Leader formation with mean-field birth and death models

G. Albi [*], M. Bongini [†], F. Rossi, [‡]F. Solombrino[§]

Abstract

We provide a mean-field description for a leader-follower dynamics with mass transfer among the two populations. This model allows the transition from followers to leaders and vice versa, with scalar-valued transition rates depending nonlinearly on the global state of the system at each time.

We first prove the existence and uniqueness of solutions for the leader-follower dynamics, under suitable assumptions. We then establish, for an appropriate choice of the initial datum, the equivalence of the system with a PDE-ODE system, that consists of a continuity equation over the state space and an ODE for the transition from leader to follower or vice versa. We further introduce a stochastic process approximating the PDE, together with a jump process that models the switch between the two populations. Using a propagation of chaos argument, we show that the particle system generated by these two processes converges in probability to a solution of the PDE-ODE system. Finally, several numerical simulations of social interactions dynamics modeled by our system are discussed.

1 Introduction

The mathematical modeling of collective behavior for systems of interacting agents has spawned an enormous wealth of literature in recent years. From the study of biological, social and economical phenomena [1, 2, 9, 22] to automatic learning [14, 35] and optimization heuristics [25, 36], these models lay at the heart of some of today's most prominent lines of research: for the latest development in the field, we point to the surveys [8, 13, 18, 48] and references therein.

The modeling of such phenomena typically starts from particle-like systems as in statistical physics. These particle models are also called Agent Based Models, and they usually consist of a set of ODEs (one for each agent) interwined in a nonlinear way. Such a modeling approach is quite useful, with one of the main advantages being the

[*] Department of Computer Science, University of Verona,Str. Le Grazie 15, Verona, IT-37134, Italy. giacomo.albi@univr.it

[†] Big Data and Marketing Analytics, CRIF, Via M. Fantin 3, Bologna, IT-40131, Italy. mattia.bongini@crif.com

[‡] Department of Mathematics "Tullio Levi-Civita", University of Padova, Via Trieste, 63, Padova, IT-35121, Italy. francesco.rossi@math.unipd.it

[§] Department of Mathematics and Applications "R. Caccioppoli", University of Naples "Federico II" Via Cintia, Monte S. Angelo - IT-80126 Naples, Italy. francesco.solombrino@unina.it

explicit description of the mutual interaction among agents, but has huge problems to treat large systems of particles, as is the case with cells, molecules and social networks' users. A classical approach to attack the problem is to pass to a continuous description of the system, which means to pass from a particle description to a kinetic descriptions where the unknown is the particle density distribution in the state space.

A useful tool in solving this problem is the *mean-field* limit [33], which amounts to replace the influence of all the other individuals in the dynamics of any given agent by a single averaged effect, a technique that goes back to [24] in plasma physics: to exemplify it, if applied to a Hegselmann-Krause-type discrete particle system over $\mathbb{R}^d$ (see [34])

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1}^{N} K(x_i(t) - x_j(t)), \qquad i = 1, \ldots, N \text{ and } t \in [0, T]$$

where K denotes the *interaction kernel* (which models the interaction between particles) it leads to a continuity equation of Vlasov type

$$\partial_t \mu_t = -\operatorname{div}((K * \mu_t)\mu_t), \qquad t \in [0, T]$$

with μ_t denoting the probability distribution of the particles over the state space $\mathbb{R}^d$ and

$$(K * \mu_t)(x) = \int_{\mathbb{R}^d} K(x - y)\, d\mu_t(y), \qquad \text{for every } x \in \mathbb{R}^d.$$

Notice how, in the process, the information of the pointwise positions $x_j(t)$ is replaced by the knowledge of the space distribution of the particles μ_t. Such approach has the advantage of reducing the computational complexity of the models (overcoming the curse of dimensionality [10]) and allows the so-called *microfundation of macromodels*, i.e., the validation of the macroscopic dynamics from the coherence with the behavior of individuals (a central issue in the field of macroeconomics). The mean-field limit of systems of interacting agents has been thoroughly studied also in conjunction with irregular interaction kernel [17, 32], control problems [3, 15, 29, 30, 38] and multiple populations [4, 5, 12, 21]. Also models where the total mass of the system is not preserved in time, due to the presence of source (or sink) terms, have been considered (see for instance [45, Sections 4-5]). In other models, the total mass of the system is preserved, but not the role of the agents, since exchanges of mass between different populations are allowed. One of these models is the leader-follower dynamics studied in [19, 27], given by

$$\begin{cases} \partial_t \mu_t^F = -\operatorname{div}\big((K^F * \mu_t^F + K^L * \mu_t^L)\mu_t^F\big) \\ \qquad\qquad - \alpha_F(\mu_t^F, \mu_t^L)\mu_t^F + \alpha_L(\mu_t^F, \mu_t^L)\mu_t^L, \\ \partial_t \mu_t^L = -\operatorname{div}\big((K^F * \mu_t^F + K^L * \mu_t^L)\mu_t^L\big) \\ \qquad\qquad + \alpha_F(\mu_t^F, \mu_t^L)\mu_t^F - \alpha_L(\mu_t^F, \mu_t^L)\mu_t^L. \end{cases} \quad t \in [0, T]. \tag{1}$$

Here, two competing populations μ_t^F and μ_t^L, of followers and leaders respectively, are in interaction. Both the masses of followers and leaders vary in time, while their sum is constant. The functionals $K^i : \mathbb{R}^d \to \mathbb{R}^d$ for $i \in \{F, L\}$ are interaction kernels, modeling their mutual spatial influence, while the transition rates $\alpha_F, \alpha_L : \mathcal{M}(\mathbb{R}^d) \times \mathcal{M}(\mathbb{R}^d) \to [0, +\infty)$ govern the exchange of mass between μ_t^F and μ_t^L. In this paper, we shall provide a thorough mean-field analysis of (1), discussing its well-posedness and rigorously deriving it from an Agent Based Model. In order to do this, we will restrict our attention to the case where the transition rates α_i are scalar-valued, that is they depend on the global state of the system at each time t, but not explicitly on the position x. The usefulness of such a simplification in our analysis is discussed later on in Remark 2.7.

In order to carry out our analysis, we shall first establish the well-posedness of (1) by means of a compactness argument in the space of finite positive measures with compact support endowed with the *generalized Wasserstein distance* $\mathcal{W}_g$, see [43]. To do so, we shall introduce an explicit Euler approximation of the dynamics and show that it converges, as the time step vanishes, to the unique solution of system (1).

We shall then prove the equivalence between (1) and another system for which we can more easily provide a particle dynamics. Intuitively, this equivalent system is introduced by defining the measures (ν_t, σ_t) as[1]

$$\nu_t := \mu_t^F + \mu_t^L, \qquad (\sigma_t(F), \sigma_t(L)) := (\mu_t^F(\mathbb{R}^d)/\nu_t(\mathbb{R}^d), \mu_t^L(\mathbb{R}^d)/\nu_t(\mathbb{R}^d)).$$

The idea of the equivalence is that, under suitable hypotheses on the initial data, one can recover μ_t^F, μ_t^L from ν_t, σ_t by the relations

$$\mu_t^F = \sigma_t(F)\nu_t, \qquad \mu_t^L = \sigma_t(L)\nu_t. \tag{2}$$

We shall show that, if the initial datum μ_0^F, μ_0^L satisfies (2) for $t = 0$, the system (1) is equivalent to

$$\begin{cases} \partial_t \nu_t = -\mathrm{div}\Big(\langle K, \nu_t \times \sigma_t \rangle \, \nu_t \Big), \\ \partial_t \sigma_t = A_{\nu_t, \sigma_t} \sigma_t, \end{cases} \qquad t \in [0, T], \tag{3}$$

where $\nu_t \times \sigma_t$ is the product measure, the vector field for ν_t is

$$\langle K, \nu_t \times \sigma_t \rangle := \sigma_t(F) K^F * \nu_t + \sigma_t(L) K^L * \nu_t, \tag{4}$$

and the birth-death transition matrix is

$$A_{\nu_t, \sigma_t} := \begin{bmatrix} -\alpha_F(\sigma_t(F)\nu_t, \sigma_t(L)\nu_t) & \alpha_L(\sigma_t(F)\nu_t, \sigma_t(L)\nu_t) \\ \alpha_F(\sigma_t(F)\nu_t, \sigma_t(L)\nu_t) & -\alpha_L(\sigma_t(F)\nu_t, \sigma_t(L)\nu_t) \end{bmatrix}. \tag{5}$$

The advantage of the measures ν_t and σ_t, with respect to μ_t^F and μ_t^L, is that they are probability measuers over $\mathbb{R}^d$ and $\{F, L\}$, respectively, and we can therefore use a propagation of chaos argument (see [46]) to show that there exists a sequence of stochastic

[1] With a slight abuse of notation, from now on we write $\sigma(F), \sigma(L)$ instead of $\sigma(\{F\}), \sigma(\{L\})$.

processes whose mean-field limit for $N \to \infty$ is system (3). We will actually provide such processes in explicit form: we denote them as $(X_t^{1,N}, Y_t^{1,N}), \ldots, (X_t^{N,N}, Y_t^{N,N})$, where for every $t \in [0,T]$ and $i = 1, \ldots, N$ we have $(X_t^{i,N}, Y_t^{i,N}) \in \mathbb{R}^d \times \{F, L\}$. Setting

$$\nu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}} \quad \text{and} \quad \sigma_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^{i,N}},$$

then their dynamics is given by

- $dX_t^{i,N} = \langle K, \nu_t^N \times \sigma_t^N \rangle (X_t^{i,N}) dt\,,$
- $Y_t^{i,N}$ obeys a jump process, with conditional transition rates for the realization of (ν_t^N, σ_t^N) at time t, given by
  - if $Y_t^{i,N} = F$ then $F \to L$ with rate $\alpha_F(\nu_t^N, \sigma_t^N)$,
  - if $Y_t^{i,N} = L$ then $L \to F$ with rate $\alpha_L(\nu_t^N, \sigma_t^N)$.

By virtue of the equivalence between (1) and (3), the mean-field limit for $N \to \infty$ of the above Agent Based Model is system (1).

The final part of our paper is devoted to numerical implementations of (1). Three model applications are considered:

- consensus dynamics for two populations μ_t^F, μ_t^L with a bounded confidence interaction kernel of Hegselmann-Krause type;
- aggregation dynamics with competition among repulsive followers μ_t^F, and attractive leaders μ_t^L;
- the problem of steering a population towards a desired position via leaders' action.

In the case of consensus we compare the effect of suitably chosen density-dependent birth and death rates, allowing the system to reach consensus, with constant ones, where instead the system ends up clustering around different states.

For the second case, observe that aggregation models are used to describe several biological phenomena, but also as building brick of social interactions such as crowd motion [11, 23]. We show that a controlled generation of leaders, with attraction kernel, is able to confine the whole density, balancing the repulsiveness of followers' interactions.

In the third case, we study the case where leaders' generation is conditioned to achievement of a desired position, in analogy with control problem for pedestrian dynamics [2, 16]. Thus leaders' motion influences the followers' density towards a specific goal, whereas followers' interactions are ruled by an aggregation equation. We show that the whole population is steered to the desired state, with the leaders' mass diminishing, and eventually vanishing, as soon as the followers are sufficiently close to the final state.

As a final remark we observe that most of our results can be straightforwardly extended to the case of a *finite hierarchy* of labels $\{L_1, \ldots, L_n\}$ instead of $\{F, L\}$ with transitions given by

$$L_1 \leftrightarrow L_2 \leftrightarrow \ldots \leftrightarrow L_n \leftrightarrow L_1.$$

All the proofs would follow along the same lines, though at the expense of notation. Actually, we conjecture that the results of the paper hold true even in the case of a countable number of labels $\{L_k\}_{k\in\mathbb{N}}$, as treated in a simplified scenario in [47].

A further issue, which falls for the moment outside the scope of our methods, and is likely to require a finer analysis, is the mean-field derivation of system (1) in the case where the birth rates take values in a functional space, for example when they explicitly depend on the position x. We plan to address these aspects in future contributions.

The structure of the paper is the following. After discussing some measure-theoretical preliminaries in Section 2, we turn our attention to system (1). We introduce a general set of assumptions and prove the existence and uniqueness of solutions, using explicit Euler approximations of the dynamics and a compactness argument in the space of positive measures with bounded mass and compact support, endowed with the generalized Wasserstein distance $\mathcal{W}_g$. This is done in Section 3, where we also establish a bijection between solutions of (1) and of (3) under certain assumptions on the initial data (Proposition 3.4). In Section 4 we derive system (3) as mean-field limit of a particle system which couples a SDE (66) for the particles' motion with a nonlinear master equation (67) for their labels. Section 5 is devoted to numerical experiments, which make use of the finite volume scheme discussed in B. In A we introduce some explicit examples of transition functionals which comply with our assumptions and are indeed used in the experiments of Section 5.

2 Preliminaries

Let X be a Radon space; we denote by $\mathcal{M}(X)$ the set of finite positive measures on X, and by $\mathcal{M}_c(X)$ the subset of finite positive measures with compact support. The space $\mathcal{P}(X)$ is the subset of $\mathcal{M}(X)$ whose elements are the probability measures on X, i.e., those $\mu \in \mathcal{M}(X)$ for which $\mu(X) = 1$. The space $\mathcal{P}_p(X)$ is the subset of $\mathcal{P}(X)$ whose elements have finite p-th moment, i.e.,

$$\int_X |x|^p \, d\mu(x) < +\infty.$$

We denote by $\mathcal{P}_c(X)$ the subset of $\mathcal{P}(X)$ which consists of all probability measures with compact support. We denote the mass of a measure as $|\mu| = \mu(X)$.

If X_1 and X_2 are Radon spaces, for any[2] $\mu \in \mathcal{M}(X_1)$ and any Borel function $f : X_1 \to X_2$, we denote by $f\#\mu \in \mathcal{M}(X_2)$ the *push-forward of* μ *through* f, defined by

$$f\#\mu(E) := \mu(f^{-1}(E)) \qquad \text{for every Borel set } E \text{ of } X_2.$$

In particular, if one considers the projection operators π_1 and π_2 defined on the product space $X_1 \times X_2$, for every $\rho \in \mathcal{P}(X_1 \times X_2)$ we call *first* (resp., *second*) *marginal* of ρ the probability measure $\pi_1\#\rho$ (resp., $\pi_2\#\rho$). Given $\mu \in \mathcal{P}(X_1)$ and $\nu \in \mathcal{P}(X_2)$, we denote

[2] more in general, also if μ is a signed Borel measure on X_1

by $\Gamma(\mu,\nu)$ the subset of all probability measures in $\mathcal{P}(X_1 \times X_2)$ with first marginal μ and second marginal ν.

We denote the weak convergence of measures as follows:

$$\mu_n \rightharpoonup \mu \qquad \text{when} \qquad \text{for all } f \in \mathcal{C}_c^\infty(\mathbb{R}^d) \text{ it holds} \int f\,d\mu_n \to \int f\,d\mu.$$

2.1 The Wasserstein distance

In this section, we recall the definition of Wasserstein distance, as well as some of its useful properties.

Definition 2.1 (Wasserstein distance)**.** For every $\mu, \nu \in \mathcal{P}_p(X)$ we define

$$\mathcal{W}_p(\mu,\nu) := \inf\left\{ \int_{X^2} |x-y|^p \, d\rho(x,y) \;:\; \rho \in \Gamma(\mu,\nu) \right\}^{1/p}. \tag{6}$$

Remark 2.2. We denote by $\Gamma_o(\mu,\nu)$ the set of optimal plans for which the infimum in (6) is attained, i.e.,

$$\rho \in \Gamma_o(\mu,\nu) \iff \rho \in \Gamma(\mu,\nu) \text{ and } \int_{\mathbb{R}^{2d}} |x-y|^p \, d\rho(x,y) = \mathcal{W}_p^p(\mu,\nu).$$

It is well-known that $\Gamma_o(\mu,\nu)$ is non-empty for every $(\mu,\nu) \in \mathcal{P}_p(X) \times \mathcal{P}_p(X)$, hence the infimum in (6) is actually a minimum, see [49].

Remark 2.3. Under suitable conditions (see [49, Theorem 5.9]), by Kantorovich-Rubinstein duality we have

$$\mathcal{W}_1(\mu,\nu) = \sup\left\{ \int_X \varphi(x) d(\mu-\nu)(x) : \mathrm{Lip}(\varphi) \le 1 \right\}. \tag{7}$$

In analogy with $\Gamma_o(\mu,\nu)$, we denote by $\Lambda(\mu,\nu)$ the set of Lipschitz maps $\varphi : X \to \mathbb{R}$ with $\mathrm{Lip}(\varphi) \le 1$, and by $\Lambda_o(\mu,\nu)$ the subset of $\Lambda(\mu,\nu)$ for which the above supremum is attained, i.e.,

$$\varphi \in \Lambda_o(\mu,\nu) \iff \varphi \in \Lambda(\mu,\nu) \text{ and } \int_X \varphi(x) d(\mu-\nu)(x) = \mathcal{W}_1(\mu,\nu).$$

Then, by [49, Theorem 5.9], it follows that $\Lambda_o(\mu,\nu)$ is non-empty.

We finally recall the following result, see e.g. [49].

Proposition 2.4. *Wasserstein distances are ordered, in the sense that $1 \le p_1 \le p_2$ implies*

$$W_{p_1}(\mu,\nu) \le W_{p_2}(\mu,\nu).$$

2.2 Solutions of transport equations

We now recall the precise definition of solutions to systems (1) and (3). Indeed, a solution of system (1) must be interpreted in the sense of distributions, as follows.

