weak measure solutions in a more general setting.

In order to prove uniqueness of gradient flow solutions, we shall now employ the convexity properties of the functional $\mathcal{F}$ described above in Section 2 in order to prove the Evolution Variational Inequalities (E. V. I.) in the spirit of [3], and consequently the $\mathcal{W}_2$ $|\lambda|$-contraction of the gradient flow. Before tackling this task, we state the differentiability of the Wasserstein distance along the gradient flow: if $\boldsymbol{\mu}_t$ is a gradient flow of the functional $\mathcal{F}$, for all $\boldsymbol{\nu} \in \mathscr{P}_2(\mathbb{R}^d)^2_\alpha$ we have

$$\frac{1}{2}\frac{d}{dt}\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu}_t,\boldsymbol{\nu}) = \int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}} \boldsymbol{v}_t(\mathbf{x})\cdot(\mathbf{x}-\mathbf{y})d\gamma_1(x_1,y_1)d\gamma_2(x_2,y_2).$$

t-a.e in $[0,T]$. Here $\boldsymbol{v}_t = (v_{1,t}, v_{2,t})$ with $v_{i,t}$ as in Definition 3.2. Once again, the proof can be omitted as it can be obtained by adapting the proof in [4, Theorem 2.21]. We are now ready to state the E. V. I. in the following Theorem.

Theorem 3.1. *Let K_{ij} be admissible potentials satisfying (Co), (SQ) and (MS), and let $\boldsymbol{\mu}_t$ be a gradient flow solution to (13) according to Definition 3.2. Then, $\boldsymbol{\mu}_t$ satisfies the following* Evolution Variational Inequality (E. V. I.)

$$\frac{1}{2}\frac{d}{dt}\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu}_t,\boldsymbol{\nu}) + \frac{\lambda}{2}\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu},\boldsymbol{\nu}) \le \mathcal{F}(\boldsymbol{\nu}) - \mathcal{F}(\boldsymbol{\mu}), \tag{19}$$

for all $\boldsymbol{\nu} \in \mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)$. In addition, given two gradient flow solutions $\boldsymbol{\mu}^1$, $\boldsymbol{\mu}^2$ corresponding to the initial data $\boldsymbol{\mu}_0^1$ and $\boldsymbol{\mu}_0^2$, we have the $|\lambda|$-contraction property in $\mathcal{W}_{2,\alpha}$

$$\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu}_t^1,\boldsymbol{\mu}_t^2) \le e^{|\lambda|t}\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu}_0^1,\boldsymbol{\mu}_0^2). \tag{20}$$

In particular, for a given initial condition in $\mathcal{W}_{2,\alpha}$ there exists a unique gradient flow solution to (13) in the sense of Definition 3.2.

Proof. Given $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathscr{P}_2(\mathbb{R}) \times \mathscr{P}_2(\mathbb{R})$, due to the $\lambda-$convexity of $\mathcal{F}$,

$$\mathcal{F}(\boldsymbol{\gamma}_t) \leq (1-t)\mathcal{F}(\boldsymbol{\mu}) + t\mathcal{F}(\boldsymbol{\nu}) - \frac{\lambda}{2}t(1-t)\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu})$$

where

$$\begin{aligned}
&\boldsymbol{\gamma}_t = (\gamma^1_t, \gamma^2_t) \\
&\gamma^1_t = ((1-t)\,\pi_1 + t\pi_2)_\#\,\gamma_1 \qquad \gamma^2_t = ((1-t)\,\pi_1 + t\pi_2)_\#\,\gamma_2, \\
&\gamma^1 \in \Gamma(\mu_1, \nu_1), \gamma^2 \in \Gamma(\mu_2, \nu_2)
\end{aligned}$$

Following a standard computation as in [3, 4],we can write

$$\frac{\mathcal{F}(\boldsymbol{\gamma}_t) - \mathcal{F}(\boldsymbol{\mu})}{t} \leq -\mathcal{F}(\boldsymbol{\mu}) + \mathcal{F}(\boldsymbol{\nu}) - \frac{\lambda}{2}(1-t)\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu}, \boldsymbol{\nu}). \tag{21}$$

Then, using the characterization of the sub-differential (15) and passing to the limit as $t \searrow 0$ in (21), we obtain:

$$\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}} \partial^0\mathcal{F}(\mathbf{x})\cdot(\mathbf{x}-\mathbf{y})d\gamma_1(x_1,y_1)d\gamma_2(x_2,y_2) + \frac{\lambda}{2}\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu},\boldsymbol{\nu}) \leq \mathcal{F}(\boldsymbol{\nu}) - \mathcal{F}(\boldsymbol{\mu}). \tag{22}$$

Then, let $\boldsymbol{\mu} = \boldsymbol{\mu}_t$ in (22) be a solution for the (13), one can reconstruct the derivative of the Wasserstein distance from the first term in the left hand side of the above equation,

$$\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}} \partial^0\mathcal{F}(\mathbf{x})\cdot(\mathbf{x}-\mathbf{y})d\gamma_1(x_1,y_1)d\gamma_2(x_2,y_2) = \frac{1}{2}\frac{d}{dt}\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu}_t,\boldsymbol{\nu})$$

and then get the E.V.I.

$$\frac{1}{2}\frac{d}{dt}\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu}_t,\boldsymbol{\nu}) + \frac{\lambda}{2}\mathcal{W}^2_{2,\alpha}(\boldsymbol{\mu},\boldsymbol{\nu}) \leq \mathcal{F}(\boldsymbol{\nu}) - \mathcal{F}(\boldsymbol{\mu}).$$

■

Remark 3.2 (Particle solutions)**.** Similarly to [20], we remark here that *particle solutions are gradient flow solutions.* More precisely, let $X_1(t), \ldots, X_N(t)$, $Y_1(t), \ldots, Y_M(t)$ solve (globally and almost everywhere in time) the system

$$\begin{cases} \dot{X}_i = -\displaystyle\sum_{k\in C_i} m^k_X \nabla K_{11}(X_i - X_k) - \sum_{j\in D_i} m^j_Y \nabla K_{12}(X_i - Y_j) & i = 1,..,N \\ \dot{Y}_j = -\displaystyle\sum_{h\in E_j} m^h_Y \nabla K_{22}(Y_j - Y_h) - \sum_{i\in F_j} m^i_X \nabla K_{12}(Y_j - X_i) & j = 1,..,M, \end{cases} \tag{23}$$

with $C_i = \{k \in \{1,\ldots,N\} : X_k \neq X_i\}$, $D_i = \{j \in \{1,\ldots,M\} : Y_j \neq X_i\}$, $E_j = \{h \in \{1,\ldots,M\} : Y_j \neq Y_h\}$, $F_j = \{i \in \{1,\ldots,N\} : Y_j \neq X_i\}$, and let

$$\mu_1(t) = \sum_{i=1}^N m^i_X \delta_{X_i(t)}, \qquad \mu_2(t) = \sum_{j=1}^M m^j_Y \delta_{Y_j(t)}. \tag{24}$$

Then, the curve $[0,+\infty) \ni t \mapsto \boldsymbol{\mu}(t) = (\mu_1(t), \mu_2(t))$ is a gradient flow solution to (13) in the sense of Definition 3.2. Conversely, let the initial condition for (13) be given by $\boldsymbol{\mu}_0 = (\mu_{1,0}, \mu_{2,0})$ with

$$\mu_{1,0} = \sum_{i=1}^N m^i_X \delta_{X_{0,i}}, \qquad \mu_{2,0} = \sum_{j=1}^M m^j_Y \delta_{Y_{0,j}},$$

then, the unique gradient flow solution to (13) is of the form (24). The proof of such two statements would be trivial if no collisions occur, see also [20, Remark 2.10]. In the case of one single species, though, only a finite number of collisions occur, which simplifies the problem of giving sense to the ODE system (23). However, as stated in the introduction, in our case collisions between particles of different species do not necessarily imply (almost never, actually) that those particles will stick together indefinitely. Therefore, infinitely many collisions are likely to occur. However, it is easily seen that they can only be of a countable number, since one can always 'restart' the particle system after each collision, and since the number of particles is finite there always exists a nonzero waiting time before the next collision occurs. Nevertheless, this argument alone cannot guarantee the global existence of particle solutions almost everywhere in time, as the sequence of collision times could accumulate at a finite time. In the Appendix B, we prove that such task can be achieved by means of a finite dimensional gradient flow structure, which ensures that the solution is almost everywhere globally defined in time (and unique!). Please notice that here we have assumed $\alpha = 1$ without restriction, since we are not requiring any normalization condition on the masses.

3.3. **Finite time blow-up phenomena.** We turn now on the studying of the large time behaviour of the symmetrizable system (13), in particular the case of *attractive non-Osgood potentials* K_{ij} in the sense of Definition 2.2. More precisely, we shall assume that, further to the conditions needed to obtain a unique gradient flow solution, K_{ij} also satisfy conditions (Rad), (Mon), and (N-Os). Similarly to the case of a single species, the Non-Osgood condition (N-Os) is responsible for the *collapse in finite time* of the particles of both species at one single point, for all compactly supported initial measures. On the other hand, we shall also prove that the *total separation* of the two species is also possible before the total collapse occurs.

The strategy follows the basic idea used in [20], namely to study the behaviour of finite particle solutions and to use the stability property (20) in order to pass to general solutions. Let us consider the particle system with two species

$$\begin{cases} \dot{X}_i = -\sum\limits_{k\in C_i} m_X^k \nabla K_{11}(X_i - X_k) - \sum\limits_{j\in D_i} m_Y^j \nabla K_{12}(X_i - Y_j) & i = 1, .., N \\ \dot{Y}_j = -\sum\limits_{h\in E_j} m_Y^h \nabla K_{22}(Y_j - Y_h) - \sum\limits_{i\in F_j} m_X^i \nabla K_{21}(Y_j - X_i) & j = 1, .., M, \end{cases} \tag{25}$$

with

$$\begin{aligned} C_i &= \{k \in \{1, \dots, N\} : X_k \neq X_i\} \\ D_i &= \{j \in \{1, \dots, M\} : Y_j \neq X_i\} \\ E_j &= \{h \in \{1, \dots, M\} : Y_j \neq Y_h\} \\ F_j &= \{i \in \{1, \dots, N\} : Y_j \neq X_i\}, \end{aligned}$$

in which both total masses satisfies

$$\sum_{k=1}^{N} m_X^k = \sum_{h=1}^{M} m_Y^h = 1.$$

We remark that, unlike in Remark 3.2, we are re-normalizing the masses and assuming the condition

$$K_{21} = \alpha K_{12} \tag{26}$$

holds for some $\alpha > 0$.