Definition 2.5 (Solution of system (1)). Let $(\overline{\mu}^F, \overline{\mu}^L) \in \mathcal{M}_c(\mathbb{R}^d) \times \mathcal{M}_c(\mathbb{R}^d)$ be given, as well as $\mu^F, \mu^L : [0,T] \to \mathcal{M}_c(\mathbb{R}^d)$. We say that the couple (μ_t^F, μ_t^L) is a solution of system (1) with initial datum $(\overline{\mu}^F, \overline{\mu}^L)$ when

1. $\mu_0^F = \overline{\mu}^F$ and $\mu_0^L = \overline{\mu}^L$;

2. for each $i \in \{F, L\}$, the function $t \to \mu_t^i$ is continuous with respect to the topology of weak convergence of measures;

3. there exists $R_T > 0$ such that $\bigcup_{t\in[0,T]} \mathrm{supp}(\mu_t^i) \subseteq B(0, R_T)$ for every $i \in \{F, L\}$;

4. for every $\varphi \in \mathcal{C}_c^1(\mathbb{R}^d)$ and $i \in \{F, L\}$ it holds

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(x) d\mu_t^i(x) = \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot \left[\sum_{j\in\{F,L\}} (K^j * \mu_t^j)(x)\right] d\mu_t^i(x) \\ - \alpha_i(\mu_t^F, \mu_t^L) \int_{\mathbb{R}^d} \varphi(x) d\mu_t^i(x) + \alpha_{\neg i}(\mu_t^F, \mu_t^L) \int_{\mathbb{R}^d} \varphi(x) d\mu_t^{\neg i}(x),$$

for almost every $t \in [0, T]$, with

$$\neg i := \begin{cases} L & \text{if } i = F, \\ F & \text{if } i = L. \end{cases}$$

Similarly, we introduce the concept of solution of system (3).

Definition 2.6 (Solution of system (3)). Let $(\overline{\nu}, \overline{\sigma}) \in \mathcal{M}_c(\mathbb{R}^d) \times \mathcal{P}(\{F, L\})$ be given, as well as $\nu : [0,T] \to \mathcal{M}_c(\mathbb{R}^d)$ and $\sigma : [0,T] \to \mathcal{P}(\{F, L\})$. We say that (ν_t, σ_t) is a solution of system (3) with initial datum $(\overline{\nu}, \overline{\sigma})$ when

1. $\nu_0 = \overline{\nu}$ and $\sigma_0 = \overline{\sigma}$;

2. the function $t \to \nu_t$ is continuous with respect to the topology of weak convergence of measures, while $t \to (\sigma_t(F), \sigma_t(L))$ is absolutely continuous[3] ;

3. there exists $R_T > 0$ such that $\cup_{t\in[0,T]} \mathrm{supp}(\nu_t) \subseteq B(0, R_T)$;

4. (ν_t, σ_t) satisfy

$$\dot{\sigma}_t(i) = A_{\nu_t,\sigma_t}\sigma_t(i)$$

[3] It is sufficient to prove absolute continuity of one component only, since $\sigma_t(F) + \sigma_t(L) = 1$.

for almost every $t \in [0,T]$, with A_{ν_t,σ_t} given in (5), as well as

$$\frac{d}{dt}\int_{\mathbb{R}^d} \varphi(x)d\nu_t(x) = \int_{\mathbb{R}^d} \nabla\varphi(x)\cdot \langle K, \nu_t \times \sigma_t\rangle (x)d\nu_t(x)$$

for every $\varphi \in \mathcal{C}_c^1(\mathbb{R}^d)$.

Remark 2.7. Throughout the paper it is assumed that the transition rates encoded by the matrix $A_{\nu,\sigma}$ only depend on the global state of the system and not on the position x. While being already useful at the level of deducing existence of solutions of (1), this restriction will be needed in order to show equivalence between solutions of (1) and (3), provided the initial datum is suitably chosen, satisfying Assumption (H1) below. This will be apparent in the proof of Proposition 3.4. As already discussed in the Introduction, this equivalence is a crucial step for the mean field derivation in Section 4.

2.3 The method of characteristics

In this section, we recall the method of characteristics to find solutions of transport equations. In particular, we recall the connection between the solutions of an ordinary differential equation with vector field v and the solution to transport equations as the evolution of the corresponding probability distribution.

We start with the classical definitions of Carathéodory functions and solutions.

Definition 2.8. A function $g : [0,T]\times\Omega \to \mathbb{R}^d$ is a Carathéodory function if

1. For all $t \in [0,T]$, the application $x \mapsto g(t,x)$ is Lipschitz.
2. For all $x \in \mathbb{R}^d$, the application $t \mapsto g(t,x)$ is measurable.
3. There exists $M > 0$ such that $|g(t,x)| \leqslant M(1+|x|)$ for all t, x.

A Carathéodory solution of

$$\dot{y}(t) = g(t,y(t)) \quad \text{for } t \in [0,T], \tag{8}$$

is an absolutely continuous function $y : [0,T] \to \mathbb{R}^d$ which satisfies (8) a.e. in $[0,T]$.

If the Lipschitz constant L_t of the function $g(t,\cdot)$ belongs to $L^1(0,T)$, existence and uniqueness of Carathéodory solutions to (8) can be shown, see e.g. [28]. From now on, we denote by Φ_t^g the flow of (8), i.e. the map $x_0 \mapsto \Phi_t^g(x_0)$ that associates to each initial data x_0 the corresponding solution of (8) at time t. Carathéodory solutions of finite dimensional systems and weak solutions of continuity equations are intimately related, as the following classical result shows.

Lemma 2.9. *Let $v : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ be a Carathéodory function and $X : [0,T] \to \mathbb{R}^d$ be a Carathéodory solution of*

$$\begin{cases} \dot{x} = v_t(x), \\ x(0) = x_0. \end{cases}$$

Then $\mu_t = \Phi^v_t \# \mu_0$ is the unique weak solution of

$$\begin{cases} \partial_t \mu_t &= -\mathrm{div}(v_t \mu_t), \\ \mu(0) &= \mu_0. \end{cases} \tag{9}$$

As a consequence, if $\mathrm{supp}(\mu_0) \subset B(0,R)$*, then for each $t > 0$ it holds*

$$\mathrm{supp}(\mu_t) \subset B(0, R + t\|v\|_{\mathcal{C}^0}). \tag{10}$$

Moreover, consider the inhomogeneous transport equation

$$\begin{cases} \partial_t \mu_t &= -\mathrm{div}(v_t \mu_t) + s_t, \\ \mu(0) &= \mu_0. \end{cases} \tag{11}$$

for s_t being a measurable family (with respect to the weak topology of meaures) of signed Borel measures such that there exist M_s, R_s with

$$|s_t^+| + |s_t^-| \leq M_s, \qquad \mathrm{supp}(s_t) \subset B(0, R_s)$$

for all $t \in [0,T]$.

Then, there exists a unique solution to (11)*, that satisfies the Duhamel's formula*

$$\mu_t = \Phi^v_t \# \mu_0 + \int_0^t \Phi^v_{(\tau,t)} \# s_\tau \, d\tau. \tag{12}$$

Here, $\Phi^v_{(\tau,t)}$ is the flow of the non-autonomous vector field v_t starting at time τ, i.e. the function $x_\tau \mapsto \Phi^v_{(\tau,t)}(x_\tau)$ that associates to x_τ the solution at time t of

$$\begin{cases} \dot{x} = v_t(x), \\ x(\tau) = x_\tau. \end{cases}$$

As a consequence, if $\mathrm{supp}(\mu_0) \subset B(0,R)$*, then for each $t > 0$ it holds*

$$\mathrm{supp}(\mu_t) \subset B(0, \max\{R, R_s\} + t\|v\|_{\mathcal{C}^0}). \tag{13}$$

Proof. For the existence of a solution to (9), which is the push-forward of the initial datum via the flow map, see e.g. [49]. Uniqueness comes from standard arguments for the linear continuity equation, see e.g. [7].

We now prove (10). For a given $t > 0$, consider a test function φ with compact support, such that $\varphi \equiv 0$ on $B(0, R + t\|v\|_{\mathcal{C}^0})$. It then holds

$$\int_{\mathbb{R}^d} \varphi(x) \, d(\Phi^v_t \# \eta_0)(x) = \int_{\mathbb{R}^d} \varphi(\Phi^v_t(x)) \, d\eta_0(x). \tag{14}$$

Recall the elementary estimate for ordinary differential equations

$$|\Phi^v_t(x) - x| = \left| \int_0^t v(s, x(s)) \, ds \right| \leq t\|v\|_{\mathcal{C}^0}.$$

Since for each $x \in \mathrm{supp}(\mu(0))$ it holds $|x| \leq R$, then $\varphi(\Phi_t^v(x)) = 0$. Thus, the integral in (14) is zero. Since this holds for any test function with support outside $B(0, R+t\|v\|_{\mathcal{C}^0})$, this implies that η_t is supported in $B(0, R+t\|v\|_{\mathcal{C}^0})$.

The proof of existence for the inhomogeneous case is similar. Duhamel's formula is a re-writing of the method of variations of constants, that can be verified with direct computations. Uniqueness can be proved with the standard method: the difference between two solutions solves (9) with $\mu_0 \equiv 0$, then its unique solution is $\mu_t \equiv 0$. The proof for (13) follows the proof for (10). □

2.4 The Generalized Wasserstein distance

The main technical issue about the transport equation (1) is that it mixes two different phenomena: on one side the non-local dynamics given by convolutions $K^i * \mu^i$; on the other side, sources and sink that make the total mass of μ^i non-constant.

It has been shown in several examples that the Wasserstein distance is a powerful tool to deal with transport equation with non-local vector fields, see e.g. [6, 24, 33, 42, 49]. Neverthelss, the Wasserstein distance is defined between measures with the same mass, hence it is not useful for problems in which the mass varies in time. This issue recently led to the development of a series of different generalizations of the Wasserstein distance to measures with different masses. See e.g. [20, 37, 39, 43].

In this article, we choose to use the generalized Wasserstein distance, that has been introduced in [43, 44]. Indeed, it has been already proved in [43] that, under suitable hypotheses written in terms of the generalized Wasserstein distance, transport equations with both non-local velocities and source terms admit existence and uniqueness of the solution.

We now recall the definition of the generalized Wasserstein distance, together with some key properties.

Definition 2.10. Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$ be two measures. Given $a, b > 0$ and $p \geq 1$, we define the functional

$$\mathcal{W}_g^{a,b,p}(\mu,\nu) := \inf_{\tilde{\mu},\tilde{\nu} \in \mathcal{M}(\mathbb{R}^d),\, |\tilde{\mu}|=|\tilde{\nu}|} \left(a^p \left(|\mu-\tilde{\mu}| + |\nu-\tilde{\nu}|\right)^p + b^p W_p^p(\tilde{\mu},\tilde{\nu})\right)^{1/p}. \tag{15}$$

Proposition 2.11. *The following properties hold:*

1. *The functional $\mathcal{W}_g^{a,b,p}$ is a distance on $\mathcal{M}(\mathbb{R}^d)$.*

2. *The distance $\mathcal{W}_g^{a,b,p}$ metrizes the weak convergence on compact sets, i.e. given μ_n, μ with $\mathrm{supp}(\mu_n), \mathrm{supp}(\mu) \subset B_R(0)$ it holds*

$$\mu_n \rightharpoonup \mu \quad \textit{if and only if} \quad \mathcal{W}_g^{a,b,p}(\mu_n,\mu) \to 0.$$

3. *The space $\mathcal{M}(\mathbb{R}^d)$ is complete with respect to $\mathcal{W}_g^{a,b,p}$.*

4. *Let v_t, w_t be two Lipschitz vector fields, with L a Lipschitz constant for both v_t, w_t. It then holds:*

$$\mathcal{W}_g^{a,b,p}(\mu, \Phi_t^v \# \mu) \le b \sup_{\tau\in[0,t]} \{\|v_\tau\|_{\mathcal{C}^0}\} t\,|\mu| \tag{16}$$

$$\mathcal{W}_g^{a,b,p}(\Phi_t^v \# \mu, \Phi_t^w \# \nu) \le e^{\frac{p+1}{p}Lt} \mathcal{W}_g^{a,b,p}(\mu,\nu) + |\mu| \frac{be^{Lt/p}(e^{Lt}-1)}{L} \sup_{\tau\in[0,t]} \{\|v_\tau - w_\tau\|_{\mathcal{C}^0}\}. \tag{17}$$

5. *It holds*

$$\mathcal{W}_g\,(\mu,\nu) \le |\mu| + |\nu|. \tag{18}$$

Proof. See [43, 44]. □

We now recall a result equivalent to the Kantorovich-Rubinstein duality for the generalized Wasserstein distance $\mathcal{W}_g^{1,1,1}$. It states that it coincides with the so-called flat distance, see e.g. [26].

Theorem 2.12. *Let $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$. Then*

$$\mathcal{W}_g^{1,1,1}(\mu,\nu) = \sup\left\{ \int f d(\mu-\nu) \mid \|f\|_\infty \le 1,\ \mathrm{Lip}(f) \le 1 \right\}.$$

As simple consequences, for $\lambda, \bar\lambda > 0$ it holds

$$\mathcal{W}_g^{1,1,1}(\lambda\mu, \bar\lambda\mu) = |\lambda - \bar\lambda|\,|\mu| \tag{19}$$

$$\mathcal{W}_g^{1,1,1}(\lambda\mu, \lambda\nu) = \lambda \mathcal{W}_g^{1,1,1}(\mu,\nu). \tag{20}$$

The proof is given in [44].

From now on, we will only deal with the generalized Wasserstein distance $\mathcal{W}_g^{1,1,1}$, i.e. with the flat distance. For this reason, we will drop the parameters, and use the notation

$$\mathcal{W}_g\,(\mu,\nu) := \mathcal{W}_g^{1,1,1}(\mu,\nu).$$

Moreover, we use the same notation for the corresponding distance on $\mathcal{M}_c(\mathbb{R}^d)\times\mathcal{M}_c(\mathbb{R}^d)$: given $\mu = (\mu^F, \mu^L)$ and $\nu = (\nu^F, \nu^L)$, we write

$$\mathcal{W}_g\,(\mu,\nu) := \mathcal{W}_g\left(\mu^F, \nu^F\right) + \mathcal{W}_g\left(\mu^L, \nu^L\right). \tag{21}$$

Finally, we use again the same notation for the supremum distance on $C([0,T], \mathcal{M}_c(\mathbb{R}^d)\times \mathcal{M}_c(\mathbb{R}^d))$: given $\mu : t \mapsto \mu_t = (\mu_t^F, \mu_t^L)$ and $\nu : t \mapsto \nu_t = (\nu_t^F, \nu_t^L)$, we write

$$\mathcal{W}_g\,(\mu,\nu) := \sup_{t\in[0,T]} \mathcal{W}_g((\mu_t^{F,k}, \mu_t^{L,k}), (\nu_t^{F,k}, \nu_t^{L,k})). \tag{22}$$

3 Well-posedness and equivalence for the leader-follower dynamics

We now turn our attention to system (1) and use the tools introduced in Section 2 to prove the existence and uniqueness of solutions. To do so, we will define a sequence of measures $(\mu^{F,k}, \mu^{L,k})$ as explicit Euler approximations of the dynamics (1) and, by a a compactness argument in the space $\mathcal{M}_c(\mathbb{R}^d)$ embedded with the generalized Wasserstein distance $\mathcal{W}_g$, we show that it converges, up to subsequences, to the unique solution (μ^F, μ^L) of system (1). Next, we shall establish a bijection between solutions of (1) and of (3) under certain assumptions on the initial data. As a byproduct of the previous results, such equivalence yields the well-posedness of (3) as well, paving the way for the mean-field analysis of the subsequent sections.

3.1 Main assumptions

In this section we discuss the set of assumptions we shall assume henceforth. These assumptions assure, in particular, the existence and uniqueness of solutions of (1), as well as the equivalence between (1) and (3), that is more amenable to a mean-field analysis, as we will show in Section 4. We warn in advance the reader that Assumption (H1) below, differently from the other ones, is not needed for the existence result in Proposition 3.2, but will be used for the equivalence result in Proposition 3.4.

(H1) There exist $\overline{\sigma} \in \mathcal{P}(\{F, L\})$ and $\overline{\nu} \in \mathcal{M}_c(\mathbb{R}^d)$ such that $\overline{\mu}^F = \overline{\sigma}(F)\overline{\nu}$ and $\overline{\mu}^L = \overline{\sigma}(L)\overline{\nu}$.

(H2) there exists a constant $L_K > 0$ such that, for every $x_1, x_2 \in \mathbb{R}^d$ and $i \in \{F, L\}$, it holds

$$|K^i(x_1) - K^i(x_2)| \leq L_K |x_1 - x_2|.$$

(H3) there exists a constant $B_K > 0$ such that, for every $x \in \mathbb{R}^d$ and $i \in \{F, L\}$, it holds

$$|K^i(x)| \leq B_K(1 + |x|).$$

(H4) there exists a constant M_α such that for every $i \in \{F, L\}$ and $(\mu^F, \mu^L) \in \mathcal{M}_c(\mathbb{R}^d) \times \mathcal{M}_c(\mathbb{R}^d)$ it holds

$$0 \leq \alpha_i(\mu^F, \mu^L) \leq M_\alpha.$$

(H5) there exists a constant $L_{\alpha,M,R}$ such that, for every $i \in \{F, L\}$ and $(\mu^F, \mu^L), (\nu^F, \nu^L) \in \mathcal{M}_c(\mathbb{R}^d) \times \mathcal{M}_c(\mathbb{R}^d)$ satisfying

$$|\mu^F| + |\mu^L| = |\nu^F| + |\nu^L| \leq M, \tag{23}$$

and

$$\operatorname{supp}(\mu^j), \operatorname{supp}(\nu^j) \subset B(0, R), \qquad j \in \{F, L\} \tag{24}$$

it holds

$$|\alpha_i(\mu^F, \mu^L) - \alpha_i(\nu^F, \nu^L)| \leq L_{\alpha,M,R}(\mathcal{W}_g\left(\mu^F, \nu^F\right) + \mathcal{W}_g\left(\mu^L, \nu^L\right)). \tag{25}$$

We now list some useful consequences of the previous hypotheses.

Proposition 3.1. *Let (H4)-(H5) hold, and μ, ν satisfy* (23)-(24)*. Then, it exists $L'_{\alpha,M,R}$ such that for each $i \in \{F, L\}$ it holds*

$$\mathcal{W}_g\left(\alpha_i(\mu^F, \mu^L)\mu^i, \alpha_i(\nu^F, \nu^L)\nu^i\right) \leq L'_{\alpha,M,R}\mathcal{W}_g\left(\mu, \nu\right) \tag{26}$$

Proof. Use the triangular inequality and (19)-(20) to write

$$\begin{aligned}
&\mathcal{W}_g\left(\alpha_i(\mu^F, \mu^L)\mu^i, \alpha_i(\nu^F, \nu^L)\nu^i\right) \leq \\
&\mathcal{W}_g\left(\alpha_i(\mu^F, \mu^L)\mu^i, \alpha_i(\nu^F, \nu^L)\mu^i\right) + \mathcal{W}_g\left(\alpha_i(\nu^F, \nu^L)\mu^i, \alpha_i(\nu^F, \nu^L)\nu^i\right) = \\
&|\alpha_i(\mu^F, \mu^L) - \alpha_i(\nu^F, \nu^L)|\,|\mu^i| + \alpha_i(\nu^F, \nu^L)\mathcal{W}_g\left(\mu^i, \nu^i\right) \leq \\
&L_{\alpha,M,R}(\mathcal{W}_g\left(\mu^F, \nu^F\right) + \mathcal{W}_g\left(\mu^L, \nu^L\right))M + M_\alpha \mathcal{W}_g\left(\mu^i, \nu^i\right),
\end{aligned}$$

from which the result easily follows. □

For $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\sigma \in \mathcal{P}_1(\{F, L\})$, we will use (as already done in (5)) the notations $\alpha_F(\nu, \sigma)$ and $\alpha_L(\nu, \sigma)$ to indicate the transition rates defined by

$$\alpha_F(\nu, \sigma) := \alpha_F(\sigma(F)\nu, \sigma(L)\nu), \quad \alpha_L(\nu, \sigma) := \alpha_L(\sigma(F)\nu, \sigma(L)\nu)\,. \tag{27}$$

If (25) holds, it easily follows from the definition (15) that, for ν_1, $\nu_2 \in \mathcal{P}_1(\mathbb{R}^d)$ satisfying (24) and σ_1, $\sigma_2 \in \mathcal{P}_1(\{F, L\})$, we have

$$|\alpha_i(\nu_1, \sigma_1) - \alpha_i(\nu_2, \sigma_2)| \leq L_{\alpha,R}(\mathcal{W}_1(\nu_1, \nu_2) + |\sigma_1(F) - \sigma_2(F)|)$$

for $i = F, L$. We additionally exploited above the inequality $\mathcal{W}_g\left(\nu_1, \nu_2\right) \leq \mathcal{W}_1(\nu_1, \nu_2)$ which immediately stems out of (15) whenever ν_1 and ν_2 are probability measures. If we endow the set $\{F, L\}$ with the usual distance on finite sets defined by

$$|y - \overline{y}|_{\{F,L\}} := \begin{cases} 0 & \text{if } y = \overline{y}, \\ 1 & \text{otherwise.} \end{cases} \tag{28}$$

we can rewrite the above inequality as

$$|\alpha_i(\nu_1, \sigma_1) - \alpha_i(\nu_2, \sigma_2)| \leq L_{\alpha,R}(\mathcal{W}_1(\nu_1, \nu_2) + \mathcal{W}_1(\sigma_1, \sigma_2))\,. \tag{29}$$

3.2 Existence and uniqueness

In this section, we prove existence and uniqueness of the solutions to Cauchy problems with dynamics given by systems (1) and (3). For the first case, we will adapt ideas from [43], while for the second we will use the equivalence of the two problems.

We first prove an existence result for (1).

Proposition 3.2. *Let an initial data* $(\mu_0^F, \mu_0^L) \in \mathcal{M}_c(\mathbb{R}^d) \times \mathcal{M}_c(\mathbb{R}^d)$ *and a time interval* $[0, T]$ *be fixed. For each* $k \in \mathbb{N}$*, define an* **explicit Euler approximation** $\mu^{F,k}, \mu^{L,k}$ *of the solution to system* (1) *as follows: fix* $\Delta t = T/2^k$ *and define*

$$(\mu_0^{F,k}, \mu_0^{L,k}) := (\mu_0^F, \mu_0^L); \tag{30}$$

$$v_{n\Delta t}^k := K^F * \mu_{n\Delta t}^{F,k} + K^L * \mu_{n\Delta t}^{L,k}, \qquad n = 0, \dots, 2^k, \tag{31}$$

$$\mu_{(n+1)\Delta t}^{F,k} := \Phi_{\Delta t}^{v_{n\Delta t}^k} \# \left(\mu_{n\Delta t}^{F,k} + \Delta t(-\alpha_F(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{F,k} + \alpha_L(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{L,k}) \right); \tag{32}$$

$$\mu_{(n+1)\Delta t}^{L,k} := \Phi_{\Delta t}^{v_{n\Delta t}^k} \# \left(\mu_{n\Delta t}^{L,k} + \Delta t(\alpha_F(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{F,k} - \alpha_L(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{L,k}) \right). \tag{33}$$

Also define the solution on intermediate times: for $\tau \in (0, 1)$ *define*

$$\mu_{(n+\tau)\Delta t}^{F,k} := \Phi_{\tau\Delta t}^{v_{n\Delta t}^k} \# \left(\mu_{n\Delta t}^{F,k} + \tau\Delta t(-\alpha_F(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{F,k} + \alpha_L(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{L,k}) \right); \tag{34}$$

$$\mu_{(n+\tau)\Delta t}^{L,k} := \Phi_{\tau\Delta t}^{v_{n\Delta t}^k} \# \left(\mu_{n\Delta t}^{L,k} + \tau\Delta t(\alpha_F(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{F,k} - \alpha_L(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k})\mu_{n\Delta t}^{L,k}) \right). \tag{35}$$

Let Hypotheses (H2)-(H3)-(H4)-(H5) hold. Let moreover be $\Delta t M_\alpha < 1$*. Then, the following properties hold:*

1. *both* $\mu_t^{F,k}$ *and* $\mu_t^{F,k}$ *are non-negative measures;*
2. *the total mass is preserved, since it satisfies*
$$|\mu_t^{F,k}| + |\mu_t^{L,k}| = |\mu_0^F| + |\mu_0^L|; \tag{36}$$
3. *the sequence has equi-bounded support, i.e there exists* $R > 0$ *such that for all* $t \in [0, T]$ *and* $k \in \mathbb{N}$ *it holds*
$$\operatorname{supp}(\mu_t^{F,k}), \operatorname{supp}(\mu_t^{L,k}) \subset B(0, R);$$
4. *the sequence* $\{(\mu_t^{F,k}, \mu_t^{L,k})\}_{k\in\mathbb{N}}$ *is uniformly bounded and uniformly Lipschitz in the* t *variable with respect to the distance* (21)*.*

As a consequence, there exists a subsequence of $(\mu^{L,k}, \mu^{F,k})$ *converging with respect to the uniform convergence, i.e. with respect to the metric* (22)*. The limit of such subsequence is a solution to* (1)*.*

Proof. We prove **Property 1**. We first prove that $\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k}$ are non-negative measures for each $n = 0, \dots, 2^k$, by induction on n. It is clear that the property holds for $n = 0$, since (30) holds.

Let now be $\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k}$ non-negative measures. We aim to prove that $\mu_{(n+1)\Delta t}^{F,k}, \mu_{(n+1)\Delta t}^{L,k}$ given by (32)-(33) are non-negative measures. We only prove it for $\mu_{(n+1)\Delta t}^{F,k}$, since the proof for $\mu_{(n+1)\Delta t}^{L,k}$ is similar. Observe that $\Delta t M_\alpha < 1$, together with (H4), implies

$$1 - \Delta t \alpha_F(\mu_{n\Delta t}^{F,k}, \mu_{n\Delta t}^{L,k}) > 0.$$

Then $\mu^{F,k}_{n\Delta t}(1-\Delta t\alpha_F(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t}))$ is a non-negative measure, as well as $\Delta t\alpha_L(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})\mu^{L,k}_{n\Delta t}$. Their sum is thus a non-negative measure, and its push-forward by $\Phi^{v^k_{n\Delta t}}_{\Delta t}$ is non-negative too. By induction, this proves that $\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t}$ are non-negative measures for each $n=0,\ldots,2^k$.

For intermediate times of the form $(n+\tau)\Delta t$, first observe that we just proved that $\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t}$ are non-negative measures. Moreover, $\tau\in(0,1)$ implies

$$1-\tau\Delta t\alpha_F(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})>0.$$

Then, following the proof of the previous case, we have that $\mu^{F,k}_{(n+\tau)\Delta t},\mu^{L,k}_{(n+\tau)\Delta t}$ are non-negative measures.

We now prove **Property 2**. We first prove that (36) holds for times of the form $n\Delta t$, again by induction on n. Definition (30) implies that (36) holds for $n=0$. If (36) holds for a given n, then it holds for $n+1$, as a consequence of (32)-(33). Indeed, by the proof of Proposition 1, we know that both $\mu^{F,k}_{n\Delta t}(1-\Delta t\alpha_F(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t}))$ and $\Delta t\alpha_L(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})\mu^{L,k}_{n\Delta t}$ are non-negative measures, and the same holds for the corresponding terms in (33). Thus, the mass of the sum is the sum of the masses, and the push-forward of a non-negative measure preserves the mass. As a consequence, it holds

$$\begin{aligned}|\mu^{F,k}_{(n+1)\Delta t}| \quad + \quad & |\mu^{L,k}_{(n+1)\Delta t}| = (1-\Delta t\alpha_F(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t}))|\mu^{F,k}_{n\Delta t}| + \Delta t\alpha_L(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})|\mu^{L,k}_{n\Delta t}| + \\ & (1-\Delta t\alpha_L(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})\mu^{L,k}_{n\Delta t})|\mu^{L,k}_{n\Delta t}| + \Delta t\alpha_F(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})|\mu^{F,k}_{n\Delta t}| = \\ & |\mu^{F,k}_{n\Delta t}| + |\mu^{L,k}_{n\Delta t}| = |\mu^F_0| + |\mu^F_0|,\end{aligned}$$

where we used homogeneity of the mass $|\lambda\mu|=\lambda|\mu|$. The proof for intermediate times is identical.

We now prove **Property 3**. First observe that, due to (H3), the hypothesis

$$\mathrm{supp}(\mu^F),\mathrm{supp}(\mu^L)\subset B(0,R)$$

implies

$$\|K^F*\mu^F+K^L*\mu^L\|_{\mathcal{C}^0}\le\|K^F\|_{\mathcal{C}^0}|\mu^F|+\|K^L\|_{\mathcal{C}^0}|\mu^L|\le B_K(1+2R)(|\mu^F|+|\mu^L|). \quad (37)$$

Choose now $R_0>0$ such that $\mathrm{supp}(\mu^F_0),\mathrm{supp}(\mu^L_0)\subset B(0,R_0)$. We now define a sequence R^k_n such that

$$\mathrm{supp}(\mu^{F,k}_{n\Delta t}),\mathrm{supp}(\mu^{L,k}_{n\Delta t})\subset B(0,R^k_n), \quad (38)$$

by induction. It first holds $R^k_0=R_0$ by (30). By definition of $v^k_{n\Delta t}$ in (31), and also using (37) and Property 1, it holds

$$\|v^k_{n\Delta t}\|_{\mathcal{C}^0}\le B_K(1+2R^k_n)(|\mu^{F,k}_{n\Delta t}|+|\mu^{L,k}_{n\Delta t}|)=B_K(1+2R^k_n)(|\mu^F_0|+|\mu^L_0|). \quad (39)$$

Apply (10) to (32)-(33): since $\mathrm{supp}(\mu^{F,k}_{n\Delta t}),\mathrm{supp}(\mu^{L,k}_{n\Delta t})\subset B(0,R^k_n)$, then

$$\mathrm{supp}(\mu^{F,k}_{(n+1)\Delta t}),\mathrm{supp}(\mu^{L,k}_{(n+1)\Delta t})\subset B(0,R^k_n+\Delta tB_K(1+2R^k_n)(|\mu^F_0|+|\mu^L_0|)). \quad (40)$$

Define now the sequence

$$R_0^k := R_0, \qquad R_{n+1}^k = (1 + \Delta t\, C)R_n^k + \Delta t\, C$$

with $C := 2B_K(|\mu_0^F| + |\mu_0^L|))$. With this choice, (38) holds. Moreover, again by applying (10) to the definition of μ_t^k at intermediate times (34)-(35), for each $\tau \in (0,1)$ it holds

$$\operatorname{supp}(\mu_{(n+\tau)\Delta t}^{F,k}), \operatorname{supp}(\mu_{(n+\tau)\Delta t}^{L,k}) \subset B(0, R_{n+1}^k). \tag{41}$$

We now recall that n runs from 0 to 2^k. Since R_n^k is an increasing sequence with respect to the parameter n, then (40)-(41) imply that for each $k \in \mathbb{N}$ and each $t \in [0,T]$ it holds

$$\operatorname{supp}(\mu_t^{F,k}), \operatorname{supp}(\mu_t^{L,k}) \subset B(0, R_{2^k}^k).$$

An explicit computation shows that

$$R_{2^k}^k = (1 + \Delta t\, C)^{2^k}(R_0 + 1) - 1 \leq e^{2^k \Delta t\, C}(R_0 + 1) < e^{TC}(R_0 + 1), \tag{42}$$

thus, supports of μ_t^k are uniformly bounded.

We now prove **Property 4**. Since we proved that $|\mu_t^{F,k}| + |\mu_t^{L,k}| = |\mu_0^F| + |\mu_0^L|$, it holds both $|\mu_t^{F,k}| \leq |\mu_0^F| + |\mu_0^L|$ and $|\mu_t^{F,k}| \leq |\mu_0^F| + |\mu_0^L|$. Then, by applying (18), it holds

$$\mathcal{W}_g\left(\mu_t^{F,k}, \mu_t^{L,k}\right) \leq |\mu_t^{F,k}| + |\mu_t^{L,k}| \leq 2(|\mu_0^F| + |\mu_0^L|),$$

then the sequence is equi-bounded.

We now prove equi-Lipschitz continuity. Let $k \in \mathbb{N}$ be fixed, and assume to have t, s such that $n\Delta t \leq t < s \leq (n+1)\Delta t$. We then want to estimate $\mathcal{W}_g(\mu_t^k, \mu_s^k)$. Observe that, by (34)-(35) and the property of composition of flows, it holds $\mu_s^{F,k} = \Phi_{s-t}^{v_{n\Delta t}^k} \# \mu_t^{F,k}$, and similarly for $\mu^{L,k}$. Apply now (16) to $v_{n\Delta t}^k$, that satisfies (39) and recall that $R_n^k \leq e^{TC}(R_0 + 1)$, as proved for Property 3. This implies

$$\mathcal{W}_g(\mu_t^{F,k}, \mu_s^{F,k}) \leq B_K(1 + 2e^{TC}(R_0 + 1))(|\mu_0^F| + |\mu_0^L|)^2 |t - s|. \tag{43}$$

The same estimate holds for $\mathcal{W}_g(\mu_t^{L,k}, \mu_s^{L,k})$, then for $\mathcal{W}_g(\mu_t^k, \mu_s^k)$ by doubling the right hand side. For general $t < s \in [0,T]$, one recovers (43) by applying the triangular inequality on each sub-interval $[t, n\Delta t]$, $[n\Delta t, (n+1)\Delta t]$, …, $[(n+k)\Delta t, s]$.

We finally prove the **existence of a solution to** (1). First observe that Property 4, together with the Arzelà-Ascoli theorem, implies the existence of a subsequence (that we do not relabel) μ^k that uniformly converges to some μ^* with respect to the metric $\mathcal{W}_g$.

We are left to prove that μ^* is a solution to (1), in the sense of Definition 2.5. Since $\mu_0^{F,k} = \bar{\mu}^F$, by uniform convergence it holds $\mu_0^{F,*} = \bar{\mu}^F$, and the same holds for $\mu_0^{L,*}$. Then, Condition 1 of Definition 2.5 is proved.

Condition 3 of uniform boundedness of the support comes from Property 3. Indeed, μ^k has uniformly bounded support in some $B(0,R)$ implies that μ^* has uniformly bounded support too, in $B(0,R+1)$. To prove this classical result, it is sufficient to test μ^* with functions having support outside $B(0,R+1)$.

We now prove Condition 2 of continuity with respect to the weak convergence of measures. It is a consequence of the fact that the sequence μ^k is equi-Lipschitz, thus μ^* is Lipschitz with respect to the distance $\mathcal{W}_g$, and such distance metrizes weak convergence on measures with equi-bounded support (Proposition 2.11, statement 2).

We now prove Condition 4. We first prove a list of auxiliary estimates. Take a function φ with extra regularity, namely $\varphi \in \mathcal{C}^2_c(\mathbb{R}^d)$, and fix $t \in [0,T]$. For each k in the subsequence $\mu^k \to \mu^*$, choose n as the largest integer satisfying $n\Delta t \leq t$. Thus, $t - n\Delta t \geq 0$. We have the following estimates:

Estimate 1. Take $m_1 := \|\varphi\|_{\mathcal{C}^1} = \|\varphi\|_{\mathcal{C}^0} + \mathrm{Lip}(\varphi)$. It then holds

$$\left| \int_{\mathbb{R}^d} \varphi \, d(\mu_t^{F,*} - \mu_t^{F,k}) \right| \leq m_1 \mathcal{W}_g(\mu_t^{F,*}, \mu_t^{F,k}).$$

This is a consequence of the Kantorovich-Rubinstein duality for the generalized Wasserstein distance, see Theorem 2.12.