For future use, for a general pair of measures $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathscr{P}(\mathbb{R}^d)^2_\alpha$, we define the *partial center or masses*

$$X_C(t) = \int x d\mu_1(x), \qquad Y_C(t) = \int y d\mu_2(y),$$

and recall the definition of *joint center of mass*

$$C_{M,\alpha} = \frac{\alpha X_C + Y_C}{2}.$$

We recall that, whereas $X_C(t)$ and $Y_C(t)$ may vary on time, the quantity $C_{M,\alpha}$ is preserved in time, as it can be easily seen by a simple computation. For simplicity in the notation, we shall often skip the subscript α in $C_{M,\alpha}$.

Without loss of generality we can assume that $C_M = 0$. In the following proposition we will show finite-time total collapse for particles under the crucial assumption (N-Os)

Proposition 3.2. *Let K_{ij} be admissible potentials satisfying (Co), (SQ), (MS), (Rad), (Mon) and (N-Os). Let the initial datum for (25) be given by $X_i(0) = X_i^0$, $Y_j(0) = Y_i^0$, with masses m_X^i and m_Y^j respectively for the particles X_i^0 and Y_j^0. Then then there exists $T^* > 0$ such that the unique solution to the ODE system (25) satisfies*

$$X_i(t) \equiv Y_j(t) \equiv 0$$

for all $t \geq T^*$, or equivalently, the unique gradient flow solution of (13) with initial datum

$$\boldsymbol{\mu}_0 = \left(\sum_{i=1}^{N} m_X^i \delta_{X_i^0}, \sum_{j=1}^{M} m_Y^j \delta_{Y_j^0} \right)$$

satisfies

$$\boldsymbol{\mu}(t) = (\delta_0, \delta_0) \quad \forall t \geq T^*,$$

and T^* only depends on the initial largest distance of the particles to the total center of mass:

$$R_0 = \max_{i,j} \left\{ |X_i^0|, |Y_j^0| \right\}.$$

Proof. Consider

$$R(t) = \max_{i,j} \{ |X_i(t)|, |Y_j(t)| \}$$

and we want compute

$$\frac{d}{dt} R^2(t)$$

Since the number of particles is finite, for all $t \geq 0$ there exist two subsets of indexes $S_X(t) \subset \{1, \ldots, N\}$ and $S_Y(t) \subset \{1, \ldots, M\}$ for which $|X_i(t)| = |Y_j(t)| = R(t)$ for all $i \in S_X(t)$ and $j \in S_Y(t)$, with one of the two subsets being possibly empty. Assuming without restriction that $S_X(t) \neq \emptyset$ for a given a time t, we get

$$\frac{d}{dt} R(t)^2 = \frac{d}{dt} |X_i|^2(t)$$

for some $i \in S_X(t)$, and therefore

$$\begin{aligned} \frac{d}{dt} R(t)^2 = & -2 \sum_{l \in C_i} m_X^l \frac{(X_i - X_l) \cdot X_i}{|X_i - X_l|} k'_{11}(|X_i - X_l|) + \\ & - 2 \sum_{j \in D_i} m_Y^j \frac{(X_i - Y_j) \cdot X_i}{|X_i - Y_j|} k'_{12}(|X_i - Y_j|). \end{aligned}$$

Notice that, while the two set of indexes $S_X(t)$ and $S_Y(t)$ are well defined for *all* times, the above identity and the estimates below are only valid almost everywhere. Both the quantities $(X_i - X_k) \cdot X_i$, $(X_i - Y_j) \cdot X_i$ are nonnegative, e. g.

$$(X_i - Y_j) \cdot X_i = |X_i|^2 - Y_j \cdot X_i \geq 0$$

due to $|Y_j| \leq R$. Hence, due to the assumption (Mon) and to $|X_i - X_k|, |X_i - Y_j| \leq 2R$ we have

$$
\begin{aligned}
&\frac{d}{dt}|X_i|^2(t) \\
&\leq -\frac{k_{11}'(2R)}{R}\sum_{l\in C_i} m_X^l (X_i - X_l)\cdot X_i - \frac{k_{12}'(2R)}{R}\sum_{j\in D_i} m_Y^j (X_i - Y_j)\cdot x_i \\
&= -\frac{k_{11}'(2R)}{R}\frac{\alpha}{\alpha}\sum_{l=1}^{N} m_X^l (X_i - X_l)\cdot X_i - \frac{k_{12}'(2R)}{R}\sum_{j=1}^{M} m_Y^j (X_i - Y_j)\cdot X_i \\
&\leq -\min\left\{\frac{K_{11}'(2R)}{\alpha R};\frac{k_{12}'(2R)}{R}\right\}\left(\alpha R^2 - \alpha X_C\cdot X_i + R^2 - Y_C\cdot X_i\right) \\
&= -\min\left\{\frac{k_{11}'(2R)}{\alpha R};\frac{k_{12}'(2R)}{R}\right\}(\alpha+1)\,R^2,
\end{aligned}
$$

and then

$$
\frac{d}{dt}R \leq -\min\left\{\frac{k_{11}'(2R)}{\alpha R};\frac{k_{12}'(2R)}{R}\right\}(\alpha+1)\,R.
$$

A similar estimate can be obtained in case the set $S_X(t)$ is empty, involving the potential K_{22} as well. Using the non-Osgood condition on the interaction potentials, one can easily see that the quantity $R(t)$ goes to zero in a finite time which depends only on the initial radius of the support R_0, (see [20, 10]).

■

We prove now that, under the additional hypothesis on the cross interaction potential

$$
|\nabla K_{12}(x)| \to 0 \text{ for } |x| \to +\infty, \tag{27}
$$

the two species produce a *two-delta separation* before the total collapse on C_M occurs. The main idea behind this fact is that, if the two species are initially separated, and if the cross-interaction kernel is weak enough at large distances, then the two species remain separated and collapse each one onto its center of mass. The two particles thus obtained will then collapse on the total center of mass C_M in finite time. Once again, a major issue is to prove that the time of *partial collapse* does not depend on the number of particles. In order to simplify the proof below, we shall also replace the assumption (N-Os) with the simpler (and more restrictive) one

(Str-Attr) $K(x) = k(|x|)$ with $k'(0^+) > 0$.

Proposition 3.3. *Assume that the admissible potentials K_{11} and K_{22} satisfy (Rad), (Mon) and (Str-Attr) and assume K_{12} is admissible and satisfies (Rad), $K_{12}(x) = k_{12}(|x|)$ and k_{12}' non-increasing, and (27). Let the initial datum for (25) be given by $X_i(0) = X_i^0$, $Y_j(0) = Y_i^0$, with masses m_X^i and m_Y^j respectively for the particles X_i and Y_j. Let X_C and Y_C denote the partial center of masses of $\mu_1^0 = \sum_i m_X^i \delta_{X_i^0}$ and $\mu_2^0 = \sum_j m_Y^j \delta_{Y_j^0}$ respectively. Then, there exist positive constants T^*, $\bar{T}$ and λ such that, if*

$$
\lambda < |X_C - Y_C|,
$$

then the unique solution to the ODE system (25) satisfies

$$
X_i(t) = C_X(t), \quad Y_j(t) = C_Y(t)
$$

for all $t \in [T^*, \bar{T}]$, or equivalently the unique gradient flow solution of (13) with initial datum

$$
\boldsymbol{\mu}_0 = (\mu_1^0, \mu_2^0)
$$

satisfies

$$\boldsymbol{\mu}_t = \left(\delta_{X_C(t)}, \delta_{Y_C(t)}\right)$$

for all $t \in [T^*, \bar{T}]$, *and moreover*

$$\boldsymbol{\mu}_t = (\delta_{C_M}, \delta_{C_M})$$

for all $t \geq \bar{T}$.

Proof. Consider the distance between the particles of a single species and the center of mass of the same species, e. g. for the X-type

$$R_X(t) = \max_{I=1,\ldots,N} |X_i - X_C|.$$

Assume the particle X_i achieves the maximum above, and compute the time evolution

$$\begin{aligned}
\frac{d}{dt}|X_i - X_C|^2(t) &= -2\sum_{k\in C_i} m_X^k \frac{(X_i - X_k)\cdot(X_i - X_C)}{|X_i - X_k|} k'_{11}(|X_i - X_k|) \\
&- 2\sum_{j\in D_i} m_Y^j \frac{(X_i - Y_j)\cdot(X_i - X_C)}{|X_i - Y_j|} k'_{12}(|X_i - Y_j|) \\
&+ 2\sum_{h=1}^{N}\sum_{j\in D_h} m_X^h m_Y^j \frac{(X_h - Y_j)\cdot(X_i - X_C)}{|X_h - Y_j|} k'_{12}(|X_h - Y_j|).
\end{aligned}$$

Grouping the terms that depends on i in the second and third summation and by considerations similar to those of the previous proposition

$$\begin{aligned}
\frac{d}{dt}|X_i - X_C|^2(t) &\leq -k'_{11}(2R_X)R_X + 2\left(1 - m_X^i\right)\sum_{j\in D_i} m_Y^j |X_i - X_C\|k'_{12}(|X_i - Y_j|)| \\
&+ 2\sum_{h\neq i}\sum_{j\in D_h} m_X^h m_Y^j |X_i - X_C\|k'_{12}(|X_h - Y_j|)|.
\end{aligned}$$