Estimate 2. Define

$$v_t^* := K^F * \mu_t^{F,*} + K^L * \mu_t^{K,*}. \tag{44}$$

It exists m_2, independent on t, k, n, such that it holds

$$\left| \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot v^k_{n\Delta t} d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t} \# \mu^{F,k}_{n\Delta t} - \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot v_t^* d\mu_t^{*,k} \right| \leq$$
$$m_2(\mathcal{W}_g(\mu_t^{F,*}, \mu_t^{F,k}) + (t - n\Delta t)). \tag{45}$$

Indeed, we first observe that (39)-(42) imply

$$\|v^k_{n\Delta t}\|_{\mathcal{C}^0} \leq B_K(1 + (2e^{TC}(R_0+1)))(|\mu_0^F| + |\mu_0^L|). \tag{46}$$

Second, recall that v^k_n in (31) is defined as a convolution. The, Lipschitz continuity of K^F, K^L given by (H2), implies equi-Lipschitz continuity of the v^k_n. Indeed, it holds

$$\begin{aligned} |(K^F * \mu)(x) - (K^F * \mu)(y)| \;\; &\leq \;\; \int_{\mathbb{R}^d} |K^F(z-x) - K^F(z-y)| \, d\mu(z) \leq \\ & \qquad L_k|x-y| \, |\mu|. \end{aligned} \tag{47}$$

Thus, equi-boundedness of masses (Property 2) implies equi-Lipschitz continuity.

Third, since $\varphi \in \mathcal{C}^2(\mathbb{R}^d)$, the family $\nabla\varphi \cdot v^k_{n\Delta t}$ is equi-bounded and equi-Lipschitz, i.e. $m_2' := \sup_{n,k} \left\{ \|\nabla\varphi \cdot v^k_{n\Delta t}\|_{\mathcal{C}^0}, \mathrm{Lip}(\nabla\varphi \cdot v^k_{n\Delta t}) \right\}$ is finite. Thus, Kantorovich-Rubinstein duality implies

$$\left| \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot v^k_{n\Delta t} d(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t} \# \mu^{F,k}_{n\Delta t} - \mu_t^{F,*}) \right| \leq m_2' \mathcal{W}_g(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t} \# \mu^{F,k}_{n\Delta t}, \mu_t^{F,*}). \tag{48}$$

We apply the triangular inequality to have

$$\mathcal{W}_g(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t}, \mu^{F,*}_t) \leq \mathcal{W}_g(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t}, \mu^{F,k}_t) + \mathcal{W}_g(\mu^{F,k}_t, \mu^{F,*}_t). \tag{49}$$

For the first term, recall the definition of (34) and apply the Kantorovich-Rubinstein duality, observing that it holds

$$\begin{aligned}
&\int_{\mathbb{R}^d} f\, d(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t} - \mu^{F,k}_t) = \\
&\int_{\mathbb{R}^d} (-f)\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# \left((t-n\Delta t)(-\alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t} + \alpha_L(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{L,k}_{n\Delta t})\right) \\
&\leq \|f\|_{\mathcal{C}^0}(t-n\Delta t)\left|-\alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t} + \alpha_L(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{L,k}_{n\Delta t})\right| \leq \\
&1(t-n\Delta t)M_\alpha(|\mu^{F,k}_{n\Delta t}| + |\mu^{L,k}_{n\Delta t}|) = (t-n\Delta t)M_\alpha(|\mu^F_0| + |\mu^L_0|).
\end{aligned}$$

We use the fact that push-forward conserves the mass, hypothesis (H4) and Property 2. Since such estimate is independent on f satisfying $\|f\|_{\mathcal{C}^0} \leq 1$, this gives

$$\mathcal{W}_g(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t}, \mu^{F,k}_t) \leq (t-n\Delta t)M_\alpha(|\mu^F_0| + |\mu^L_0|).$$

Merging it with (48)-(49), we have (45).

Estimate 3. Define

$$s^k_{n\Delta t} := -\alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t} + \alpha_L(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{L,k}_{n\Delta t},$$

and similarly

$$s^*_t := -\alpha_F(\mu^{F,*}_t, \mu^{L,*}_t)\mu^{F,*}_t + \alpha_L(\mu^{F,*}_t, \mu^{L,*}_t)\mu^{L,*}_t.$$

It exists m_3, independent on t, k, n, such that it holds

$$\left|\int_{\mathbb{R}^d} \varphi\, d\left(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#s^k_{n\Delta t} - s^*_t\right)\right| \leq m_3(\mathcal{W}_g(\mu^*_t, \mu^k_t) + (t-n\Delta t)). \tag{50}$$

Indeed, we first consider the negative parts of the measures $\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#s^k_{n\Delta t}$ and s^*_t. They satisfy

$$\begin{aligned}
&C_- := \left|\int_{\mathbb{R}^d} \varphi(x)\, d(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#(\alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t}) - \alpha_F(\mu^{F,*}_t, \mu^{L,*}_t)\mu^{F,*}_t)\right| \leq \\
&m_1\mathcal{W}_g(\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#(\alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t}), \alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t}) + \\
&m_1\mathcal{W}_g(\alpha_F(\mu^{F,k}_{n\Delta t}, \mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t}, \alpha_F(\mu^{F,*}_t, \mu^{L,*}_t)\mu^{F,k}_{n\Delta t}) + \\
&m_1\mathcal{W}_g(\alpha_F(\mu^{F,*}_t, \mu^{L,*}_t)\mu^{F,k}_{n\Delta t}, \alpha_F(\mu^{F,*}_t, \mu^{L,*}_t)\mu^{F,*}_t).
\end{aligned}$$

where we used the definition of m_1 in Estimate 1, the Kantorovich-Rubinstein duality and the triangular inequality. For the first term, use (16) together with the estimate (46) for $\|v^k_{n\Delta t}\|_{\mathcal{C}^0}$, as well as (H4). For the second and third terms, use (H5): since (23)-(24) hold, then (25) holds with some $L_{\alpha,M,R}$. Moreover, use (19) for the second term and (20) for the third one. It then holds

$$\begin{aligned}
&C_- \leq m_1(t-n\Delta t)\|v^k_{n\Delta t}\|_{\mathcal{C}^0}|\alpha_F(\mu^{F,k}_{n\Delta t},\mu^{L,k}_{n\Delta t})\mu^{F,k}_{n\Delta t}| + \\
&m_1 L_{\alpha,M,R}\mathcal{W}_g\left(\mu^k_{n\Delta t},\mu^*_t\right) + m_1\alpha_F(\mu^{F,*}_t,\mu^{L,*}_t)\mathcal{W}_g(\mu^{F,k}_{n\Delta t},\mu^{F,*}_t) \leq \\
&m_3'(t-n\Delta t)M_\alpha|\mu^{F,k}_{n\Delta t}| + m_1 L_{\alpha,M,R}\mathcal{W}_g(\mu^k_{n\Delta t},\mu^*_t) + m_1 M_\alpha \mathcal{W}_g(\mu^{F,k}_{n\Delta t},\mu^{F,*}_t),
\end{aligned}$$

for some m_3' independent on t,k,n. Recall that masses are equi-bounded (Property 2). Also apply the triangular inequality and uniform Lipschitz continuity of the μ^k (Property 4) to write

$$\mathcal{W}_g(\mu^{F,k}_{n\Delta t},\mu^{F,*}_t) \leq \mathcal{W}_g(\mu^{F,k}_{n\Delta t},\mu^{F,k}_t) + \mathcal{W}_g(\mu^{F,k}_t,\mu^{F,*}_t) \leq L'|t-n\Delta t| + \mathcal{W}_g(\mu^{F,k}_t,\mu^{F,*}_t),$$

for some L', and similarly for $\mathcal{W}_g(\mu^{L,k}_{n\Delta t},\mu^{L,*}_t)$. It then exists m_3'' such that

$$C_- \leq m_3''((t-n\Delta t) + \mathcal{W}_g(\mu^{F,k}_t,\mu^{F,*}_t) + \mathcal{W}_g(\mu^{L,k}_t,\mu^{L,*}_t)).$$

An equivalent estimate holds for the positive parts of the measures $\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# s^k_{n\Delta t}$ and s^*_t. We then recover (50).

Estimate 4. There exists m_4 independent on t,k,n such that it holds

$$\left|\int_{\mathbb{R}^d}\nabla\varphi(x)\cdot v^k_{n\Delta t}\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# s^k_{n\Delta t}\right| \leq m_4 \tag{51}$$

Indeed, first recall that $\|v^k_{n\Delta t}\|_{\mathcal{C}^0}$ is uniformly bounded on the support of $\mu^k_{n\Delta t}$. Moreover,

$$\left|\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# s^k_{n\Delta t}\right| = |s^k_{n\Delta t}| = |(s^k_{n\Delta t})^+| + |(s^k_{n\Delta t})^-|$$

is uniformly bounded, as a consequence of (H4) and of uniform boundedness of masses (Property 2). This proves (51).

Estimate 5. We now prove that μ^k_t solves an approximated version of (1). By the definition (34) of $\mu^{F,k}_t$, and applying elementary properties of derivation as well as Lemma 2.9, it holds

$$\begin{aligned}
&\frac{d}{dt}\int_{\mathbb{R}^d}\varphi\, d\mu^{F,k}_t = \frac{d}{dt}\int_{\mathbb{R}^d}\varphi\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t} + \frac{d}{dt}\int_{\mathbb{R}^d}\varphi\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#((t-n\Delta t)s^k_n) = \\
&\int_{\mathbb{R}^d}\nabla\varphi\cdot v^k_{n\Delta t} d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t} + \frac{d}{dt}\left[(t-n\Delta t)\int_{\mathbb{R}^d}\varphi\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# s^k_n\right] = \\
&\int_{\mathbb{R}^d}\nabla\varphi\cdot v^k_{n\Delta t} d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\#\mu^{F,k}_{n\Delta t} + \int_{\mathbb{R}^d}\varphi\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# s^k_n + \\
&(t-n\Delta t)\int_{\mathbb{R}^d}\nabla\varphi\cdot v^k_{n\Delta t}\, d\Phi^{v^k_{n\Delta t}}_{t-n\Delta t}\# s^k_n.
\end{aligned} \tag{52}$$

for all $t \neq n\Delta t$. The equivalent estimate for $\mu_t^{L,k}$ holds too, by replacing $\mu_t^{F,k}$ with $\mu_t^{L,k}$ and s_n^k with $-s_n^k$.

One can write (52) in integral form too, as follows: for each $t \in [0,T]$ and $k \in \mathbb{N}$, choose the largest n [4] satisfying $n\Delta t = n2^{-k}T \leq t$. For each $\bar{t} \in [0,T]$, it holds

$$\int_0^{\bar{t}} dt \int_{\mathbb{R}^d} \varphi \, d\mu_t^{F,k} - \left(\int_{\mathbb{R}^d} \nabla\varphi \cdot v_{n\Delta t}^k d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# \mu_{n\Delta t}^{F,k} + \right.$$
$$\left. \int_{\mathbb{R}^d} \varphi \, d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# s_n^k + (t - n\Delta t) \int_{\mathbb{R}^d} \nabla\varphi \cdot v_{n\Delta t}^k \, d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# s_n^k . \right) = 0. \qquad (53)$$

We are now ready to prove Condition 4, that we prove in the equivalent integral form: for every $\bar{t} \in [0,T]$, the measure $\mu^{F,*}$ satisfies

$$\int_0^{\bar{t}} dt \left(\int_{\mathbb{R}^d} \varphi \, d\mu_t^{F,*} - \int_{\mathbb{R}^d} \nabla\varphi \cdot v_t^* \, d\mu_t^{F,*} - \int_{\mathbb{R}^d} \varphi \, ds_t^* \right) = 0, \qquad (54)$$

and a similar expression holds for $\mu^{L,*}$.

Assume that $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d)$. We then prove that (54) holds by writing

$$C^* := \left| \int_0^{\bar{t}} dt \left(\int_{\mathbb{R}^d} \varphi \, d\mu_t^{F,*} - \int_{\mathbb{R}^d} \nabla\varphi \cdot v_t^* \, d\mu_t^{F,*} - \int_{\mathbb{R}^d} \varphi \, ds_t^* \right) \right| \leq$$
$$\int_0^{\bar{t}} dt \left(\left| \int_{\mathbb{R}^d} \varphi \, d(\mu_t^{F,*} - \mu_t^{F,k}) \right| + \left| \int_{\mathbb{R}^d} \nabla\varphi \cdot v_t^* \, d\mu_t^{F,*} - \int_{\mathbb{R}^d} \nabla\varphi \cdot v_{n\Delta t}^k d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# \mu_{n\Delta t}^{F,k} \right| \right.$$
$$+ \left| \int_{\mathbb{R}^d} \varphi \, d(s_t^* - \Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# s_n^k) \right| + \left| \int_{\mathbb{R}^d} \varphi \, d\mu_t^{F,k} - \int_{\mathbb{R}^d} \nabla\varphi \cdot v_{n\Delta t}^k d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# \mu_{n\Delta t}^{F,k} - \right.$$
$$\left. \int_{\mathbb{R}^d} \varphi \, d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# s_n^k \right| \leq \int_0^{\bar{t}} dt \, \left((m_1 + m_2)\mathcal{W}_g(\mu_t^{F,*}, \mu_t^{F,k}) + m_3 \mathcal{W}_g(\mu_t^*, \mu_t^k) + \right.$$
$$\left. (m_2 + m_3)(t - n\Delta t) + \left| (t - n\Delta t) \int_{\mathbb{R}^d} \nabla\varphi \cdot v_{n\Delta t}^k \, d\Phi_{t-n\Delta t}^{v_{n\Delta t}^k} \# s_n^k \right| \right).$$

We used here Estimates 1, 2, 3, as well as Estimate 5 in its integral form (53). Recall now the definition of $\mathcal{W}_g(\mu^*, \mu^k)$ in (22) and use Estimate 4 for the last term. Also observe that it holds $t - n\Delta t \leq \Delta t = T2^{-k}$ by the choice of n. By defining $m := m_1 + m_2 + m_3 + m_4$, it holds

$$C^* \leq \bar{t}(m\mathcal{W}_g(\mu^*, \mu^k) + mT2^{-k}).$$

Since such estimate holds for any k in the converging subsequence, it holds $C^* = 0$.

We have then proved that (54) is satisfied for any $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d)$. Since for any μ^F, s^* the three operators $\varphi \to \int_{\mathbb{R}^d} \varphi \, d\mu^F$, $\int_{\mathbb{R}^d} \nabla\varphi \cdot v^* \, d\mu^F$, $\int_{\mathbb{R}^d} \varphi \, ds^*$ are continuous with respect to the norm $\mathcal{C}^1$, and $\mathcal{C}_c^2(\mathbb{R}^d)$ is dense in $\mathcal{C}_c^1(\mathbb{R}^d)$ with respect to such norm, then (54) is satisfied for any $\varphi \in \mathcal{C}_c^1(\mathbb{R}^d)$. □

[4] the dependence of n on t and k is omitted for the sake of notation.

We now prove existence and uniqueness of the solution to (1).

Proposition 3.3. *Let an initial data $(\mu_0^F,\mu_0^L)\in\mathcal{M}_c(\mathbb{R}^d)\times\mathcal{M}_c(\mathbb{R}^d)$ and a time interval $[0,T]$ be fixed. Let (H2)-(H3)-(H4)-(H5) hold. Then, there exists a unique solution to (1).*

Proof. Existence of a solution was proved in Propostion 3.2. We now prove uniqueness.

Let μ,ν be two solutions of (1), in the sense of Definition 2.5. They are both continuous with respect to the topology of weak convergence of measures (Condition 2) and have equi-bounded support (Condition 3). By choosing $\varphi\in\mathcal{C}_c^1(\mathbb{R}^d)$ satisfying $\varphi\equiv 1$ on such equi-bounded support and using (H4), it holds

$$\partial_t|\mu_t^F|\leq 0+M_\alpha(|\mu_t^F|+|\mu_t^L|),$$

and similarly for $|\mu_t^L|$. This implies $|\mu_t^F|+|\mu_t^L|\leq e^{2M_\alpha t}(|\mu_0^F|+|\mu_0^L|)$, hence masses are equi-bounded too.

For the given solution μ_t, define the corresponding vector field and source term

$$w_t:=\sum_{j\in\{F,L\}}K^j*\mu_t^j,\qquad s_t:=-\alpha_F(\mu_t^F,\mu_t^L)\mu_t^F+\alpha_L(\mu_t^F,\mu_t^L)\mu_t^L.$$

Consider them as time-varying operators, not depending on μ. By construction, it holds

$$\partial_t\mu_t^F=-\mathrm{div}(w_t\mu_t^F)+s_t. \tag{55}$$

Observe that w_t is a time-varying vector field, continuous with respect to the time variable and uniformly Lipschitz with respect to the space variable, due to (H3), (47) and equi-boundedness of $|\mu_t|$. It is then a Carathéodory function. Moreover, s_t is continuous with respect to time, with uniformly bounded mass due to (H4) and with uniformly bounded support due to (H5). Then, hypotheses of Lemma 2.9 are satisfied, hence μ_t^F is the unique solution of (55) and it satisfies the Duhamel's formula (12). It is clear that the previous properties hold for μ^L too, with the same vector field w_t and source $-s_t$. Moreover, the same properties hold for ν^F too, with vector field and source term

$$w_t':=\sum_{j\in\{F,L\}}K^j*\nu_t^j,\qquad s_t':=-\alpha_F(\nu_t^F,\nu_t^L)\nu_t^F+\alpha_L(\nu_t^F,\nu_t^L)\nu_t^L,$$

as well as for ν_t^L, with w_t' and $-s_t'$.