Thanks to the assumption (27), we can choose a constant $\lambda > 0$ such that

$$k'_{11}(0^+) + k'_{22}(0^+) \geq 4k'_{12}(\lambda/2). \tag{28}$$

Then, we choose

$$R_X(0) + R_Y(0) < \lambda/2$$

and

$$|X_C - Y_C| \geq 2\lambda.$$

Then, at least in a small time interval $(0, T^*)$, we have

$$\lambda < |X_C(t) - Y_C(t)|$$

and

$$\lambda > R_X(t) + R_Y(t).$$

We observe that, for all h, j,

$$\lambda \leq |X_C(t) - Y_C(t)| \leq R_X + |X_h - Y_j| + R_Y,$$

which implies

$$|X_h - Y_j| \geq \lambda - (R_X + R_Y).$$

the monotonicity assumption on k'_{12} yields

$$\frac{d}{dt}R_X^2(t) \leq -k'_{11}(0^+)R_X + 4\left(1 - m_X^i\right)R_X k'_{12}(\lambda - (R_X + R_Y))$$

and then

$$\frac{d}{dt}R_X(t) \leq -k'_{11}(0^+) + 4\left(1 - m_X^i\right)k'_{12}(\lambda - (R_X + R_Y)).$$

A similar computation holds on the Y-particles, leading to

$$\frac{d}{dt}R_Y(t) \leq -k'_{22}(0^+) + 4\left(1 - m_Y^j\right)k'_{12}(\lambda - (R_X + R_Y)).$$

Consider now the sum of the two distance $f(t) = R_X + R_Y$, which satisfies the differential inequality

$$\frac{d}{dt}f(t) \leq -c + 4k'_{12}(\lambda - f(t)),$$

with $c = k'_{11}(0^+) + k'_{22}(0^+)$. We consider the differential equation

$$\frac{d}{dt}y(t) = -c + 4k'_{12}(\lambda - y(t)).$$

Now, our choice of λ ensures that $y(t) \leq y(0)$ as long as $y(0) < \lambda/2$. Hence, as k'_{12} is a decreasing function, we get

$$\frac{d}{dt}y(t) \leq -c + 4k'_{12}(\lambda - y_0) < 0,$$

and this is true as long as $y(0) < \lambda/2$ and a fixed, small enough time interval. Therefore, $y(t) = 0$, for $t \geq \bar{t}(y_0)$ and the time $\bar{t}(y_0)$ depends continuously on y_0 and satisfies $\bar{t}(0) = 0$. By comparison principle, the assertion is proven. Please notice that blow up time does not depend on the number of particles, but only on the quantities $R_X(0)$ and $R_Y(0)$, apart from the distance $|X_C - Y_C|$.

■

Theorem 3.2. *Let K_{ij} be admissible potentials satisfying (Co), (SQ), (MS), and let $\boldsymbol{\mu}(t)$ the unique gradient flow solution to (13) with initial datum $\boldsymbol{\mu}_0 = (\mu_{0,1}, \mu_{0,2})$.*

(1) *Assume that all the potentials K_{ij} satisfy (Rad), (Mon) and (N-Os), and assume $\boldsymbol{\mu}_0$ is supported in $\bar{B}(X_C, R_0) \times \bar{B}(Y_C, R_0)$. Then, there exists T^* depending only on R_0 such that $\boldsymbol{\mu}(t) = (\delta_{C_M}, \delta_{C_M}) \quad \forall t \geq T^*$.*

(2) *Assume that K_{11} and K_{22} satisfy (Rad), (Mon) and (Str-Attr) and K_{12} satisfies (Rad) and (27). Then that there exist $0 < \bar{T} < T^*$ and $0 < R_1 < R_2$ such that, if:*
   - *$\mu_{0,1}, \mu_{0,2}$ are supported in $\bar{B}(x_C, R_1)$ and $\bar{B}(y_C, R_1)$ respectively,*
   - *$d\left(\bar{B}(x_C, R_1), \bar{B}(y_C, R_0)\right) \geq R_2$;*

   then

$$\boldsymbol{\mu}(t) = (\delta_{x_C}, \delta_{y_C}) \quad \forall t \in [\bar{T}, T^*],$$

   and $\boldsymbol{\mu}(t) = (\delta_{C_M}, \delta_{C_M}) \quad \forall t \geq T^*$.

Proof. The proof can be obtained as in [20, Theorem 4.3, Corollary 4.7]. For an arbitrary constant $\eta > 0$, approximate the initial datum with

$$\boldsymbol{\nu}_0 = \left(\sum_{i=1}^N m_x^i \delta_{x_i^0}, \sum_{j=1}^M m_y^j \delta_{y_j^0}\right)$$

with

$$\left\{\left(x_i^0, y_j^0\right)\right\}_{i,j} \in \bar{B}(x_C, R_0) \times \bar{B}(y_C, R_0)$$

such that $\mathcal{W}_{2,\alpha}^2(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0) \leq \eta$ using the stability results 3.1 and the Propositions 3.2, 3.3, we get the desired result.

■

4. Non symmetric system

In this section we address the existence and uniqueness theory of measure solutions for (4) without assuming any correlation between the cross-interaction kernels. More precisely, condition (5) does not necessarily hold in this section. Admissible kernels with possible singularities as in condition (MS) in Definition 2.2 are still of interest, as they are still expected to produce collapse of solutions in a finite time. Nevertheless, when no correlation such as (5) is assumed, the system (4) cannot be endowed with a gradient flow structure, not even at a particle level. Hence, one cannot benefit of convexity properties of the functional providing stability properties such as (20), and uniqueness of solutions becomes then a non-trivial task. However, we can adopt the JKO scheme to prove existence of solutions in a quite general set of assumptions for the potentials, possibly including kernels leading to finite time collapse. We shall perform this task in Subsection 4.1. Uniqueness of solutions can be achieved in the case of smooth kernels. This will be proven in Subsection 4.2.

4.1. **Existence theory.**

Let us consider the general system

$$\begin{cases}\partial_t\mu_1 = \operatorname{div}\left(\mu_1\nabla H_1 * \mu_1 + \mu_1\nabla K_1 * \mu_2\right)\\ \partial_t\mu_2 = \operatorname{div}\left(\mu_2\nabla H_2 * \mu_2 + \mu_2\nabla K_2 * \mu_1\right)\end{cases} \tag{29}$$

with H_i and K_i admissible potentials, H_1 and H_2 satisfying (MS), and furthermore

- (GL) H_i and K_i are globally Lipschitz on $\mathbb{R}^d$, $i = 1, 2$,
- (RK) ∇K_1 and ∇K_2 are continuous on $\mathbb{R}^2$.

We state our definition of weak measure solution for (29).

Definition 4.1. *A curve $\mu(\cdot) = (\mu_1(\cdot), \mu_2(\cdot)) : [0, +\infty) \to \mathscr{P}(\mathbb{R}^d)_2^2$ is a* weak measure solution *to (29) is, for all $\phi, \psi \in C_c^\infty(\mathbb{R}^d)$, we have*

$$\begin{aligned}\frac{d}{dt}\int\phi(x)d\mu_1(x,t) &= -\frac{1}{2}\iint\nabla H_1(x-y)\cdot(\nabla\phi(x) - \nabla\phi(y))d\mu_1(x)d\mu_1(y)\\ &\quad-\iint\nabla K_1(x-y)\cdot\nabla\phi(x)d\mu_1(x)d\mu_2(y)\\ \frac{d}{dt}\int\psi(x)d\mu_2(x,t) &= -\frac{1}{2}\iint\nabla H_2(x-y)\cdot(\nabla\psi(x) - \nabla\psi(y))d\mu_2(x)d\mu_2(y)\\ &\quad-\iint\nabla K_2(x-y)\cdot\psi(x)d\mu_2(x)d\mu_1(y).\end{aligned}$$

Please notice that the definition of weak solution 4.1 uncovers the lack of symmetry of system (29). Indeed, the cross interaction terms cannot be symmetrized as the self-interaction terms can. As a consequence of that, a notion of solution with *atoms* in case either ∇K_1 or ∇K_2 are not continuous at zero cannot be recovered straightforwardly. This fact explains the need of a slightly stronger regularity (RK) assumed for the cross-interaction kernels.

Our strategy to prove global existence of weak measure solutions for (29) relies on a *semi-implicit* version of the JKO scheme. As the system cannot be recovered as a gradient flow with respect to the Wasserstein space, we shall solve the JKO scheme by *freezing* the non symmetric part of the system. In order to perform this task, we need to introduce the following *relative interaction energy* functional. Let $\boldsymbol{\nu} \in \mathscr{P}_2(\mathbb{R}^d)^2$ be a fixed, time independent reference measure. For all $\boldsymbol{\mu} \in \mathscr{P}(\mathbb{R}^d)^2$ we set

$$\mathcal{F}[\boldsymbol{\mu}|\boldsymbol{\nu}] = \frac{1}{2}\int_{\mathbb{R}^d} H_1 * \mu_1 d\mu_1 + \int_{\mathbb{R}^d} K_1 * \nu_2 d\mu_1 + \frac{1}{2}\int_{\mathbb{R}^d} H_2 * \mu_2 d\mu_2 + \int_{\mathbb{R}^d} K_2 * \nu_1 d\mu_2.$$

In some sense, $\mathcal{F}[\boldsymbol{\mu}|\boldsymbol{\nu}]$ is the combination of an interaction energy functional and of an external conservative force field constructed via $\boldsymbol{\nu}$. We now construct the following semi-implicit JKO scheme recursively. Let $\tau > 0$ be a fixed time step, and let $\boldsymbol{\mu}_0 = (\mu_{0,1}, \mu_{0,2}) \in \mathscr{P}(\mathbb{R}^d)^2$ be a fixed initial pair of probability measures. For a given $\boldsymbol{\mu}_n^\tau \in \mathscr{P}(\mathbb{R}^d)^2$, we define the sequence $\boldsymbol{\mu}_{n+1}^\tau$ as

$$\boldsymbol{\mu}_\tau^{n+1} \in \operatorname{argmin}_{\boldsymbol{\mu}\in\mathscr{P}_2(\mathbb{R}^d)\times\mathscr{P}_2(\mathbb{R}^d)} \left\{ \frac{1}{2\tau}\mathcal{W}_2^2(\boldsymbol{\mu}_\tau^n, \boldsymbol{\mu}) + \mathcal{F}\left[\boldsymbol{\mu}|\boldsymbol{\mu}_\tau^n\right] \right\}.$$

By re-tracing the arguments in [20, Lemma 2.3, Proposition 2.5] it is very easy to prove that the above sequence is well defined. For a given choice of the sequence $\boldsymbol{\mu}_\tau^n = (\mu_{1,\tau}^n, \mu_{2,\tau}^n)$, we define the piecewise constant interpolation

$$\bar{\mu}_{i,\tau}(t) = \mu_{i,\tau}^n \quad t \in ((n-1)t, nt]\,.$$

Proposition 4.1. *Let $T > 0$. There exists an absolutely continuous curve $\bar{\mu} : [0,T] \to \mathscr{P}(\mathbb{R}^d)_2^2$ such that the family $\boldsymbol{\mu}_\tau^n$ (up to a converging subsequence) satisfies $\boldsymbol{\mu}_\tau^n \to \bar{\mu}$ as $\tau \searrow 0$ uniformly on $[0,T]$.*

Proof. Using the notation $\boldsymbol{\mu}_\tau^n = (\mu_{1,\tau}^n, \mu_{2,\tau}^n)$, $\boldsymbol{\mu}_\tau^{n+1} = (\mu_{1,\tau}^{n+1}, \mu_{2,\tau}^{n+1})$, and by definition of the minimizing scheme, we have

$$\begin{aligned}
&\frac{1}{2\tau}\mathcal{W}_2^2(\boldsymbol{\mu}_\tau^n, \boldsymbol{\mu}_\tau^{n+1}) \leq \mathcal{F}\left[\boldsymbol{\mu}_\tau^n|\boldsymbol{\mu}_\tau^n\right] - \mathcal{F}\left[\boldsymbol{\mu}_\tau^{n+1}|\boldsymbol{\mu}_\tau^n\right] \\
&= \sum_{i=1}^2 \frac{1}{2}\left(\int_{\mathbb{R}^d} H_i * \mu_{i,\tau}^n d\mu_{i,\tau}^n - \frac{1}{2}\int_{\mathbb{R}^d} H_i * \mu_{i,\tau}^{n+1} d\mu_{i,\tau}^{n+1}\right) \\
&+ \sum_{j\neq i}\left(\int_{\mathbb{R}^d} K_i * \mu_{j,\tau}^n d\mu_{i,\tau}^n - \int_{\mathbb{R}^d} K_i * \mu_{j,\tau}^n d\mu_{i,\tau}^{n+1}\right) = \sum_{i=1}^2 I_i + \sum_{i\neq j} J_{ij}.
\end{aligned} \tag{30}$$

Let us compute

$$\begin{aligned}
J_{ij} &= \int_{\mathbb{R}^d} K_i * \mu_{j,\tau}^n d\mu_{i,\tau}^n - \int_{\mathbb{R}^d} K_i * \mu_{j,\tau}^n d\mu_{i,\tau}^{n+1} = \\
&= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^{2d}} \left(K_i(x-y) - K_i(x-z)\right) d\gamma_{i,\tau}^n(y,z)\right) d\mu_{j,\tau}^n
\end{aligned}$$

for all $\gamma_{i,\tau}^n(y,z) \in \Gamma_o(\mu_{i,\tau}^n, \mu_{i,\tau}^{n+1})$. Then,

$$\begin{aligned}
|J_{ij}| &\leq \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^{2d}} \left|\bar{K}_i(x-y) - \bar{K}_i(x-z)\right| d\gamma_{i,\tau}^n(y,z)\right) d\mu_{j,\tau}^n(x) \leq \\
&\leq \operatorname{Lip}(K_i) W_2(\mu_{i,\tau}^n, \mu_{i,\tau}^{n+1}) \leq \frac{1}{4\tau} W_2^2(\mu_{i,\tau}^n, \mu_{i,\tau}^{n+1}) + C\tau
\end{aligned}$$

for some constant $C > 0$ independent of τ. Combining the previous estimate with (30) we get

$$\frac{1}{4\tau}\mathcal{W}_2^2(\boldsymbol{\mu}_\tau^n, \boldsymbol{\mu}_\tau^{n+1}) \leq \sum_{i=1}^2 \frac{1}{2}\int_{\mathbb{R}^d} H_i * \mu_{i,\tau}^n d\mu_{i,\tau}^n - \frac{1}{2}\int_{\mathbb{R}^d} H_i * \mu_{i,\tau}^{n+1} d\mu_{i,\tau}^{n+1} + C\tau.$$

Taking the sum with respect to n, we obtain a telescopic sum, and therefore

$$\begin{aligned}
&\frac{1}{4\tau}\sum_{k=m}^n \mathcal{W}_2^2(\boldsymbol{\mu}_\tau^k, \boldsymbol{\mu}_\tau^{k+1}) \\
&\leq \sum_{i=1}^2 \frac{1}{2}\int_{\mathbb{R}^d} H_i * \mu_{i,\tau}^m d\mu_{i,\tau}^m - \frac{1}{2}\int_{\mathbb{R}^d} H_i * \mu_{i,\tau}^{n+1} d\mu_{i,\tau}^{n+1} + C(n-m)\tau.
\end{aligned}$$

Similarly to [3, Section 3.3], using assumption (SQ) for the potentials H_i, we can easily obtain the uniform estimate for the interpolant $\bar{\mu}_\tau$ for $t \in ((n-1)\tau, n\tau]$ by triangulation with the initial condition μ_0

$$\mathcal{W}_2^2(\bar{\boldsymbol{\mu}}_\tau(0), \bar{\boldsymbol{\mu}}_\tau(t)) \leq \left(C(\mu_0) + \frac{1}{2}\mathcal{W}_2^2(\bar{\boldsymbol{\mu}}_\tau(0), \bar{\boldsymbol{\mu}}_\tau(t)) \right) n\tau + Cm^2\tau^2$$

Hence, the second moment of $\bar{\mu}_\tau(t)$ is uniformly bounded on compact time intervals. Consider now two times $0 \leq s < t$ with $m = \left[\frac{|t-s|}{\tau}\right]$. We similarly get

$$\mathcal{W}_2^2(\bar{\boldsymbol{\mu}}_\tau(s), \bar{\boldsymbol{\mu}}_\tau(t)) \leq Cm\tau + Cm^2\tau^2$$

Hence, we can apply the refined version of Ascoli's theorem in [3, Section 3] to obtain the uniform narrow compactness of $\bar{\mu}_\tau$ on compact time intervals.

■

We now prove that the approximating sequence $\bar{\mu}_\tau$ constructed above converges to a weak measure solution to (29). The strategy of the proof relies on the well known technique developed in [43].

Theorem 4.1. *Let $\boldsymbol{\mu}_0 \in \mathscr{P}(\mathbb{R}^d)_2^2$ be fixed. There exists an absolutely continuous curve $\mu(\cdot) : [0, +\infty) \to \mathscr{P}(\mathbb{R}^d)_2^2$ such that $\mu(0) = \mu_0$ and $\mu(t)$ is a weak measure solution to (29) in the sense of Definition 4.1. Such solution can be constructed as the limit (up to subsequences) of the approximating curve $\bar{\mu}_\tau$.*

Proof. From the minimizing property of $\boldsymbol{\mu}_\tau^{n+1}$, for all $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathscr{P}_2(\mathbb{R}^d)_2^2$ we get

$$0 \leq \frac{1}{2\tau}\mathcal{W}_2^2(\boldsymbol{\mu}_\tau^n, \boldsymbol{\mu}) + \mathcal{F}\left[\boldsymbol{\mu}|\boldsymbol{\mu}_\tau^n\right] - \frac{1}{2\tau}\mathcal{W}_2^2(\boldsymbol{\mu}_\tau^n, \boldsymbol{\mu}_\tau^{n+1}) + \mathcal{F}\left[\boldsymbol{\mu}_\tau^{n+1}|\boldsymbol{\mu}_\tau^n\right] \tag{31}$$

In order to recover the notion of weak solution for the system (29), we consider $\boldsymbol{\mu}$ to be the push forward of $\boldsymbol{\mu}_\tau^{n+1}$ via a perturbation of the identity map on each component $i = 1, 2$. More precisely, for a given $i = 1, 2$ we set

$$\begin{aligned} \boldsymbol{\mu}_i &= (T_{i,\sharp}^1 \mu_{1,\tau}^{n+1}, T_{i,\sharp}^2 \mu_{2,\tau}^{n+1}) \\ T_{i,\sharp}^j(x) &= x + \delta_{ij}\epsilon\nabla\zeta_i(x), \end{aligned}$$

where δ_{ij} is the Kronecker delta, $\epsilon > 0$ is a small constant, and $\zeta_1, \zeta_2 \in C_c^\infty(\mathbb{R}^d)$. We now evaluate separately all the contributions in (31). Let us consider for simplicity the case $i = 1$. The self interaction term involving H_2 gives a null contribution. As for the other self-interaction term, we get

$$\begin{aligned} &\frac{1}{2}\int_{\mathbb{R}^d} H_1 * \mu_1 d\mu_1 - \frac{1}{2}\int_{\mathbb{R}^d} H_1 * \mu_{1,\tau}^{n+1} d\mu_{1,\tau}^{n+1} \\ &= \frac{1}{2}\int_{\mathbb{R}^{2d}} \left(H_1(T_1^1(x) - T_1^1(y)) - H_1(x-y)\right) d\mu_{i,\tau}^{n+1}(y) d\mu_{i,\tau}^{n+1}(x) \\ &= \frac{1}{2}\int_{\mathbb{R}^{2d}} \left(H_1(x - y + \epsilon(\nabla\zeta_1(x) - \nabla\zeta_1(y))) - H_1(x-y)\right) d\mu_{i,\tau}^{n+1}(y) d\mu_{i,\tau}^{n+1}(x). \end{aligned} \tag{32}$$