We now compute $\mathcal{W}_g(\mu_t,\nu_t)$ by using the Duhamel's formula and the Kantorovich-Rubinstein duality. Take f such that $\|f\|_{\mathcal{C}^0},\mathrm{Lip}(f)\leq 1$ and compute

$$\int_{\mathbb{R}^d}f\,d(\mu_t^F-\nu_t^F)=\int_{\mathbb{R}^d}f\,d(\Phi_t^{w_t}\#\mu_0^F-\Phi_t^{w_t'}\#\nu_0^F)+$$
$$\int_0^t d\tau\int_{\mathbb{R}^d}f\,d(\Phi_{(\tau,t)}^{w}\#s_\tau-\Phi_{(\tau,t)}^{w'}\#s_\tau')\leq\mathcal{W}_g\left(\Phi_t^{w_t}\#\mu_0^F,\Phi_t^{w_t'}\#\nu_0^F\right)+$$
$$\int_0^t d\tau\int_{\mathbb{R}^d}f\,d\left(-\Phi_{(\tau,t)}^{w}\#\alpha_F(\mu_\tau^F,\mu_\tau^L)\mu_\tau^F+\Phi_{(\tau,t)}^{w'}\#\alpha_F(\nu_\tau^F,\nu_\tau^L)\nu_\tau^F\right)+$$

$$\begin{aligned}
&\int_0^t d\tau \int_{\mathbb{R}^d} f\, d\left(\Phi^{w}_{(\tau,t)}\#\alpha_L(\mu^F_\tau,\mu^L_\tau)\mu^L_\tau - \Phi^{w'}_{(\tau,t)}\#\alpha_L(\nu^F_\tau,\nu^L_\tau)\nu^L_\tau\right) \leq \\
&e^{2Lt}\mathcal{W}_g\left(\mu^F_0,\nu^F_0\right) + |\mu^F_0|\frac{e^{2Lt}(e^{Lt}-1)}{L}\sup_{\tau\in[0,t]}\{\|w_\tau - w'_\tau\|_{\mathcal{C}^0}\} + \\
&\int_0^t d\tau e^{2L(t-\tau)}\left(\mathcal{W}_g\left(\alpha_F(\mu^F_\tau,\mu^L_\tau)\mu^F_\tau,\alpha_F(\nu^F_\tau,\nu^L_\tau)\nu^F_\tau\right)+\right. \\
&\qquad\qquad \left.\mathcal{W}_g\left(\alpha_L(\mu^F_\tau,\mu^L_\tau)\mu^L_\tau,\alpha_L(\nu^F_\tau,\nu^L_\tau)\nu^L_\tau\right)\right) + \\
&\int_0^t d\tau(|\alpha_F(\mu^F_0,\mu^L_0)\mu^F_0| + |\alpha_L(\mu^F_0,\mu^L_0)\mu^L_0|)\frac{e^{2L(t-\tau)}(e^{L(t-\tau)}-1)}{L}\sup_{\tau'\in[\tau,t]}\{\|w_{\tau'}-w'_{\tau'}\|_{\mathcal{C}^0}\},
\end{aligned}\tag{56}$$

where L is a Lipschitz constant for both w_τ, w'_τ, that exists by (47), and where we also used (26). Observe that it holds

$$\begin{aligned}
|(K^F * \mu^F_t)(x) - (K^F * \nu^F_t)(x)| \;\leq\; & \left|\int_{\mathbb{R}^d} K^F(z-x)\, d(\mu^F_t(z) - \nu^F_t(z))\right| \leq \\
& L_K \mathcal{W}_g\left(\mu^F_t,\nu^F_t\right)
\end{aligned}\tag{57}$$

where we used the Kantorovich-Rubinstein duality and (H2). The same estimate holds for K^L, thus $|w_t(x) - w'_t(x)| \leq L_K \mathcal{W}_g(\mu_t,\nu_t)$.

Going back to (56), recall that $\mathcal{W}_g(\mu_0,\nu_0) = 0$, since the initial data coincide. Also apply the estimate (26) and hypothesis (H4). Define

$$\varepsilon(t) := \sup_{\tau\in[0,t]} \mathcal{W}_g(\mu_\tau,\nu_\tau)$$

and observe that it holds

$$\begin{aligned}
&\int_{\mathbb{R}^d} f\, d(\mu^F_t - \nu^F_t) \leq 0 + |\mu^F_0|\frac{e^{2Lt}(e^{Lt}-1)}{L}L_K\varepsilon(t) + \frac{e^{2Lt}-1}{L}L'_{\alpha,M,R}\varepsilon(t) + \\
&M_\alpha(|\mu^F_0| + |\mu^L_0|)\frac{(e^{2Lt}-1)(e^{Lt}-1)}{L}L_K\varepsilon(t)
\end{aligned}$$

Since the left hand side does not depend on f, one can take the supremum over f satisfying $\|f\|_{\mathcal{C}^0}, \mathrm{Lip}(f) \leq 1$, i.e. replace it with $\mathcal{W}_g\left(\mu^F_t,\nu^F_t\right)$. The equivalent estimate holds for $\mathcal{W}_g\left(\mu^L_t,\nu^L_t\right)$. Merging them, it holds

$$\mathcal{W}_g(\mu_t,\nu_t) \leq C_t\varepsilon(t), \tag{58}$$

with

$$C_t := (|\mu^F_0| + |\mu^L_0|)\frac{(e^{Lt}-1)}{L}L_K\left(e^{2Lt} + 2M_\alpha(e^{2Lt}-1)\right) + 2\frac{e^{2Lt}-1}{L}L'_{\alpha,M,R}.$$

Since the right hand side in (58) is an increasing function with respect to t, one can replace $\mathcal{W}_g(\mu_t,\nu_t)$ with $\varepsilon(t)$ on the left hand side. It then holds

$$\varepsilon(t) \leq C_t\varepsilon(t).$$

Since $\lim_{t\to 0} C_t = 0$ and C_t is continuous, it holds $\varepsilon(t) = 0$ for t sufficiently small. By iterating the estimate, this holds for any $t \in [0,T]$, thus $\mu_t = \nu_t$ for all $t \in [0,T]$.

□

3.3 Equivalence between systems (1) and (3)

We now prove that, if (H1) is satisfied, then systems (1) and (3) are equivalent, in the sense that there exists a bijection between solutions. We also use this equivalence to prove existence and uniqueness of solutions to system (3).

Proposition 3.4. *Let (μ_t^F, μ_t^L) be a solution to system* (1)*, such that (μ_0^F, μ_0^L) satisfies (H1). Assume that hypotheses (H2)-(H3)-(H4)-(H5) hold. Define*

$$\nu_t := \mu_t^F + \mu_t^L, \qquad (\sigma_t(F), \sigma_t(L)) := \left(\frac{|\mu_t^F|}{|\nu_t|}, \frac{|\mu_t^L|}{|\nu_t|}\right). \tag{59}$$

Then, (ν_t, σ_t) is a solution to system (3)*.*

Conversely, let the hypotheses (H2)-(H3)-(H4)-(H5) hold and let (ν_t, σ_t) be a solution to system (3)*. Define*

$$\mu_t^F = \sigma_t(F)\nu_t, \qquad \mu_t^L = \sigma_t(L)\nu_t. \tag{60}$$

Then, (μ_t^F, μ_t^L) is a solution to system (1)*.*

Proof. We prove **Statement 1**.Take (μ_t^F, μ_t^L) a solution to system (1) with (μ_0^F, μ_0^L) satisfying (H1). Define (ν_t, σ_t) according to (59). By a direct computation, it holds

$$\partial_t \nu_t = -\mathrm{div}((K^F * \mu_t^F + K^L * \mu_t^L)\nu_t). \tag{61}$$

This also implies that $|\nu_t|$ is constant. Define now σ_t according to (59), and compute

$$\begin{aligned}\partial_t \sigma_t(F) &= \frac{\partial_t |\mu_t^F|}{|\nu_t|} = \frac{-\alpha_F(\mu_t^F, \mu_t^L)|\mu_t^F| + \alpha_L(\mu_t^F, \mu_t^L)|\mu_t^L|}{|\nu_t|} = \\ &= -\alpha_F(\mu_t^F, \mu_t^L)\sigma_t(F) + \alpha_L(\mu_t^F, \mu_t^L)\sigma_t(L). \end{aligned} \tag{62}$$

We used the fact that $|\nu_t|$ is constant, as a consequence of (61), and the definition of σ_t. One easily recovers $\sigma_t(L) = 1 - \sigma_t(F)$, hence $\partial_t \sigma_t(L) = -\partial_t \sigma_t(F)$. The difficulty is now to prove that it holds $\mu_t^F = \sigma_t(F)\nu_t, \mu_t^L := \sigma_t(L)\nu_t$ for all times.

Since μ_t^F, μ_t^L are given, one can define the non-autonomous vector field and the coefficients for the source term

$$v_t := K^F * \mu_t^F + K^L * \mu_t^L, \qquad h_t^F := \alpha_F(\mu_t^F, \mu_t^L), \qquad h_t^L := \alpha_L(\mu_t^F, \mu_t^L).$$

Define

$$\tilde{\mu}_t^F := \sigma_t(F)\nu_t, \qquad \tilde{\mu}_t^L := \sigma_t(L)\nu_t$$

Observe that it holds $\tilde{\mu}_0^F = \mu_0^F$ and $\tilde{\mu}_0^L = \mu_0^L$, as a consequence of (H1). Using (61)-(62), it holds

$$\begin{aligned}\partial_t \tilde{\mu}_t^F &= -\mathrm{div}((K^F * \mu_t^F + K^L * \mu_t^L)\sigma_t(F)\nu_t) - \alpha_F(\mu_t^F, \mu_t^L)\sigma_t(F)\nu_t + \\ &\quad \alpha_L(\mu_t^F, \mu_t^L)\sigma_t(L)\nu_t = -\mathrm{div}(v_t \tilde{\mu}_t^F) - h_t^F \tilde{\mu}_t^F + h_t^L \tilde{\mu}_t^L.\end{aligned}$$

One similarly has $\partial_t\tilde{\mu}^L_t = -\mathrm{div}(v_t\tilde{\mu}^L_t) + h^F_t\tilde{\mu}^F_t - h^L_t\tilde{\mu}^L_t$. By construction, it also holds

$$\partial_t\mu^F_t = -\mathrm{div}(v_t\mu^F_t) - h^F_t\mu^F_t + h^L_t\mu^L_t, \qquad \partial_t\mu^L_t = -\mathrm{div}(v_t\mu^L_t) + h^F_t\mu^F_t - h^L_t\mu^L_t.$$

Take f such that $\|f\|_{\mathcal{C}^0}, \mathrm{Lip}(f) \leq 1$, and apply the Duhamel's formula for both $\mu_t, \tilde{\mu}_t$. It holds

$$\int_{\mathbb{R}^d} f\, d(\mu^F_t - \tilde{\mu}^F_t) = \int_{\mathbb{R}^d} f\, d(\Phi^{v_t}_t\#\mu^F_0 - \Phi^{v_t}_t\#\tilde{\mu}^F_0) + \tag{63}$$

$$\int_0^t d\tau \left[h^F_\tau \int_{\mathbb{R}^d} f\, d(\Phi^{v_t}_{(\tau,t)}\#\tilde{\mu}^F_\tau - \Phi^{v_t}_{(\tau,t)}\#\mu^F_\tau) + h^L_\tau \int_{\mathbb{R}^d} f\, d(\Phi^{v_t}_{(\tau,t)}\#\mu^L_\tau - \Phi^{v_t}_{(\tau,t)}\#\tilde{\mu}^L_\tau) \right] \leq$$

$$0 + \int_0^t d\tau (h^F_\tau \mathcal{W}_g\left(\Phi^{v_t}_{(\tau,t)}\#\mu^F_\tau, \Phi^{v_t}_{(\tau,t)}\#\tilde{\mu}^F_\tau\right) + h^L_\tau \mathcal{W}_g\left(\Phi^{v_t}_{(\tau,t)}\#\mu^L_\tau, \Phi^{v_t}_{(\tau,t)}\#\tilde{\mu}^L_\tau\right). \tag{64}$$

Here we used the fact that $\mu^F_0 = \tilde{\mu}^F_0$ implies $\Phi^{v_t}_t\#\mu^F_0 = \Phi^{v_t}_t\#\tilde{\mu}^F_0$, as well as the Kantorovich-Rubinstein duality. Denote with L a Lipschitz constant for v_t, that exists due to (47), and apply (17). Observe that (64) does not depend on f, thus one can take the supremum in the left hand side of (63) with $\|f\|_{\mathcal{C}^0}, \mathrm{Lip}(f) \leq 1$, i.e. replace it with $\mathcal{W}_g\left(\mu^F_\tau, \tilde{\mu}^F_\tau\right)$. Also observe that (H4) implies $|h^F_\tau|, |h^L_\tau| \leq M_\alpha$. By defining $\varepsilon(t) := \sup_{\tau\in[0,t]} \mathcal{W}_g\left(\mu_\tau, \tilde{\mu}_\tau\right)$, it holds

$$\mathcal{W}_g\left(\mu^F_\tau, \tilde{\mu}^F_\tau\right) \leq te^{2Lt} M_\alpha \varepsilon(t),$$

and the same holds for $\mathcal{W}_g\left(\mu^L_\tau, \tilde{\mu}^L_\tau\right)$.

Observe that the right hand side is increasing with respect to t, thus one can replace the left hand side with $\varepsilon(t)$. It then holds $\varepsilon(t) \leq 2te^{2Lt}M_\alpha\varepsilon(t)$, thus $\varepsilon(t) = 0$ for t sufficiently small. Applying then the result iteratively, it holds $\varepsilon(t) = 0$ for all $t \in [0, T]$, then $\mu_t = \tilde{\mu}_t$, thus

$$\mu^F_t = \sigma_t(F)\nu_t \qquad \text{and} \qquad \mu^L_t = \sigma_t(L)\nu_t.$$

We prove **Statement 2**. Since (ν_t, σ_t) is a solution to system (3) in the sense of Definition 2.6, then (μ^F_t, μ^L_t) defined by (60) satisfies Conditions 2 and 3 of Definition 2.5. Condition 1 is also satisfied, by trivially choosing $\bar{\mu}^F = \mu^F_0$ and $\bar{\mu}^L = \mu^L_0$. We are left to prove that Condition 4 is satisfied: the proof is direct, by computing derivatives. □

As a corollary to Proposition 3.4, we prove existence and uniqueness of solutions to system (3).

Corollary 3.5. *Let the hypotheses (H2)-(H3)-(H4)-(H5) hold. Then, for each initial data $(\overline{\nu}, \overline{\sigma}) \in \mathcal{M}(\mathbb{R}^d) \times \mathcal{P}(\{F, L\})$, there exists a unique solution to system* (3).

Proof. For the existence part, define

$$\bar{\mu}^F := \overline{\sigma}(F)\overline{\nu}, \qquad \bar{\mu}^L := \overline{\sigma}(F)\overline{\nu}$$

and consider the corresponding solution (μ_t^F, μ_t^L) to (1), that exists due to Proposition 3.3. Then, there exists a corresponding solution (ν_t, σ_t) to system (3), due to the first statement of Proposition 3.4. Such solution satisfies $(\nu_0, \sigma_0) = (\overline{\nu}, \overline{\sigma})$, by construction.

For the uniqueness part, assume that there exist two solutions $(\nu_t, \sigma_t), (\tilde{\nu}_t, \tilde{\sigma}_t)$ to (3) with the same initial data $(\overline{\nu}, \overline{\sigma})$. Due to the second statement of Proposition 3.4, for each of the two solutions to (3) there exists a solution $(\mu_t^F, \mu_t^L), (\tilde{\mu}_t^F, \tilde{\mu}_t^L)$ to system (1). It clearly holds $(\mu_0^F, \mu_0^L) = (\tilde{\mu}_0^F, \tilde{\mu}_0^L)$, then such two solutions coincide, due to uniqueness of the solution to system (1). Since the relation (60) is invertible, this implies $(\nu_t, \sigma_t) = (\tilde{\nu}_t, \tilde{\sigma}_t)$. □

Remark 3.6. By inspection of our proofs, other types of measure-dependent velocity fields can be encompassed by our approach, as long as the dependence is Lipschitz with repect to $\mathcal{W}_g$ (see, e.g., [43]). For instance, instead of the convolution term $K^i * \mu^i$ for $i = F$ or $i = L$, one could simply consider a weighted velocity of the form $|\mu^i| K^i(x)$ which still allows for proving the existence, uniqueness and equivalence results of this section. Accordingly, in the equivalent system to be considered in Proposition 3.4 one has to consider a velocity field of the form (if $i = L$ in the equation above)

$$\sigma_t(F) K^F * \nu_t + \sigma_t(L) K^L$$

for which also the mean-field derivation of Section 4 can be performed without changing the proofs.

4 A mean-field description of the leader-follower dynamics

In this section we shall provide a mean-field description of system (3). To do this, we shall first introduce for every $N \in \mathbb{N}$ a particle system which consists of a transport part for the evolution over the state space $\mathbb{R}^d$ and a jump part for the change of label in $\{F, L\}$.

The connection between systems of interacting particles and nonlinear evolution equations has been studied by many authors, going back to McKean [40]; for detailed expositions on this topic, the reader may consult Sznitman [46] or Méléard [41]. A central point of this connection is the introduction of a nonlinear *averaged* particle system associated with the original one, whose marginal laws appear explicitly (and nonlinearly) in the generator of its dynamics. When the interactions are regular and the particles are exchangeable, a unique nonlinear process exists, and it describes the limiting behavior of one particle of the original system when their number tends to infinity. One further has the propagation of chaos property, which is in this case equivalent to a trajectorial law of large numbers and yields the final mean-field limit result.

4.1 Definition of the stochastic processes

Throughout this section, we shall fix $\overline{\nu} \in \mathcal{P}_1(\mathbb{R}^d)$ and $\overline{\sigma} \in \mathcal{P}_1(\{F, L\})$, as well as, for every $N > 0$, a sample of N particles from $\overline{\nu} \times \overline{\sigma}$, i.e.,

$$(X_0^{i,N}, Y_0^{i,N})_{i=1}^N \sim \overline{\nu} \times \overline{\sigma}.$$

We assume that $\overline{\nu}$ has compact support in $\mathbb{R}^d$.

We introduce the stochastic processes $(X_t^{1,N}, Y_t^{1,N}), \dots, (X_t^{N,N}, Y_t^{N,N})$ defined for every $t \in [0,T]$ and $i = 1, \dots, N$ as follows

- the initial conditions are $(X_0^{i,N}, Y_0^{i,N})_{i=1}^N$,
- we set

$$\nu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}} \quad \text{and} \quad \sigma_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^{i,N}}, \tag{65}$$

- it holds

$$dX_t^{i,N} = \langle K, \nu_t^N \times \sigma_t^N \rangle (X_t^{i,N}) dt \tag{66}$$

- the conditional transition rates for $Y^{i,N}$ at time t, for a realization of (ν_t^N, σ_t^N), are given by
  - if $Y_t^{i,N} = F$ then $F \to L$ with rate $\alpha_F(\nu_t^N, \sigma_t^N)$,
  - if $Y_t^{i,N} = L$ then $L \to F$ with rate $\alpha_L(\nu_t^N, \sigma_t^N)$,

  where we used the shorthand notation (27) for α_F and α_L.

More formally, we define the processes $(Y_t^{1,N}, \dots, Y_t^{N,N})$ to be the jump processes such that $\mathrm{Law}(Y_t^{i,N})$ satisfy the system of ODEs

$$\frac{d}{dt}\mathrm{Law}(Y_t^{i,N}) = \mathbb{E}\left(A_{\nu_t^N, \sigma_t^N}\right) \mathrm{Law}(Y_t^{i,N}) \quad i = 1, \dots, N. \tag{67}$$

Notice that (67) clearly stems out of the above definitions and the law of total probability, averaging on all realizations of (ν_t^N, σ_t^N).

Remark 4.1. We shortly discuss the well-posedness of the above defined processes, leaving the details to the reader.

For a realization of $(Y_t^{1,N}, \dots, Y_t^{N,N})$ in the space of cádlág functions, the applications $\langle K, \nu_t^N \times \sigma_t^N \rangle$ and $A_{\nu_t^N, \sigma_t^N}$ are both measurable and bounded in time. Thus, (67) has a right-hand side which is measurable and bounded with respect to t and Lipschitz continuous with resect to X, uniformly for $t \in [0,T]$. Hence, the existence of Lipschitz continuous solutions to (66) uniquely determined by the initial data follows directly from the general theory in [28].

Concerning the stochastic processes $Y_t^{i,N}$ with law given by (67), they can be, for instance, realised as limit of discrete-in-time processes of the form

$$\begin{pmatrix} \mathbb{P}\left\{Y_{t+h}^{i,N} = F | (\nu_t^N, \sigma_t^N) = (\tilde{\nu}, \tilde{\sigma})\right\} \\ \mathbb{P}\left\{Y_{t+h}^{i,N} = L | (\nu_t^N, \sigma_t^N) = (\tilde{\nu}, \tilde{\sigma})\right\} \end{pmatrix} = (I + hA_{\tilde{\nu}, \tilde{\sigma}}) \begin{pmatrix} \mathbb{P}\left\{Y_t^{i,N} = F\right\} \\ \mathbb{P}\left\{Y_t^{i,N} = L\right\} \end{pmatrix} \quad i = 1, \dots, N$$

for I being the identity matrix and $h > 0$ a vanishing time step. In the equation above notice that, since by construction the vector $(1,1)^T$ belongs to the kernel of the transpose $A^*_{\tilde{\nu},\tilde{\sigma}}$ of $A_{\tilde{\nu},\tilde{\sigma}}$ for every realization $(\tilde{\nu},\tilde{\sigma})$ of (ν_t^N, σ_t^N), the left-hand side above is well-defined as a conditional probability law on $\{F, L\}$.

Remark 4.2. Since $\overline{\nu}$ has compact support in $\mathbb{R}^d$, and using (4), standard arguments (see, e.g. [30, Appendix]) entail that it exists $R_T > 0$ such that, for all $N \in \mathbb{N}$, $i = 1, \ldots, N$ and $t \in [0,T]$, it holds

$$|X_t^{i,N}| \leq R_T, \text{ so that } \operatorname{supp}(\nu_t^N) \subset B(0, R_T)\,. \tag{68}$$

This inclusion has clearly to be understood as being verified with probability 1.

Our next step is defining, for fixed i and N, an auxiliary averaged process $(\overline{X}_t^{i,N}, \overline{Y}_t^{i,N})$ having the solutions ν_t and σ_t of system (3) as laws. To this purpose, we need some preparation which will be useful also in the sequel.

Proposition 4.3. *Let (ν_t, σ_t) be a solution of (3) and define a process $(\overline{X}_t, \overline{Y}_t)$ as follows*

- $\overline{X}_0 \sim \nu_0$ *and* $\overline{Y}_t \sim \sigma_0$,
- $d\overline{X}_t = \langle K, \nu_t \times \sigma_t \rangle (\overline{X}_t) dt$,
- *the transition rates at time t are given by*
  - *if* $\overline{Y}_t = F$ *then* $F \to L$ *with rate* $\alpha_F(\nu_t, \sigma_t)$,
  - *if* $\overline{Y}_t = L$ *then* $L \to F$ *with rate* $\alpha_L(\nu_t, \sigma_t)$.

Then $\nu_t = \operatorname{Law}(\overline{X}_t)$ *and* $\sigma_t = \operatorname{Law}(\overline{Y}_t)$.

Proof. Define $\eta_t = \operatorname{Law}(\overline{X}_t)$ and let $\varphi \in \mathcal{C}_c^1(\mathbb{R}^d)$ be any test function. For $v_t = \langle K, \nu_t \times \sigma_t \rangle$, by definition of $\overline{X}_t$ and linearity of the expected values it holds that

$$\partial_t \langle \varphi, \eta_t \rangle = \partial_t \mathbb{E}[\varphi(\overline{X}_t)] = \langle \varphi, -\operatorname{div}(v_t \eta_t) \rangle.$$

The initial condition $\eta_0 = \operatorname{Law}(\overline{X}_0) = \nu_0$ holds by definition. Hence, $\operatorname{Law}(\overline{X}_t)$ is a solution to the PDE

$$\begin{cases} \partial_t \eta_t = -\operatorname{div}(\langle K, \nu_t \times \sigma_t \rangle \eta_t), \\ \eta(0) = \nu_0, \end{cases}$$

which is unique by Lemma 2.9. Since ν_t solves the same problem, we get $\nu_t = \operatorname{Law}(\overline{X}_t)$. Moreover, as both σ_t and $\operatorname{Law}(\overline{Y}_t)$ are solutions of

$$\dot{\eta}_t = A_{\nu_t, \sigma_t} \eta_t$$

with initial condition σ_0, then again by uniqueness we have $\sigma_t = \operatorname{Law}(\overline{Y}_t)$. □

We are now in a position to define, for fixed i and N, the processes $\overline{X}_t^{i,N}$ and $\overline{Y}_t^{i,N}$ through the following dynamics:

- $\overline{X}_0^{i,N} = X_0^{i,N}$ and $\overline{Y}_0^{i,N} = Y_0^{i,N}$,
- $\mathrm{Law}(\overline{X}_t^{i,N}) = \nu_t$ and $\mathrm{Law}(\overline{Y}_t^{i,N}) = \sigma_t$,
- $d\overline{X}_t^{i,N} = \langle K, \nu_t \times \sigma_t \rangle (\overline{X}_t^{i,N}) dt$,
- the transition rates at time t are given by
  - if $\overline{Y}_t^{i,N} = F$ then $F \to L$ with rate $\alpha_F(\nu_t, \sigma_t)$,
  - if $\overline{Y}_t^{i,N} = L$ then $L \to F$ with rate $\alpha_L(\nu_t, \sigma_t)$.

The well-posedness of such processes is indeed a corollary of our previous results.

Corollary 4.4. *The processes* $\overline{X}_t^{i,N}$ *and* $\overline{Y}_t^{i,N}$ *exist for every* $N \in \mathbb{N}$ *and every* $i = 1, \dots, N$.

Proof. Follows from Proposition 4.3 and the existence of (ν_t, σ_t) from Corollary 3.5. □

For every fixed t, the processes $\overline{X}_t^{i,N}$ with $i = 1, \dots, N$ are clearly independent of each other, and so are the processes $\overline{Y}_t^{i,N}$.

Now, all the above constructions still leave one free to choose how to couple the processes $Y_t^{i,N}$ and $\overline{Y}_t^{i,N}$ in their product space[5] : we namely assume that

$$\mathrm{Law}\left(Y_t^{i,N}, \overline{Y}_t^{i,N}\right) \in \Gamma_o(\mathrm{Law}(Y_t^{i,N}), \sigma_t) .$$

Here optimality of the transportation plans is meant with respect of the distance $\mathcal{W}_1$ on $\mathcal{P}_1(\{F, L\})$, where (as everywhere in what follows) the set $\{F, L\}$ is endowed with the distance (28). With the above choice, by the definition of $\mathcal{W}_1$ we have

$$\mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| = \mathcal{W}_1(\mathrm{Law}(Y_t^{i,N}), \sigma_t) \tag{69}$$

for all i, N, and t. Since $Y_t^{i,N}$ and $\overline{Y}_t^{i,N}$ are random variables on the discrete space $\{F, L\}$, a simple computation using (7) together with (69) entails that

$$\mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| = \left| \mathbb{P}\{Y_t^{i,N} = F\} - \mathbb{P}\{\overline{Y}_t^{i,N} = F\} \right| . \tag{70}$$

Remark 4.5. The relationship between the empirical mean of the independent processes $(\overline{X}_t^{i,N}, \overline{Y}_t^{i,N})_{i=1,\dots,N}$ given by

$$\overline{\nu}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{\overline{X}_t^{1,N}} \qquad \text{and} \qquad \overline{\sigma}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{\overline{Y}_t^{i,N}} \tag{71}$$

[5] The coupling between $X_t^{i,N}$ and $\overline{X}_t^{i,N}$ is as usual tacitly defined by asking that $X_t^{i,N} - \overline{X}_t^{i,N}$ solves the SDE obtained as difference of the ones for $X_t^{i,N}$ and $\overline{X}_t^{i,N}$, respectively.

and (ν_t, σ_t) solution of system (3) is clear: by the Glivenko-Cantelli's theorem, $(\overline{\nu}_t^N, \overline{\sigma}_t^N)$ converges weakly to (ν_t, σ_t) as $N \to +\infty$. The rate of convergence can actually be quantified thanks to [31, Theorem 1] (which holds for ν_t and σ_t, since their support is uniformly compact in time): we may apply it once for σ_t and the values $p = d = 1, q = 3$ to get for every $t \in [0, T]$

$$\mathbb{E}\left[\mathcal{W}_1(\overline{\sigma}_t^N, \sigma_t)\right] \leq K_1(N^{-1/2} + N^{-2/3}), \tag{72}$$

for a given constant $K_1 > 0$. If we apply it for ν_t and the values $p = 2d, q = 3p$ (where d here denotes the dimension of the state space $\mathbb{R}^d$) we get for every $t \in [0, T]$

$$\mathbb{E}\left[\mathcal{W}_{2d}(\overline{\nu}_t^N, \nu_t)^{2d}\right] \leq K_2(N^{-1/2} + N^{-2/3}),$$

for some $K_2 > 0$. Since it is well-known that $\mathcal{W}_1 \leq \mathcal{W}_p$ for any $p \geq 1$, an application of Jensen's inequality yields the estimate

$$\mathbb{E}\left[\mathcal{W}_1(\overline{\nu}_t^N, \nu_t)\right] \leq K_2^{1/2d}(N^{-1/4d} + N^{-1/3d}). \tag{73}$$

Setting

$$\Theta(N) := K_1(N^{-1/2} + N^{-2/3}) + K_2^{1/2d}(N^{-1/4d} + N^{-1/3d}),$$

and putting together (73) and (72) we obtain

$$\mathbb{E}\left[\mathcal{W}_1(\overline{\nu}_t^N, \nu_t)\right] + \mathbb{E}\left[\mathcal{W}_1(\overline{\sigma}_t^N, \sigma_t)\right] \leq \Theta(N). \tag{74}$$

Remark 4.6 (Exchangeability of processes)**.** Notice a fundamental property of the processes $(\overline{X}_t^{i,N}, \overline{Y}_t^{i,N})_{i=1,\ldots,N}$: for every $i, j = 1, \ldots, N$ and every $t \geq 0$ we have

$$\mathbb{E}\left|X_t^{i,N} - \overline{X}_t^{i,N}\right| = \mathbb{E}\left|X_t^{j,N} - \overline{X}_t^{j,N}\right| \quad \text{and} \quad \mathbb{E}\left|Y_t^{i,N} - \overline{Y}_t^{i,N}\right| = \mathbb{E}\left|Y_t^{j,N} - \overline{Y}_t^{j,N}\right|.$$

After noticing that both identities hold trivially at $t = 0$, this clearly follows from the simmetry of the processes $(X_t^{i,N}, Y_t^{i,N})$ and the fact that $(\overline{X}_t^{i,N}, \overline{Y}_t^{i,N})_{i=1,\ldots,N}$ are independent. In particular, this exchangeability implies that

$$\frac{1}{N}\sum_{j=1}^N \mathbb{E}\left|X_t^{j,N} - \overline{X}_t^{j,N}\right| = \mathbb{E}\left|X_t^{i,N} - \overline{X}_t^{i,N}\right|,$$

as well as

$$\frac{1}{N}\sum_{j=1}^N \mathbb{E}\left|Y_t^{j,N} - \overline{Y}_t^{j,N}\right| = \mathbb{E}\left|Y_t^{i,N} - \overline{Y}_t^{i,N}\right|.$$

4.2 The mean-field limit

The main goal of this section is to show that, for N large, the random empirical distributions ν_t^N and σ_t^N associated to the processes $(X_t^{1,N}, Y_t^{1,N}), \dots, (X_t^{N,N}, Y_t^{N,N})$ defined in the previous subsection are close, in a probabilistic sense, to the deterministic measures ν_t and σ_t, solutions of system (3) with initial datum $(\overline{\nu}, \overline{\sigma})$. The result we aim to prove is namely the following.

Theorem 4.7. *Under Assumptions (H2) and (H5), there exists a function* $\Psi : \mathbb{N} \to \mathbb{R}_+$ *satisfying* $\lim_{N\to+\infty} \Psi(N) = 0$ *such that*

$$\sup_{t\geq 0} \mathbb{P}\left(\mathcal{W}_1(\nu_t^N, \nu_t) + \mathcal{W}_1(\sigma_t^N, \sigma_t) > \varepsilon\right) \leq \varepsilon^{-1}\Psi(N).$$

For the proof of Theorem we will need a key intermediate result that we state below.

Lemma 4.8 (Uniform propagation of chaos)**.** *Define the empirical measures* ν_t^N*,* σ_t^N*,* $\overline{\nu}_t^N$*, and* $\overline{\sigma}_t^N$ *through* (65)*, and* (71)*, respectively. Under Assumptions (H2) and (H5), there exists a function* $\Phi : \mathbb{N} \to \mathbb{R}_+$ *satisfying* $\lim_{N\to+\infty} \Phi(N) = 0$ *such that*

$$\sup_{t\geq 0} \mathbb{E}\left[\left|X_i^N(t) - \overline{X}_i^N(t)\right| + \left|Y_i^N(t) - \overline{Y}_i^N(t)\right|\right] \leq \Phi(N). \tag{75}$$

The proof of this result is postponed to the next section. Let us first show how Theorem 4.7 can be easily derived, once Lemma 4.8 is established.