Now, from the assumptions on H_1, we get

$$\frac{H_1(x - y + \epsilon(\nabla\zeta_1(x) - \nabla\zeta_1(y))) - H_1(x-y)}{\epsilon} \to \nabla H_1(x-y) \cdot (\nabla\zeta_1(x) - \nabla\zeta_1(y))$$

as $\epsilon \searrow 0$ for all $(x,y) \in \mathbb{R}^{2d}$. Since the above left hand side is uniformly bounded with respect to ϵ, by Egorov's theorem one easily gets that

$$\int_{\mathbb{R}^{2d}} \left(\frac{H_1(x-y+\epsilon(\nabla\zeta_1(x)-\nabla\zeta_1(y))) - H_1(x-y)}{\epsilon} \right) d\mu_{i,\tau}^{n+1}(y)d\mu_{i,\tau}^{n+1}(x)$$
$$\to \int_{\mathbb{R}^{2d}} \nabla H_1(x-y)\cdot(\nabla\zeta_1(x)-\nabla\zeta_1(y))d\mu_{i,\tau}^{n+1}(y)d\mu_{i,\tau}^{n+1}(x).$$

Therefore, the last term in (32) can be written as

$$\frac{\epsilon}{2}\int_{\mathbb{R}^{2d}} \nabla H_1(x-y)\cdot(\nabla\zeta_1(x)-\nabla\zeta_1(y))\, d\mu_{i,\tau}^{n+1}(y)d\mu_{i,\tau}^{n+1}(x) + o(\epsilon).$$

We now compute the term in (31) involving the cross-interaction potentials, with the above choice of $\boldsymbol{\mu}$. Once again, as the perturbation of the identity is directed only in the first component, the term involving K_2 cancels out, and we are left with the contribution

$$\begin{aligned}
&\int_{\mathbb{R}^d} K_1 * \mu_{2,\tau}^n d\mu_1 - \int_{\mathbb{R}^d} K_1 * \mu_{2,\tau}^n d\mu_{1,\tau}^{n+1}\\
&= \int_{\mathbb{R}^{2d}} \left(K_1(x+\epsilon\nabla\zeta_1(x)-y) - K_1(x-y)\right) d\mu_{2,\tau}^n(y)d\mu_{1,\tau}^{n+1}(x)\\
&= \epsilon\int_{\mathbb{R}^{2d}} \nabla K_1(x-y)\cdot\nabla\zeta_1(x)d\mu_{2,\tau}^n(y)d\mu_{2,\tau}^{n+1}(x) + o(\epsilon),
\end{aligned}$$

where the last step can be justified as before. We consider now the terms involving the Wasserstein distance. Let $\gamma_{1,\tau}^n \in \Gamma_o(\mu_{1,\tau}^n, \mu_{1,\tau}^{n+1})$. Since $\mu_2 = \mu_{2,\tau}^{n+1}$, the only contribution is given by

$$\begin{aligned}
&\frac{1}{2\tau}W_2^2(\mu_{1,\tau}^n,\mu_1) - \frac{1}{2\tau}W_2^2(\mu_{1,\tau}^n,\mu_{1,\tau}^{n+1})\\
&\quad\le \frac{1}{2\tau}\int_{\mathbb{R}^{2d}} \left(|x-y-\epsilon\nabla\zeta_1(y)|^2 - |x-y|^2\right) d\gamma_{1,\tau}^n(x,y)
\end{aligned}$$

Summing up all the contributions, dividing by ϵ and sending $\epsilon \to 0$, we obtain

$$\begin{aligned}
0 \le\ & \frac{1}{\tau}\int_{\mathbb{R}^{2d}} (x-y)\,\nabla\zeta_1(y)d\gamma_{1,\tau}^n(x,y)\\
&+ \frac{1}{2}\int_{\mathbb{R}^{2d}} \nabla H_1(x-y)\cdot(\nabla\zeta_1(x)-\nabla\zeta_1(y))d\mu_{1,\tau}^{n+1}(y)d\mu_{1,\tau}^{n+1}(x)\\
&+ \int_{\mathbb{R}^{2d}} \nabla K_1(x-y)\cdot\nabla\zeta_1(x)d\mu_{2,\tau}^n(y)d\mu_{1,\tau}^{n+1}(x).
\end{aligned}$$

Performing again the same computation with $\epsilon < 0$, we obtain in fact the equality in the above formula. Since $(x-y)\cdot\nabla\zeta(y) = \zeta(x)-\zeta(y)+o(|x-y|^2)$, using the definition of the piecewise constant interpolation $\bar{\mu}_\tau$, with $0 \le s < t$ and

$$m = \left[\frac{s}{\tau}\right]+1, \quad n = \left[\frac{t}{\tau}\right],$$

we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} \zeta_1 d\bar{\mu}_{1,\tau}(t) - \int_{\mathbb{R}^d} \zeta_1 d\bar{\mu}_{1,\tau}(s) + o(\tau)\\
&= -\frac{1}{2}\int_t^s\int_{\mathbb{R}^{2d}} \nabla H_1(x-y)\cdot(\nabla\zeta_1(x)-\nabla\zeta_1(y))d\bar{\mu}_{1,\tau}(\sigma)d\bar{\mu}_{1,\tau}(\sigma)d\sigma\\
&- \frac{1}{2}\int_t^s\int_{\mathbb{R}^{2d}} \nabla K_1 * d\bar{\mu}_{2,\tau}(\sigma)\cdot\nabla\zeta_1 d\bar{\mu}_{1,\tau}(\sigma)d\sigma.
\end{aligned}$$

We pass to the limit for $\tau \to 0$ using the fact that $\bar{\mu}_{1,\tau}$ is tight, due to Proposition 4.1, to obtain the first equation in Definition 4.1. In a similar way we can prove the second equation holds as well, and the assertion is proven.

■

4.2. **Uniqueness for smooth kernels.** In this section we restrict to the case in which all the kernels satisfy

$$H_1, H_2, K_1, K_2 \in W^{2,\infty}(\mathbb{R}^d), \qquad \sum_{i=1}^{2} (|\nabla H_i| + |\nabla K_i|) \leq C(1 + |x|). \tag{33}$$

We use a modified version of the strategy in [19], which is basically a bootstrap version of the characteristic method.

Consider a solution $\boldsymbol{\mu}_t = (\mu_{1,t}, \mu_{2,t})$ to (29), and define the integral operator Ψ, which maps a given pair of probability measures $\boldsymbol{\mu}$ into a pair of vector fields $\Psi[\boldsymbol{\mu}] = (\Psi_1[\boldsymbol{\mu}], \Psi_2[\boldsymbol{\mu}]) \in L^2(\boldsymbol{\mu})$, as follows:

$$\Psi[\boldsymbol{\mu}] := (\nabla H_1 * \mu_1 + \nabla K_1 * \mu_2, \nabla H_2 * \mu_2 + \nabla K_2 * \mu_1). \tag{34}$$

We define the system of characteristics for (29). Assume the initial condition is given by $\boldsymbol{\mu}^0 = (\mu_1^0, \mu_2^0)$. Let $\mathbf{x}^0 = (x_1^0, x_2^0)$ be fixed. Then, assuming the solution $\boldsymbol{\mu}_t$ to (29) with initial condition $\boldsymbol{\mu}^0$ is known, we define the map

$$\boldsymbol{X}_{\boldsymbol{\mu}_t}(\mathbf{x}^0, t) = (\boldsymbol{X}^1_{\boldsymbol{\mu}_t}(x_1^0, t), \boldsymbol{X}^2_{\boldsymbol{\mu}_t}(x_2^0, t))$$

as the unique solution to the Cauchy problem

$$\begin{cases} \frac{d}{dt}\boldsymbol{X}^1_{\boldsymbol{\mu}_t}(x_1^0, t) = \Psi_1\left[\boldsymbol{\mu}_t\right](\boldsymbol{X}^1_{\boldsymbol{\mu}_t}(x_1^0, t)) \\ \frac{d}{dt}\boldsymbol{X}^2_{\boldsymbol{\mu}_t}(x_2^0, t) = \Psi_2\left[\boldsymbol{\mu}_t\right](\boldsymbol{X}^2_{\boldsymbol{\mu}_t}(x_2^0, t)) \end{cases}, \tag{35}$$

with initial condition

$$\boldsymbol{X}^i_{\boldsymbol{\mu}_t}(x_i^0, 0) = x_0^i. \tag{36}$$

In view of the assumption (33), the vector field $\Psi\left[\boldsymbol{\mu}_t\right]$ is C^2 with respect to x for all $t \geq 0$, and the linear control for the gradients in (33) ensures the system (35) admits a unique global-in-time solution for all initial conditions $\mathbf{x}^0$. Moreover, for the same reason one can prove that the flow cures corresponding to the dynamical system (35) do not intersect. Hence, it is an easy exercise to prove that the solution $\boldsymbol{\mu}_t$ to (29) satisfies

$$\mu_{i,t} = \left(\boldsymbol{X}^i_{\boldsymbol{\mu}_t}(\cdot, t)\right)_\sharp \mu_i^0, \tag{37}$$

see for instance [3, Chapter 8]. Summing up, for an arbitrary weak measure solution $\boldsymbol{\mu}_t$ to (29) with initial condition $\boldsymbol{\mu}^0$, we have proven that $\boldsymbol{\mu}_t$ can be represented as (37), with the time-dependent vector field $\boldsymbol{X}_{\boldsymbol{\mu}_t}(\cdot, t)$ being the flow map of the non-autonomous system of differential equations (35) with initial conditions (36), where the velocity vector field $\Psi[\boldsymbol{\mu}] = (\Psi_1[\boldsymbol{\mu}], \Psi_2[\boldsymbol{\mu}])$ is defined via the nonlocal operator (34).