Proof of Theorem 4.7. By the triangular inequality it follows that

$$\begin{aligned} \mathcal{W}_1(\nu_t^N, \nu_t) &\leq \mathcal{W}_1(\nu_t^N, \overline{\nu}_t^N) + \mathcal{W}_1(\overline{\nu}_t^N, \nu_t), \\ \mathcal{W}_1(\sigma_t^N, \sigma_t) &\leq \mathcal{W}_1(\sigma_t^N, \overline{\sigma}_t^N) + \mathcal{W}_1(\overline{\sigma}_t^N, \sigma_t). \end{aligned} \tag{76}$$

Since $\nu_t^N, \overline{\nu}_t^N, \sigma_t^N$ and $\overline{\sigma}_t^N$ are all atomic measures, by the properties of the Wasserstein distance we have

$$\mathcal{W}_1(\nu_t^N, \overline{\nu}_t^N) \leq \frac{1}{N}\sum_{i=1}^N \left|X_t^{i,N} - \overline{X}_t^{i,N}\right| \quad \text{and} \quad \mathcal{W}_1(\sigma_t^N, \overline{\sigma}_t^N) \leq \frac{1}{N}\sum_{i=1}^N \left|Y_t^{i,N} - \overline{Y}_t^{i,N}\right|. \tag{77}$$

Therefore, by taking expectations in (76) and plugging (74) and (75) on the the right-hand side, we obtain

$$\mathbb{E}\left[\mathcal{W}_1(\nu_t^N, \nu_t) + \mathcal{W}_1(\sigma_t^N, \sigma_t)\right] \leq \Phi(N) + \Theta(N).$$

If we set

$$\Psi(N) := \Phi(N) + \Theta(N),$$

an application of Markov's inequality concludes the proof. □

4.3 Proof of Lemma 4.8

First, we start from the term $\mathbb{E}|X_t^{i,N} - \overline{X}_t^{i,N}|$. By integrating the dynamics of $X_t^{i,N}$ from 0 to t we obtain

$$X_t^{i,N} = X_0^{i,N} + \int_0^t \langle K, \nu_s^N \times \sigma_s^N \rangle (X_s^{i,N}) \, \mathrm{d}s,$$

and similarly for $\overline{X}_t^{i,N}$ we get

$$\overline{X}_t^{i,N} = X_0^{i,N} + \int_0^t \langle K, \nu_s \times \sigma_s \rangle (\overline{X}_s^{i,N}) \, \mathrm{d}s.$$

Above we used that, by definition, $\overline{X}_0^{i,N} = X_0^{i,N}$. Therefore, adding and subtracting the terms $\langle K, \overline{\nu}_s^N \times \overline{\sigma}_s^N \rangle (X_s^{i,N})$ and $\langle K, \overline{\nu}_s^N \times \overline{\sigma}_s^N \rangle (\overline{X}_s^{i,N})$, we get the estimate

$$\begin{aligned}
&\mathbb{E}|X_t^{i,N} - \overline{X}_t^{i,N}| = \\
&\quad = \mathbb{E}\left| X_0^{i,N} + \int_0^t \langle K, \nu_s^N \times \sigma_s^N \rangle (X_t^{i,N}) \, \mathrm{d}s - X_0^{i,N} - \int_0^t \langle K, \nu_s \times \sigma_s \rangle (\overline{X}_s^{i,N}) \, \mathrm{d}s \right| \\
&\quad \leq \underbrace{\mathbb{E}|X_0^{i,N} - X_0^{i,N}|}_{=0} + \underbrace{\mathbb{E}\left[\int_0^t |\langle K, \nu_s^N \times \sigma_s^N \rangle (X_s^{i,N}) - \langle K, \overline{\nu}_s^N \times \overline{\sigma}_s^N \rangle (X_s^{i,N})| \, \mathrm{d}s \right]}_{I_1} \\
&\qquad + \underbrace{\mathbb{E}\left[\int_0^t \left| \langle K, \overline{\nu}_s^N \times \overline{\sigma}_s^N \rangle (X_s^{i,N}) - \langle K, \overline{\nu}_s^N \times \overline{\sigma}_s^N \rangle (\overline{X}_s^{i,N}) \right| \, \mathrm{d}s \right]}_{I_2} \\
&\qquad + \underbrace{\mathbb{E}\left[\int_0^t \left| \langle K, \overline{\nu}_s^N \times \overline{\sigma}_s^N \rangle (\overline{X}_s^{i,N}) - \langle K, \nu_s \times \sigma_s \rangle (\overline{X}_s^{i,N}) \right| \, \mathrm{d}s \right]}_{I_3}. \qquad (78)
\end{aligned}$$

We shall now estimate from above the terms I_1, I_2 and I_3. Recall that (68) and property (3) in Definition 2.6 hold. This latter also gives that

$$|\overline{X}_t^{i,N}| \leq R_T, \text{ so that } \operatorname{supp}(\overline{\nu}_t^N) \subset B(0, R_T), \qquad (79)$$

with probability 1. With this, (4), and [30, Lemma A.7], for I_1 we have

$$\begin{aligned}
I_1 &\leq \mathbb{E}\left[\int_0^t L_K(\mathcal{W}_1(\nu_s^N, \overline{\nu}_s^N) + \mathcal{W}_1(\sigma_s^N, \overline{\sigma}_s^N)) \, \mathrm{d}s \right] \\
&= \int_0^t L_K(\mathbb{E}\left[\mathcal{W}_1(\nu_s^N, \overline{\nu}_s^N)\right] + \mathbb{E}\left[\mathcal{W}_1(\sigma_s^N, \overline{\sigma}_s^N)\right]) \, \mathrm{d}s \\
&\leq L_K \int_0^t \left(\frac{1}{N} \sum_{j=1}^N \mathbb{E}|X_s^{j,N} - \overline{X}_s^{j,N}| + \frac{1}{N} \sum_{j=1}^N \mathbb{E}|Y_s^{j,N} - \overline{Y}_s^{j,N}| \right) \mathrm{d}s \qquad (80) \\
&= L_K \int_0^t (\mathbb{E}|X_s^{i,N} - \overline{X}_s^{i,N}| + \mathbb{E}|Y_s^{i,N} - \overline{Y}_s^{i,N}|) \, \mathrm{d}s,
\end{aligned}$$

where we additionally used the inequalities (77) and Remark 4.6. With (4) and the same argument in (57) we deduce

$$I_2 \leq L_K \int_0^t \mathbb{E}|X_s^{i,N} - \overline{X}_s^{i,N}| \, \mathrm{d}s. \tag{81}$$

Finally, using (79) within the same steps used for I_1, together with (74), yields

$$\begin{aligned} I_3 &\leq \int_0^t L_K(\mathbb{E}\left[\mathcal{W}_1(\overline{\nu}_s^N, \nu_s)\right] + \mathbb{E}\left[\mathcal{W}_1(\overline{\sigma}_s^N, \sigma_s)\right]) \, \mathrm{d}s \\ &\leq L_K \Theta(N) t. \end{aligned} \tag{82}$$

By plugging (80), (81) and (82) into (78) we finally obtain

$$\begin{aligned} \mathbb{E}|X_t^{i,N} - \overline{X}_t^{i,N}| &\leq L_K \Theta(N) t \\ &\quad + 2L_K \int_0^t (\mathbb{E}|X_s^{i,N} - \overline{X}_s^{i,N}| + \mathbb{E}|Y_s^{i,N} - \overline{Y}_s^{i,N}|) \, \mathrm{d}s. \end{aligned} \tag{83}$$

We now turn to the term $\mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}|$. Using (67) and (70), and since $Y_0^{i,N} = \overline{Y}_0^{i,N}$ we have

$$\begin{aligned} &\mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| \leq \\ &\quad \leq \int_0^t \left| \frac{\mathrm{d}}{\mathrm{d}s} \mathbb{P}\{Y_s^{i,N} = F\} - \frac{\mathrm{d}}{\mathrm{d}s} \mathbb{P}\{\overline{Y}_s^{i,N} = F\} \right| \, \mathrm{d}s \\ &\quad \leq \int_0^t \left| \mathbb{E}(\alpha_F(\nu_s, \sigma_s)) \mathbb{P}\{\overline{Y}_s^{i,N} = F\} - \mathbb{E}(\alpha_F(\nu_s^N, \sigma_s^N)) \mathbb{P}\{Y_s^{i,N} = F\} \right| \, \mathrm{d}s \\ &\quad + \int_0^t \left| \mathbb{E}(\alpha_L(\nu_s^N, \sigma_s^N)) \mathbb{P}\{Y_s^{i,N} = L\} - \mathbb{E}(\alpha_L(\nu_s, \sigma_s)) \mathbb{P}\{\overline{Y}_s^{i,N} = L\} \right| \, \mathrm{d}s \\ &\quad \leq \int_0^t \mathbb{E}\left| \alpha_F(\nu_s, \sigma_s) - \alpha_F(\nu_s^N, \sigma_s^N) \right| \, \mathrm{d}s + \int_0^t \mathbb{E}\left| \alpha_L(\nu_s, \sigma_s) - \alpha_L(\nu_s^N, \sigma_s^N) \right| \, \mathrm{d}s \\ &\quad + \int_0^t \left(|\mathbb{E}(\alpha_F(\nu_s^N, \sigma_s^N))| + |\mathbb{E}(\alpha_L(\nu_s^N, \sigma_s^N))| \right) \left| \mathbb{P}\{Y_s^{i,N} = F\} - \mathbb{P}\{\overline{Y}_s^{i,N} = F\} \right| \mathrm{d}s, \end{aligned}$$

where we additionally exploited that clearly

$$\left| \mathbb{P}\{Y_s^{i,N} = F\} - \mathbb{P}\{\overline{Y}_s^{i,N} = F\} \right| = \left| \mathbb{P}\{Y_s^{i,N} = L\} - \mathbb{P}\{\overline{Y}_s^{i,N} = L\} \right| .$$

By Assumption (H4) there exists a uniform constant $L'_\alpha > 0$ such that $|\mathbb{E}(\alpha_F(\nu_s^N, \sigma_s^N))| + |\mathbb{E}(\alpha_L(\nu_s^N, \sigma_s^N))| \leq L'_\alpha$. We also recall that, by (68) and property (3) in Definition 2.6, ν_s^N and ν_s have by construction support contained in a compact set $B(0, R_T) \subset \mathbb{R}^d$ independent of N and s. We can therefore use (29) and (70), and continue the above estimate to get

$$\mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| \leq$$

$$\leq 2L_\alpha \int_0^t \left(\mathbb{E}\left[\mathcal{W}_1(\nu_s^N,\nu_s)\right] + \mathbb{E}\left[\mathcal{W}_1(\sigma_s^N,\sigma_s)\right]\right)\,\mathrm{d}s + L'_\alpha \int_0^t \mathbb{E}|Y_s^{i,N} - \overline{Y}_s^{i,N}|\,\mathrm{d}s\,.$$

With (76) and (77), plugging (74), and with Remark 4.6, we then have

$$\begin{aligned}
&\mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| \leq \\
&\quad \leq 2L_\alpha \Theta(N)t + 2L_\alpha \int_0^t \mathbb{E}|X_s^{i,N} - \overline{X}_s^{i,N}|\,\mathrm{d}s + (2L_\alpha + L'_\alpha)\int_0^t \mathbb{E}|Y_s^{i,N} - \overline{Y}_s^{i,N}|\,\mathrm{d}s \\
&\quad \leq 2L_\alpha \Theta(N)t + (2L_\alpha + L'_\alpha)\int_0^t \Big(\mathbb{E}|X_s^{i,N} - \overline{X}_s^{i,N}| + \mathbb{E}|Y_s^{i,N} - \overline{Y}_s^{i,N}|\Big)\,\mathrm{d}s\,.
\end{aligned}$$

Summing the above estimate to (83), we obtain the integral inequality

$$\begin{aligned}
\mathbb{E}|X_t^{i,N} - \overline{X}_t^{i,N}| + \mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| &\leq (L_K + 2L_\alpha)\Theta(N)t \\
&\quad + (2L_K + 2L_\alpha + L'_\alpha)\int_0^t \Big(\mathbb{E}|X_s^{i,N} - \overline{X}_s^{i,N}| + \mathbb{E}|Y_s^{i,N} - \overline{Y}_s^{i,N}|\Big)\,ds.
\end{aligned}$$

Hence, an application of Gronwall's inequality to the function $\mathbb{E}|X_t^{i,N} - \overline{X}_t^{i,N}| + \mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}|$ inside the interval $[0,T]$ yields

$$\mathbb{E}|X_t^{i,N} - \overline{X}_t^{i,N}| + \mathbb{E}|Y_t^{i,N} - \overline{Y}_t^{i,N}| \leq (L_K + 2L_\alpha)T\mathrm{e}^{(2L_K + 2L_\alpha + L'_\alpha)T} \cdot \Theta(N)\,,$$

which is the desired statement.

5 Numerical experiments and applications

We finally provide some practical applications of the present study, by numerically implementing some examples of social interaction dynamics. We will discuss the well-posedness of these examples according to theoretical assumptions. In particular, we remark that in all examples we account a bounded computational domain, therefore condition (H3) will be automatically satisfied. Numerical simulations are performed with a first-order finite volume scheme. Details of the implementation are reported in B.

5.1 Test I: Consensus dynamics

We aim to show the evolution of the mean-field system when the measures μ_t^F, μ_t^L have a bounded confidence interaction kernel. Therefore, we consider the Hegselmann-Krause type interactions

$$K^i(x) = a^i(x)x, \qquad a^i(x) = \chi_{\{|x| \leq C^i\}}(x), \qquad i \in \{F, L\} \tag{84}$$

where $C^F, C^L > 0$ are the *confidence thresholds* respectively for the followers, and leaders. We remark that Assumptions (H2) would require to replace the indicator functions a_i's with Lipschitz approximations thereof. When a^F, and a^L are two bounded,

Table 1: Computational parameters for Test I.

Test	C^F	C^L	α_F	α_L	$\sigma_0(F)$	$\sigma_0(L)$	δ_F	δ_L
Ia	0.2	0.6	0.1	0.95	0.75	0.25	–	–
Ib	0.2	0.6	(86)	(87)	0.75	0.25	0.35	0.2

Lipschitz continuous functions, a direct computation shows indeed that the functions $K^i(x) = a^i(x)x$ satisfy Assumptions (H2) inside $B(0, R_T)$, which is enough since our measures are compactly supported. On the other hand, the experiments are not affected by such a smoothing procedure, hence we keep the definition (84) throughout this section.

We want to solve numerically the evolution of the mean-field interaction dynamics, observing the impact of different choices of birth rates functions α_F, α_L. We select the computational domain $\Omega = [-1, 1]$ and system (1) complemented by zero-flux boundary conditions.

Let $\sigma_t(F)$ and $\sigma_t(L)$ be the total mass of followers and leaders at time t, respectively. Since the total mass is preserved, by renormalizing at the initial time, it holds $\sigma_t(L) + \sigma_t(F) = 1$. We assume that at time $t = 0$ the initial data is uniformly distributed in Ω with initial density $\sigma_0(L) = 1 - \sigma_t(F) = 0.75$.

We report in Table 1 the model parameters for two different test cases. In both cases, we assume that leaders have larger range of influence than followers, with $C^F = 0.2$ and $C^L = 0.6$. For the numerical discretization, we select $N = 80$ space grid points, time step $\Delta t = 0.0127$ and final time $T = 25$.

Test Ia: constant rates. We have reported in Figure 1 the evolution of the mean-field system with different simulations, when constant transition rates α_F, α_L are selected. According to Table 1, we selected $\alpha_F = 0.1, \alpha_L = 0.95$. Notice that in the case of constant rates the total masses $\sigma_t(F), \sigma_t(L)$ converge to

$$\sigma_\infty(F) = \frac{\alpha_L}{\alpha_F + \alpha_L}, \qquad \sigma_\infty(L) = \frac{\alpha_F}{\alpha_F + \alpha_L}.$$

Figure 1 depicts the density $\nu_t(x)$ in the time interval $\Omega \times [0, T]$, and the time evolution of $\sigma_t(F)$ and $\sigma_t(L)$. In Figure 2 we report different time frame of the densities μ_t^F, μ_t^L. We observe that at final time the system has clustered around three states.

Test Ib: Density-dependent rates. We consider birth rates depending on the densities μ_t^F, μ_t^L. We consider the variance measure of μ_t^L defined as follows

$$\mathcal{V}(\mu_t^L) = \frac{1}{|\sigma_t(L)|^2} \int_{\Omega\times\Omega} |x - y|^2 \, d\mu_t^L(x) \, d\mu_t^L(y), \tag{85}$$

which measures the spread of the solution μ_t^L over Ω. The birth rate of leaders α_F is selected as a switching function with respect to the dispersion measure (85), such that creation is activated only when the dispersion is above a certain threshold $\delta_F \geq 0$. Thus,

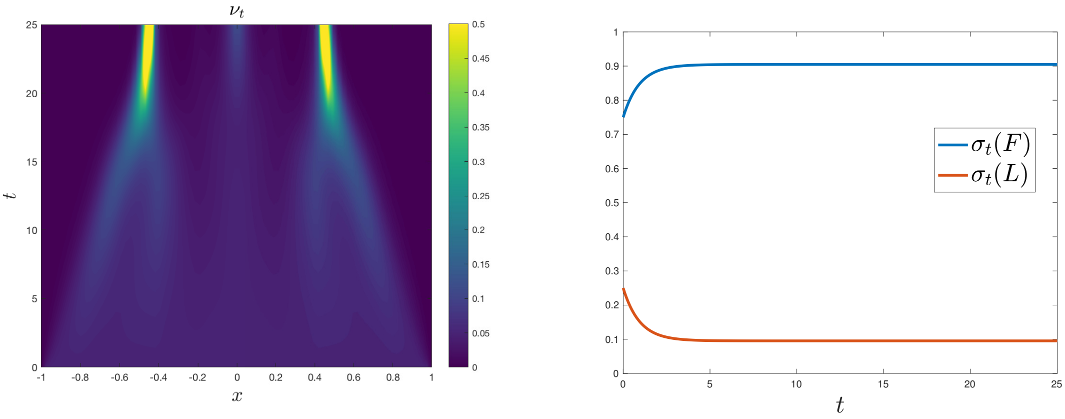

Figure 1: *Test Ia.* Left: the total density ν_t in the space time domain $[-1, 1] \times [0, T]$. Right: the followers' and leaders' mass, $\sigma_t(F), \sigma_t(L)$, with a monotonic evolution in time induced by the constant rates $\alpha_F = 0.15$ and $\alpha_L = 0.95$.