Theorem 4.2 (Stability)**.** *Assume that all the kernels H_i, K_i are C^2 and consider two initial measures $\boldsymbol{\mu}_0, \boldsymbol{\nu}_0 \in \mathscr{P}_2(\mathbb{R}^d)^2$ with compact support and the related weak measure solutions of (29) $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ respectively. Then, there exists a constant $\tilde{C} > 0$ such that*

$$\mathcal{W}_2\left(\boldsymbol{\mu}_t, \boldsymbol{\nu}_t\right) \leq e^{\tilde{C}t} \mathcal{W}_2\left(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0\right) \quad t \geq 0. \tag{38}$$

Consequently, for a given initial condition $\boldsymbol{\mu_0} \in \mathscr{P}(\mathbb{R}^d)^2$, there exists a unique weak measure solution to (29).

Proof. Using the notation above,

$$\begin{aligned}\mathcal{W}_2(\boldsymbol{\mu}_t,\boldsymbol{\nu}_t) &= \mathcal{W}_2\left((\boldsymbol{X}_{\boldsymbol{\mu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0,(\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\boldsymbol{\nu}^0\right)\\ &\le \mathcal{W}_2\left((\boldsymbol{X}_{\boldsymbol{\mu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0,(\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0\right)+\mathcal{W}_2\left((\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0,(\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\boldsymbol{\nu}^0\right).\end{aligned}\tag{39}$$

A general property of the Wasserstein distance (see e. g. [64]) states that, for all $\mu\in\mathscr{P}_2(\mathbb{R})$ and for all Borel maps Φ_1 and Φ_2, one has

$$\begin{aligned}W_2^2\left((\Phi_1)_{\#}\mu,(\Phi_2)_{\#}\mu\right) &\le \int_{\mathbb{R}^d}|\Phi_1(x)-\Phi_2(x)|^2d\mu(x)\\ &\le \|\Phi_1-\Phi_2\|^2_{L^\infty(supp(\mu))}.\end{aligned}$$

Therefore, we can estimate the first term in the right hand side of (39) as follows:

$$\mathcal{W}_2\left((\boldsymbol{X}_{\boldsymbol{\mu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0,(\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0\right)\le\|\boldsymbol{X}_{\boldsymbol{\mu}_t}(\cdot,t)-\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t)\|_{L^\infty(supp(\boldsymbol{\mu}_0))}.$$

Now, we compute

$$\begin{aligned}\frac{d}{dt}|X^i_{\boldsymbol{\mu}_t}(x_0^i,t)-X^i_{\boldsymbol{\nu}_t}(x_0^i,t)| &\le \left|\Psi_i[\boldsymbol{\mu}_t](X^i_{\boldsymbol{\mu}_t}(x_0^i,t))-\Psi_i[\boldsymbol{\nu}_t](X^i_{\boldsymbol{\nu}_t}(x_0^i,t))\right|\\ &\le \left|\Psi_i[\boldsymbol{\mu}_t](X^i_{\boldsymbol{\mu}_t}(x_0^i,t))-\Psi_i[\boldsymbol{\mu}_t](X^i_{\boldsymbol{\nu}_t}(x_0^i,t))\right|\\ &+\left|\Psi_i[\boldsymbol{\mu}_t](X^i_{\boldsymbol{\nu}_t}(x_0^i,t))-\Psi_i[\boldsymbol{\nu}_t](X^i_{\boldsymbol{\nu}_t}(x_0^i,t))\right|\\ &\le \mathrm{Lip}(\Psi_i[\boldsymbol{\mu}_t])|X^i_{\boldsymbol{\mu}_t}(x_0^i,t)-X^i_{\boldsymbol{\nu}_t}(x_0^i,t)|+\|\Psi_i[\boldsymbol{\mu}_t]-\Psi_i[\boldsymbol{\nu}_t]\|_{L^\infty}.\end{aligned}\tag{40}$$

Now, in view of assumption (33),

$$\mathrm{Lip}(\Psi_i[\boldsymbol{\mu}_t])\le\|D^2\Psi_i[\boldsymbol{\mu}_t]\|_{L^\infty}\le\max_{i=1,2}\left[\|D^2H_i\|_{L^\infty}+\|D^2K_i\|_{L^\infty}\right].\tag{41}$$

Moreover, similarly to the proof of Proposition 4.1, we can choose $\gamma_{i,t}\in\Gamma_o(\mu_{i,t},\nu_{i,t})$ and compute, for $j\neq i$,

$$\begin{aligned}|\Psi_i[\boldsymbol{\mu}_t](x)-\Psi_i[\boldsymbol{\nu}_t](x)| &= \left|\int\nabla H_i(x-y)d\mu_{i,t}(y)+\int\nabla K_i(x-y)d\mu_{j,t}(y)\right.\\ &\left.-\int\nabla H_i(x-y)d\nu_{i,t}(y)-\int\nabla K_i(x-y)d\nu_{j,t}(y)\right|\\ &\le\iint|\nabla H_i(x-y)-\nabla H_i(x-z)|\,d\gamma_{i,t}(y,z)\\ &+\iint|\nabla K_i(x-y)-\nabla K_i(x-z)|\,d\gamma_{j,t}(y,z)\\ &\le C\mathcal{W}_2(\boldsymbol{\mu}_t,\boldsymbol{\nu}_t).\end{aligned}\tag{42}$$

Therefore, we can combine the estimates in (40), (41), and (42) to obtain

$$\frac{d}{dt}|X^i_{\boldsymbol{\mu}_t}(x_0^i,t)-X^i_{\boldsymbol{\nu}_t}(x_0^i,t)|\le C|X^i_{\boldsymbol{\mu}_t}(x_0^i,t)-X^i_{\boldsymbol{\nu}_t}(x_0^i,t)|+C\mathcal{W}_2(\boldsymbol{\mu}_t,\boldsymbol{\nu}_t),$$

which can be integrated in time to obtain

$$|X^i_{\boldsymbol{\mu}_t}(x_0^i,t)-X^i_{\boldsymbol{\nu}_t}(x_0^i,t)|\le C\int_0^t e^{C(t-\tau)}\mathcal{W}_2(\boldsymbol{\mu}_\tau,\boldsymbol{\nu}_\tau)d\tau,$$

and hence

$$\mathcal{W}_2\left((\boldsymbol{X}_{\boldsymbol{\mu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0,(\boldsymbol{X}_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\boldsymbol{\mu}^0\right)\le C\int_0^t e^{C(t-\tau)}\mathcal{W}_2(\boldsymbol{\mu}_\tau,\boldsymbol{\nu}_\tau)d\tau,\tag{43}$$

where $C>0$ only depends on the interaction kernels. As for the second term on the right hand side of (39), by taking $\pi_i\in\Gamma_o(\mu_0^i,\nu_0^i)$ and

$$\gamma_i:=\left(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t)\times\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t)\right)_{\#}\pi_i\quad\Rightarrow\quad\gamma_i\in\Gamma\left((\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\mu_0^i,(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t))_{\#}\nu_0^i\right).$$

Using a standard computation, one can easily recover that

$$W_2^2\left(\left(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t)\right)_{\#}\mu_i^0,\left(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t)\right)_{\#}\mu_i^0\right)\le\int_{\mathbb{R}^{2d}}|x-y|^2d\gamma_i(x,y)$$
$$=\int_{\mathbb{R}^{2d}}|\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(x,t)-\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(y,t)|^2d\pi_i(x,y)$$
$$\le\mathrm{Lip}(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t))^2\int_{\mathbb{R}^{2d}}|x-y|^2d\pi_i(x,y)=\mathrm{Lip}(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t))^2W_2^2\left(\mu_0^i,\nu_0^i\right). \tag{44}$$

Using the Gronwall inequality and the assumption (33) on the interaction kernels, we can integrate (35) in time to compute the difference quotients for $\boldsymbol{X}_{\mu_t}(\cdot,t)$ to get the estimate

$$\mathrm{Lip}(\boldsymbol{X}^i_{\boldsymbol{\nu}_t}(\cdot,t))\le Ce^{Ct}, \tag{45}$$

with C only depending on the interaction kernels. Combining the results in (39), (43), (45), and (44), we obtain

$$\mathcal{W}_2\left(\boldsymbol{\mu}_t,\boldsymbol{\nu}_t\right)\le C\int_0^t e^{C(t-\tau)}\mathcal{W}_2(\boldsymbol{\mu}_\tau,\boldsymbol{\nu}_\tau)d\tau+e^{Ct}\mathcal{W}_2\left(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0\right).$$

By Gronwall's lemma,

$$\mathcal{W}_2\left(\boldsymbol{\mu}_t,\boldsymbol{\nu}_t\right)\le e^{Ct}\mathcal{W}_2\left(\boldsymbol{\mu}_0,\boldsymbol{\nu}_0\right).$$

■

Remark 4.1 (Particle solutions in the non symmetric case)**.** Consider the particle system

$$\begin{cases}\dot{X}_i=-\sum_{j=1}^N m_X^j\nabla H_1(X_i-X_j)-\sum_{k=1}^M m_Y^k\nabla K_1(X_i-Y_k)\\ \dot{X}_j=-\sum_{k=1}^M m_Y^k\nabla H_2(Y_j-Y_k)-\sum_{h=1}^N m_X^h\nabla K_2(Y_j-X_h)\end{cases}.$$

Under the smoothness assumption (33), it is obvious that the above system admits a unique global solution, which coincides with the unique weak measure solution provided in Theorem 4.2 after passing to empirical measures. In the more general framework of assumptions (GL) and (RK) used to prove Theorem 4.1, it is hopeless to produce a unique particle solution, as very simple cases with Peano phenomena can be produced with repulsive kernels, see [17] in the case of one species. Now, in the symmetrizable case treated in Section 3, such problem can be overcome by assuming a suitable convexity assumption, which enables to fine a unique gradient flow solution which is global in time. Here, this strategy no longer applies. However, it would be interesting to see wether examples of non-uniqueness can be produces even in case of all *attractive* kernels (still featuring a singular behaviour in their gradient which allows for non-uniqueness). We were not able to produce them, and we therefore state the following open problem: is it possible to relax significantly the assumptions in Theorem 4.2 in order to obtain a unique solution, even at level of particle solutions?

As a simple corollary to the above stability result, we prove a simple confinement property in the attractive case for the solutions to (29). Since the system is not symmetrizable and we do not have conserved quantities, we can not use standard computation as the evolution of moments. On the other hand, in the case of attractive kernels, we can use a similar procedure to the one used in the Section 3.3, i.e. to show the property for particles and then use the stability result of the previous Theorem to move to the case of general measures.

Corollary 4.1. *Assume that all the (admissible) interaction potentials are radially symmetric and attractive, and satisfy* (33)*. Let $\boldsymbol{\mu}_0\in\mathscr{P}(\mathbb{R}^d)^2$ be a supported on*

$B(0,R)\times B(0,R)$. *Then, the unique solution* $\boldsymbol{\mu}_t$ *to* (29) *satisfies*

$$\mathrm{supp}[\boldsymbol{\mu}_t]\subset B(0,R)\times B(0,R), \tag{46}$$

for all $t\geq 0$.

Proof. Let us first restrict to the case of particles

$$\boldsymbol{\mu}^0=\Big(\sum_{i=1}^{N} m_X^i\delta_{X_i},\sum_{j=1}^{M} m_Y^j\delta_{Y^j}\Big).$$

Exactly as done in Section 3.3, consider

$$R(t)=\max_{i,j}\{|X_i(t)|,|Y_j(t)|\}$$

and we are looking for the time evolution of R. First we compute

$$\frac{d}{dt}R^2(t)=\max_{i:|X_i|=R,j:|Y_j|=R}\left\{\frac{d}{dt}|X_i|^2(t);\frac{d}{dt}|Y_j|^2(t)\right\}.$$

Assuming that, for a certain, at least small, time interval the max is achieved on $|X_i|$, we have

$$\begin{aligned}\frac{d}{dt}|X_i|^2(t)&=-2\sum_{l\neq i} m_X^l\frac{(X_i-X_l)\cdot X_i}{|X_i-X_l|}h_1'(|X_i-X_l|)+\\ &-2\sum_{j=1}^{M} m_Y^j\frac{(X_i-Y_j)\cdot X_i}{|X_i-Y_j|}k_1'(|X_i-Y_j|)\leq 0\end{aligned}$$

since H_1 and K_1 are both attractive. The result for general measures can be proven via atomization by means of the stability property (38).

■

Acknowledgments

MDF acknowledges support from the CIG Marie Curie grant 'DifNonLoc, Diffusive Partial Differential Equations with Nonlocal Interaction in Biology and Social Sciences' of the University of Bath (UK). MDF also acknowledges partial support from the Ramon y Cajal Sub-programme (MICINN-RYC) of the Spanish Ministry of Science and Innovation, 'Partial Differential Equations in Biology, Medicine and Social Sciences'. Ref. RYC-2010-06412 (PI) and from the Ministerio de Ciencia y Innovación, grant MTM2011-27739-C04-02. SF acknowledges support from the Universitat Autónoma de Barcelona during his visit in 2012.

Appendix A. An illustrative toy model

Let $F:\mathbb{R}^N\times\mathbb{R}^N$ be a smooth functional and let $\alpha>0$. Consider the dynamical system

$$\begin{cases}\frac{d}{dt}X(t)=-\nabla_X F(X(t),Y(t))\\ \frac{d}{dt}Y(t)=-\alpha\nabla_Y F(X(t),Y(t))\end{cases} \tag{47}$$

Clearly, if $\alpha=1$ the above system (47) is a gradient flow of F on the standard Euclidean space $\mathbb{R}^{2N}$. If $\alpha\neq 1$, one has to slightly modify the Euclidean structure in order to produce a gradient flow. We show in a simple computation how to do that. We shall adopt the notation $\mathbf{X}=(X,Y)\in\mathbb{R}^{2N}$ with $X,Y\in\mathbb{R}^N$. We guess as a possible candidate for the new scalar product the following quantity

$$<\mathbf{X}_1,\mathbf{X}_2>_\beta=X_1\cdot X_2+\beta Y_1\cdot Y_2,\qquad \|\mathbf{X}\|_\beta^2=|X|^2+\beta|Y|^2$$

with $\beta > 0$ to be determined. Now, since all the norms in finite dimensions are equivalent, the first order Taylor expansion of F centered at a point $\mathbf{X}_0$ with increment $\mathbf{X}_1$ gives

$$F(\mathbf{X}_0+\mathbf{X}_1)-F(\mathbf{X}_0)=\nabla_X F(X_0,Y_0)\cdot X_1+\nabla_Y F(X_0,Y_0)\cdot Y_1+o(\|\mathbf{X}_1\|_\beta).$$

Now, the notion of gradient of F in the new metric space, that we denote by $\mathrm{grad}F(\mathbf{X_0})$, should be set in order to satisfy

$$F(\mathbf{X}_0+\mathbf{X}_1)-F(\mathbf{X}_0)=<\mathrm{grad}F(\mathbf{X_0}),\mathbf{X}_1>_\beta+o(\|\mathbf{X}_1\|_\beta),$$

which yields

$$\mathrm{grad}F(\mathbf{X_0})=(\nabla_X F(X_0,Y_0),\frac{1}{\beta}\nabla_Y F(X_0,Y_0)).$$

Therefore, the correct choice for interpreting (47) as a gradient flow of F in the new metric space $(\mathbb{R}^{2N},\|\cdot\|_\beta)$ is

$$\beta=\frac{1}{\alpha}.$$

Notice that the constant speed geodesics are linear convex combinations of points no matter what α is.

Appendix B. A finite dimensional gradient flow structure

Consider the particles $X_1,\dots,X_N$ with masses $n_1,\dots,n_N$, and $Y_1,\dots,Y_M$ with masses $m_1,\dots,m_M$. Assuming the (admissible) potentials K_{ij} satisfy conditions (Co) and (MS) above, we want to prove the global-in-time existence of solutions for the ODE system

$$\begin{cases}\dot{X}_i=-\sum_{k\in C_i}m_X^k\nabla K_{11}(X_i-X_k)-\sum_{j\in D_i}m_Y^j\nabla K_{12}(X_i-Y_j) & i=1,..,N\\ \dot{Y}_j=-\sum_{h\in E_j}m_Y^h\nabla K_{22}(Y_j-Y_h)-\sum_{i\in F_j}m_X^i\nabla K_{12}(Y_j-X_i) & j=1,..,M,\end{cases}\tag{48}$$

with $C_i=\{k\in\{1,\dots,N\}:|X_k\neq X_i\}$, $D_i=\{j\in\{1,\dots,M\}:|Y_j\neq X_i\}$, $E_j=\{h\in\{1,\dots,M\}:|Y_j\neq Y_h\}$, $F_j=\{i\in\{1,\dots,N\}:|Y_j\neq X_i\}$. To perform this task, we consider the set of masses n_i and m_j to be fixed, and we introduce the weighted, finite dimensional Hilbert space

$$\mathcal{P}_w^2:=\left(\mathbb{R}^{dN}\times\mathbb{R}^{dM},<\cdot,\cdot>_w\right),$$

with the notation $\mathbf{X}=(X,Y)\in\mathbb{R}^{dN}\times\mathbb{R}^{dM}$, $X=(X_1,\dots,X_N)$, $Y=(Y_1,\dots,Y_M)$, and

$$<\mathbf{X}^1,\mathbf{X}^2>_w=\sum_{i=1}^N n_iX_i^1\cdot X_i^2+\sum_{j=1}^M m_jY_j^1\cdot Y_j^2.$$

We then consider the discrete interaction energy functional

$$\mathcal{F}[\mathbf{X}]=\frac{1}{2}\sum_{i\neq j}n_in_jK_{11}(X_i-X_j)+\frac{1}{2}\sum_{i\neq j}m_im_jK_{22}(Y_i-Y_j)+\sum_{i\neq j}m_in_jK_{12}(Y_i-X_j).$$

It is left as exercise to prove that the functional $\mathcal{F}$ is λ convex on the Hilbert space $\mathcal{P}_w^2$ in the usual sense. Moreover, we also leave as an exercise to prove that the Frechet minimal sub-differential $\partial^0\mathcal{F}[\mathbf{X}]$ of $\mathcal{F}$ at a point $\mathbf{X}$ on the space $\mathcal{P}_w^2$ is given by the right-hand side of the system (48), so that (48) can be reformulated as

$$\dot{\mathbf{X}}(t)=\partial^0\mathcal{F}[\mathbf{X}].$$

Hence, the classical result in [16, Theorem 3.17] ensures the existence of a global-in-time solution for the ODE system (48) which is almost everywhere defined in time.

References

[1] G. Aletti, G. Naldi, and G. Toscani. First-order continuous models of opinion formation. *SIAM J. Appl. Math.*, 67(3):837–853 (electronic), 2007.

[2] D. Ambrosi, F. Bussolino, and L. Preziosi. A review of vasculogenesis models. *J. Theor. Med.*, 6(1):1–19, 2005.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[4] L. Ambrosio and G. Savaré. Gradient flows of probability measures. In *Handbook of differential equations: evolutionary equations. Vol. III*, Handb. Differ. Equ., pages 1–136. Elsevier/North-Holland, Amsterdam, 2007.

[5] C. Appert-Rolland, P. Degond, and S. Motsch. Two-way multi-lane traffic model for pedestrians in corridors. *Netw. Heterog. Media*, 6(3):351–381, 2011.

[6] J. Bedrossian and N. Masmoudi. Existence, uniqueness and lipschitz dependence for patlak-keller-segel and navier-stokes in r^2 with measure-valued initial data. 2012.

[7] E. Ben-Naim, P. L. Krapivsky, F. Vazquez, and S. Redner. Unity and discord in opinion dynamics. *Phys. A*, 330(1-2):99–106, 2003. Randomness and complexity (Eilat, 2003).

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

[9] A. L. Bertozzi and J. Brandman. Finite-time blow-up of L^∞-weak solutions of an aggregation equation. *Commun. Math. Sci.*, 8(1):45–65, 2010.

[10] A. L. Bertozzi, J. A. Carrillo, and T. Laurent. Blow-up in multidimensional aggregation equations with mildly singular interaction kernels. *Nonlinearity*, 22(3):683–710, 2009.

[11] A. L. Bertozzi and T. Laurent. The behavior of solutions of multidimensional aggregation equations with mildly singular interaction kernels. *Chin. Ann. Math. Ser. B*, 30(5):463–482, 2009.

[12] A. L. Bertozzi, T. Laurent, and J. Rosado. L^p theory for the multidimensional aggregation equation. *Comm. Pure Appl. Math.*, 64(1):45–83, 2011.

[13] G. Biler, P.and Karch and P. Laurençot. Blowup of solutions to a diffusive aggregation model. *Nonlinearity*, 22(7):1559–1568, 2009.

[14] 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.

[15] S. Boi, V. Capasso, and D. Morale. Modeling the aggregative behavior of ants of the species *polyergus rufescens*. *Nonlinear Anal. Real World Appl.*, 1(1):163–176, 2000. Spatial heterogeneity in ecological models (Alcalá de Henares, 1998).

[16] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[17] M. Burger and M. Di Francesco. Large time behavior of nonlocal aggregation models with nonlinear diffusion. *Netw. Heterog. Media*, 3(4):749–785, 2008.

[18] M. Burger, M. Di Francesco, and M. Franek. Stationary states of quadratic diffusion equations with long-range attraction. 2012.

[19] J. A. Cañizo, J. A. Carrillo, and J. Rosado. A well-posedness theory in measures for some kinetic models of collective motion. *Math. Models Methods Appl. Sci.*, 21(3):515–539, 2011.

[20] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.*, 156(2):229–271, 2011.

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

[22] J.A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions, and finite-time aggregation for nonlocal interaction equations. *Duke Mathematical Journal*, 156:229–271, 2011.

[23] R. M. Colombo, M. Garavello, and M. Lécureux-Mercier. A class of nonlocal models for pedestrian traffic. *Math. Models Methods Appl. Sci.*, 22(4):1150023, 34, 2012.

[24] R. M. Colombo and M. Lécureux-Mercier. Nonlocal crowd dynamics models for several populations. *Acta Math. Sci. Ser. B Engl. Ed.*, 32(1):177–196, 2012.

[25] C. Conca, E. Espejo, and K. Vilches. Remarks on the blowup and global existence for a two species chemotactic Keller-Segel system in $\mathbb{R}^2$. *European J. Appl. Math.*, 22(6):553–580, 2011.

[26] G. Crippa and M. Lécureux-Mercier. Existence and uniqueness of measure solutions for a system of continuity equations with non-local flow. 2011.

[27] E. Cristiani, P. Frasca, and B. Piccoli. Effects of anisotropic interactions on the structure of animal groups. *J. Math. Biol.*, 62(4):569–588, 2011.

[28] R. L. Dobrušin. Vlasov equations. *Funktsional. Anal. i Prilozhen.*, 13(2):48–58, 96, 1979.

[29] J. Dolbeault and C. Schmeiser. The two-dimensional Keller-Segel model after blow-up. *Discrete Contin. Dyn. Syst.*, 25(1):109–121, 2009.

[30] B. Düring, P. Markowich, J.-F. Pietschmann, and M.-T. Wolfram. Boltzmann and Fokker-Planck equations modelling opinion formation in the presence of strong leaders. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 465(2112):3687–3708, 2009.

[31] C. Escudero, F. Macià, and J. J. L. Velázquez. Two-species-coagulation approach to consensus by group level interactions. *Phys. Rev. E (3)*, 82(1), 2010.

[32] E. E. Espejo, A. Stevens, and J. J. L. Velázquez. Simultaneous finite time blow-up in a two-species model for chemotaxis. *Analysis (Munich)*, 29(3):317–338, 2009.

[33] E. E. Espejo, A. Stevens, and J. J. L. Velázquez. A note on non-simultaneous blow-up for a drift-diffusion model. *Differential Integral Equations*, 23(5-6):451–462, 2010.

[34] F. Golse. The mean-field limit for the dynamics of large particle systems. In *Journées "Équations aux Dérivées Partielles"*, pages Exp. No. IX, 47. Univ. Nantes, Nantes, 2003.

[35] M. Z. Guo, G. C. Papanicolaou, and S. R. S. Varadhan. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Commun. Math. Phys.*, 118:31–59, 1988.

[36] D. Helbing, I. J. Farkas, P. Molnar, and T. Vicsek. Simulation of pedestrian crowds in normal and evacuation situations. *in: M. Schreckenberg and S. D. Sharma (eds.)* Pedestrian and Evacuation Dynamics *(Springer, Berlin).*, pages 21–58, 2002.

[37] D. Helbing, A. Johansson, and H. Z. Al-Abideen. The dynamics of crowd disasters: An empirical study. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 75(4), 2007.

[38] D. Helbing, W. Yu, and H. Rauhut. Self-organization and emergence in social systems: modeling the coevolution of social environments and cooperative behavior. *J. Math. Sociol.*, 35(1-3):177–208, 2011.

[39] Dirk Helbing. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.*, 73(4):1067–1141, Dec 2001.

[40] D. Horstmann. Generalizing the Keller-Segel model: Lyapunov functionals, steady state analysis, and blow-up results for multi-species chemotaxis models in the presence of attraction and repulsion between competitive interacting species. *J. Nonlinear Sci.*, 21(2):231–270, 2011.

[41] D. Horstmann and M. Lucia. Nonlocal elliptic boundary value problems related to chemotactic movement of mobile species. In *Mathematical analysis on the self-organization and self-similarity*, RIMS Kôkyûroku Bessatsu, B15, pages 39–72. Res. Inst. Math. Sci. (RIMS), Kyoto, 2009.

[42] W. Jäger and S. Luckhaus. On explosions of solutions to a system of partial differential equations modelling chemotaxis. *Trans. Amer. Math. Soc.*, 329(2):819–824, 1992.

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

[44] I. Josek. Meinungsdynamik mit heterogenen agenten. Diploma thesis, Institut für Numerische und Angewandte Mathematik, WWU Münster, 2009. (Engl. translation: Opinion Dynamics with Heterogeneous Agents).

[45] E. F. Keller and L. A. Segel. Initiation of slide mold aggregation viewed as an instability. *J Theor Biol*, 26:399–415, 1970.

[46] H. Li and G. Toscani. Long-time asymptotics of kinetic models of granular flows. *Arch. Ration. Mech. Anal.*, 172(3):407–428, 2004.

[47] S. Luckhaus, Y. Sugiyama, and J. J. L. Velázquez. Measure Valued Solutions of the 2D Keller–Segel System. *Arch. Ration. Mech. Anal.*, 206(1):31–80, 2012.

[48] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *J. Math. Biol.*, 38(6):534–570, 1999.

[49] A. Mogilner, L. Edelstein-Keshet, L. Bent, and A. Spiros. Mutual interactions, potentials, and individual distance in a social aggregation. *J. Math. Biol.*, 47(4):353–389, 2003.

[50] C. B. Morrey, Jr. On the derivation of the equations of hydrodynamics from statistical mechanics. *Comm. Pure Appl. Math.*, 8:279–326, 1955.

[51] K. Oelschläger. A sequence of integro-differential equations approximating a viscous porous medium equation. *Z. Anal. Anwendungen*, 20(1):55–91, 2001.

[52] A. Okubo and S. A. Levin. *Diffusion and ecological problems: modern perspectives*, volume 14 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2001.

[53] L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev. (2)*, 65:117–149, 1944.

[54] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[55] C.S. Patlak. Random walk with persistence and external bias. *Bull. Math. Biophys.*, 15:311–338, 1953.
[56] F. Poupaud. Diagonal defect measures, adhesion dynamics and Euler equation. *Methods Appl. Anal.*, 9(4):533–561, 2002.
[57] L. Preziosi and L. Graziano. Multiphase models of tumor growth: general framework and particular cases. In *Mathematical modelling & computing in biology and medicine*, volume 1 of *Milan Res. Cent. Ind. Appl. Math. MIRIAM Proj.*, pages 622–628.
[58] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *Int. J. Mod. Phys. C*, 11:1157–1166, 2000.
[59] J. I. Tello and M. Winkler. Stabilization in a two-species chemotaxis system with a logistic source. *Nonlinearity*, 25(5):1413–1425, 2012.
[60] C. M. Topaz and A. L. Bertozzi. Swarming patterns in a two-dimensional kinematic model for biological groups. *SIAM J. Appl. Math.*, 65(1):152–174, 2004.
[61] G. Toscani. Kinetic and hydrodynamic models of nearly elastic granular flows. *Monatsh. Math.*, 142(1-2):179–192, 2004.
[62] G. Toscani. Kinetic models of opinion formation. *Commun. Math. Sci.*, 4(3):481–496, 2006.
[63] S. R. S. Varadhan. Scaling limits for interacting diffusions. *Commun. Math. Phys.*, 135:313–353, 1991.
[64] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[65] G. Wolansky. Multi-components chemotactic system in the absence of conflicts. *European J. Appl. Math.*, 13(6):641–661, 2002.

Marco Di Francesco - Department of Mathematical and Statistical Sciences, University of Bath, Claverton Down, Bath (UK), BA2 7AY
E-mail address: m.difrancesco@bath.ac.uk

Simone Fagioli - DISIM - Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Via Vetoio 1 (Coppito) 67100 L'Aquila (AQ) - Italy
E-mail address: simone.fagioli@dm.univaq.it