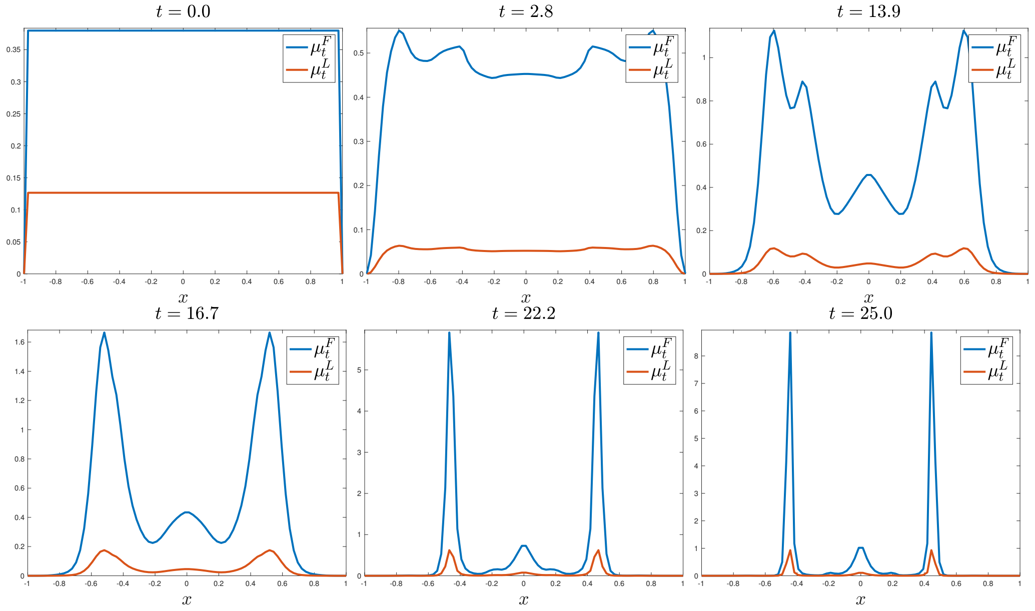

Figure 2: *Test Ia.* From left to right, and top to bottom row we show the emergence of consensus with leaders' interaction confidence level $C^L = 0.6$, and followers' interaction confidence level $C^F = 0.2$. Consensus state is not yet reached and three main clusters emerge.

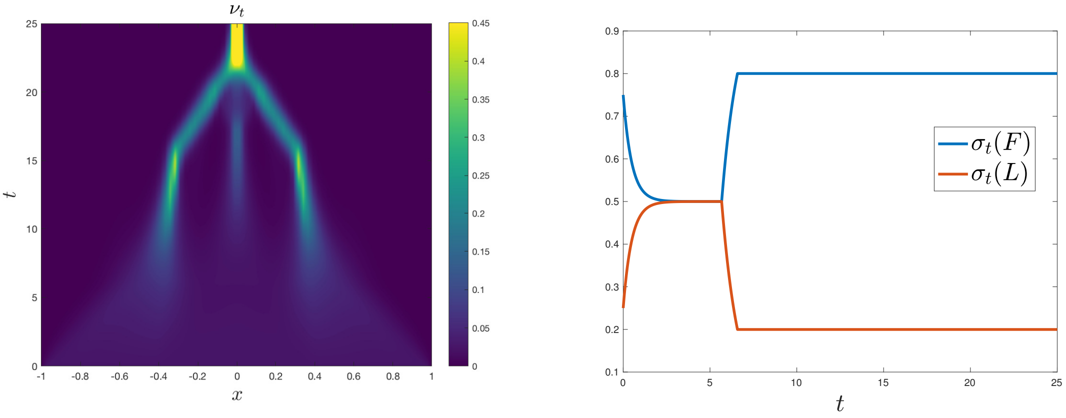

Figure 3: *Test Ib.* Left: the total density ν_t in the space time domain $[-1, 1] \times [0, T]$. Right: the non-linear evolution of the followers' and leaders' mass, $\sigma_t(F), \sigma_t(L)$.

we consider the following Lipschitz approximation of the indicator function

$$\alpha_F(\mu_t^F, \mu_t^L) = \frac{1}{1 + e^{c_F(\delta_F - \mathcal{V}(\mu_t^L))}} \tag{86}$$

with $c_F \gg 1$, here we select $c_F = 1000$, and $\delta_F = 0.15$.

Note that function (85) is exactly of the form (94), with $f(x) = |x|^2$. At the same time equation (86) complies with Assumption (H5), as shown in A, as long as $|\sigma_t(L)| \geq \epsilon$ for a fixed threshold $\epsilon > 0$. This last condition can be easily checked along the evolution.

The creation of followers given by rate α_L is instead determined by the following switching function

$$\alpha_L(\mu_t^F, \mu_t^L) = \frac{1}{1 + e^{c_L(\delta_L - |\sigma_t(L)|)}}, \tag{87}$$

namely when the total mass of leaders is above a threshold δ_L. Here we selected $\delta_L = 0.25$, and $c_L = 1000$.

Similarly to the previous test, we show in Figure 3 the total density $\nu_t(x)$ on $[0, T] \times \Omega$ and the time evolution of $\sigma_t(F)$ and $\sigma_t(L)$. In this case we observe the emergence of a consensus state before final time $T = 25$. This is explained by the large amount of leaders, whose mass increases until the total mass is too spread over the domain Ω (and so the measure $\mathcal{V}(\mu_t^L)$ is above the threshold δ_F). As soon as the threshold is reached, the creation of leaders is stopped and $\sigma_t(F), \sigma_t(L)$ converge to an asymptotic state thanks to the concentration of the total mass. In Figure 4 we show some frames of the time evolution of μ_t^F, μ_t^L.

5.2 Test II: Aggregation dynamics

We consider an aggregation dynamics ruled by an attraction towards the population of leaders, and repulsion towards the followers. Hence, we assume the following interaction

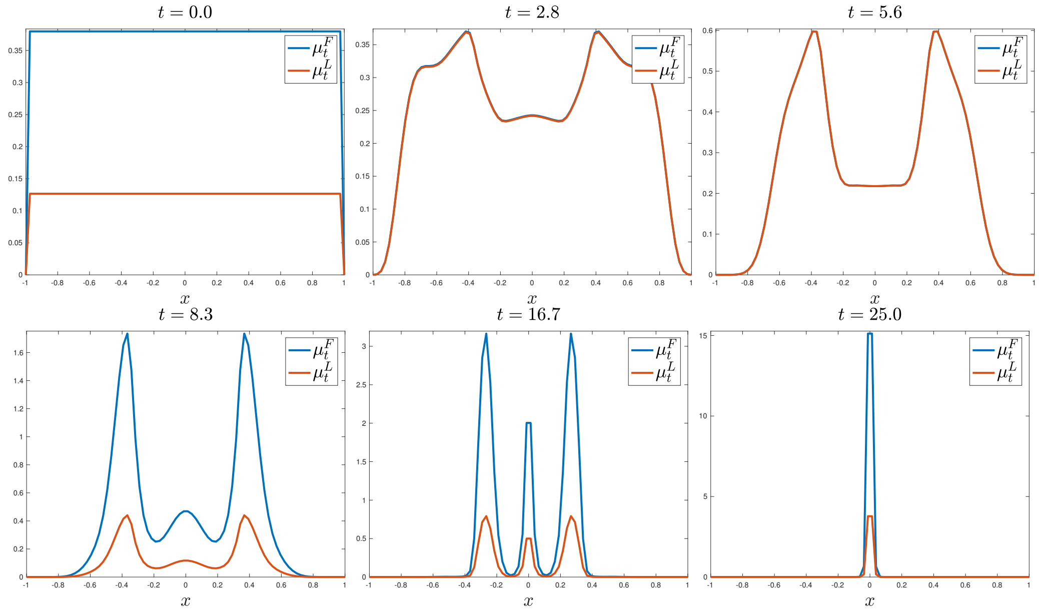

Figure 4: *Test Ib.* From left to right, and top to bottom row we show the emergence of consensus with leaders' interaction confidence level $C^L = 0.6$, and followers' interaction confidence level $C^F = 0.2$. Consensus in $x = 0$ at final time is reached.

kernels

$$K^F(x) = a^F(x)x, \qquad a^F(x) = -\frac{\ell_R}{(\epsilon + |x|)^{c_R}},$$
$$K^L(x) = a^L(x)x, \qquad a^L(x) = (\epsilon + |x|)^{c_A},$$

with non-negative parameters ℓ_R, c_R, c_A and $\epsilon = 0.001$.

The exchange of mass between leaders and followers is described as follows: we consider a constant rate α_L, whereas leaders' birth rate α_F depends non-linearly on the followers' density. Similarly to Test I we use the variance measure (85) for μ_t^F as follows

$$\mathcal{V}(\mu_t^F) = \frac{1}{|\sigma_t(F)|^2} \int_{\Omega\times\Omega} |x-y|^2 \, d\mu_t^F(x)\, d\mu_t^F(y). \tag{88}$$

The birth rate α_F is the switching function (86), modified as follows

$$\alpha_F(\mu_t^F, \mu_t^L) = \frac{1}{1 + e^{c_F(\delta_F - \mathcal{V}(\mu_t^F))}}, \tag{89}$$

with $c_F \gg 1$ and $\delta_F \geq 0$. Hence, we expect the total mass of leaders to increase when the followers' density is too spread over the domain Ω, and to decrease when followers' density is sufficiently concentrated.

Note that this choice controls the competition between the repulsive action of followers' kernel and the attraction of the leaders' one. In order to show the richness of

Table 2: Computational parameters for Test II.

Test	c_A	c_R	ℓ_R	α_F	δ_F	α_L	$\sigma_0(F)$	$\sigma_0(L)$
IIa	3	0.75	0.1	(89)	0.15	0.25	0.75	0.25
IIb	2	0.5	0.1	(89)	0.2	0.25	0.75	0.25

this setting we consider two different cases. The choice of the parameters are reported in Table 2.

For the numerical solution of the mean-field dynamics we fix the computational domain $\Omega = [-1, 1]$ with zero-flux boundary conditions, discretized with $N = 80$ space grid points, and time step $\Delta t = 0.0063$ and final time $T = 25$.

Test IIa: Uniform initial data. We consider an initial configuration where leaders and followers occupy the same domain's portion identified by the function

$$h(x) = \frac{1}{u-d}\chi_{[-\frac{u}{2},-\frac{d}{2}]\cup[\frac{d}{2},\frac{u}{2}]}(x)$$

with $d = 0.3$ and $u = 1.3$. The initial data of (1) is defined as follows

$$\mu_0^F(x) = \sigma_0(F)h(x), \quad \mu_0^L(x) = \sigma_0(L)h(x). \tag{90}$$

We report in Figure 5 the evolution of the system, observing an oscillating behavior of the total mass of leaders and followers towards a stable configuration of the densities' profiles. Indeed, initially the density of leaders increases to balance the spread of the initial mass (90), up to the moment when the birth rate α_F is switched off. Subsequently, the mass of followers starts to increase, together with the intensity of the repulsion force. Therefore, the dispersion measure (88) increases again, until the reactivation of the birth rate function α_F. At final time $T = 25$, the system has reached a stationary configuration of the densities μ_t^L, μ_t^F as well as of the total masses $\sigma_t(L), \sigma_t(F)$.

Test IIb: Confinement. We consider a confinement setting, where the leaders' density surrounds the initial density of followers. In this particular situation, differently from the previous cases, Assumption (H1) on the initial data is not plausible anymore, therefore we renounce to it. We however recall the reader that an existence and uniqueness theory for system (1) is still available, since Propositions 3.2 and 3.3 do not require (H1) to be fulfilled.

We introduce the Gaussian function

$$G(x;\varsigma^2) = \frac{1}{\sqrt{2\pi\varsigma^2}}e^{-\frac{x^2}{\varsigma^2}},$$

then we define the initial data as follows

$$\mu_0^F(x) = \sigma_0(F)G(x;1/30), \qquad \mu_0^L(x) = \frac{\sigma_0(L)}{2}\left(G(x-0.6;1/90) + G(x+0.6;1/90)\right).$$

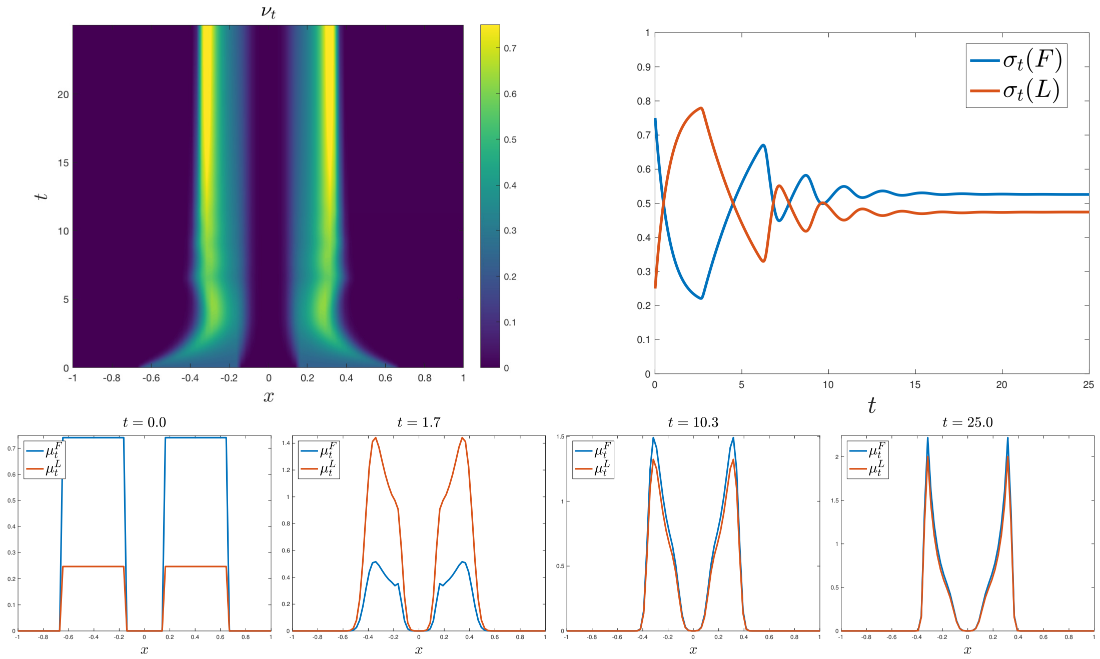

Figure 5: *Test IIa.* Top line: the left-hand picture shows the total density ν_t, the right-hand picture shows the evolution of the masses σ_t^F, σ_t^L. Bottom line: From left to right we depict the evolution of the leaders' and followers' densities from the initial data to the final stationary state.

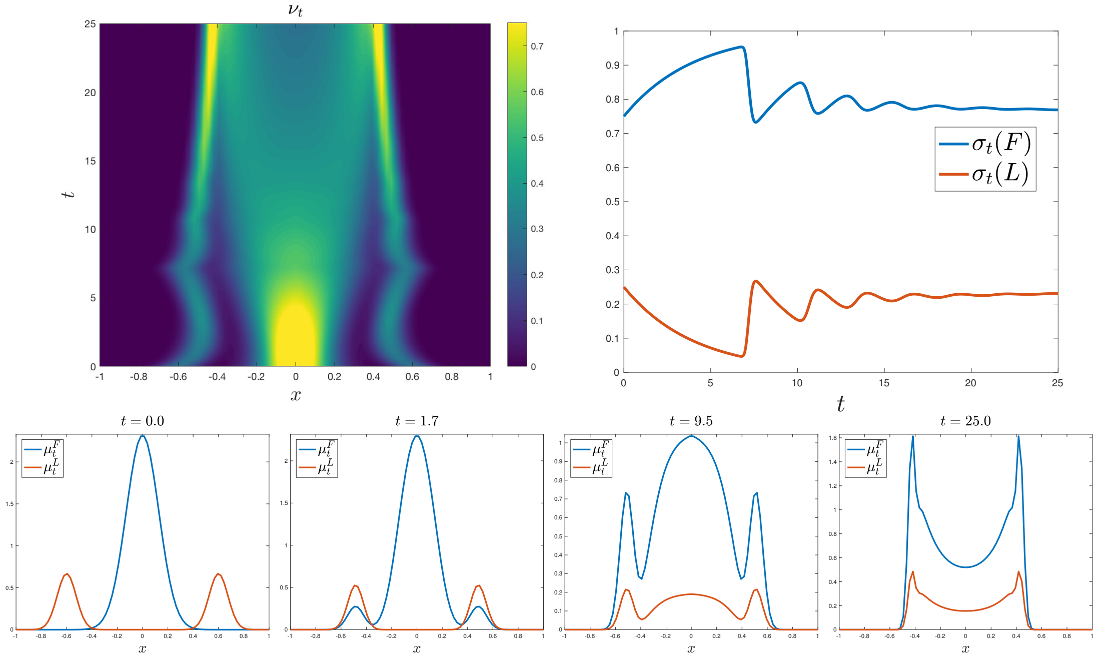

Figure 6: *Test IIb.* Top line: the left-hand picture shows the total density ν_t, the right-hand picture shows the evolution of the masses σ_t^F, σ_t^L. Bottom line: From left to right we depict the evolution of the leaders' and followers' densities from the initial data to the final state.

In this setting, the initial dispersion of followers is not large enough to activate the birth rate α_F, (89). Indeed we can observe from the first two frames of Figure 6-bottom row that the density of followers starts to grow on the support of ν_t, while the creation of leaders is not inhibited. In a second step, when the interaction becomes too repulsive, the spread of μ_t^F activates the creation of leaders, and eventually stabilizes the total density towards a stable configuration, with the masses $\sigma_t(L), \sigma_t(F)$ converging towards a stationary value.

5.3 Test III: Leaders with steering action

We study a population of leaders aiming to reach a desired position $\hat{x} \in \Omega$, and how their motion influences the followers' density. The followers' dynamics is governed by an aggregation equation of the type

$$K^F(x) = a^F(x)x, \qquad a^F(x) = (\epsilon + |x|)^{c_A} - \frac{\ell_R}{(\epsilon + |x|)^{c_R}}.$$

Leaders have a steering drift towards $\hat{x}$ of the form

$$|\mu^L| K^L(x) = \sigma(L)\,(\hat{x} - x)$$

Table 3: Computational parameters for Test III.

Test	c_A	c_R	ℓ_R	ϵ	α_F	δ_F	α_L	$\sigma_0(F)$	$\sigma_0(L)$	$\hat{x}$
III	2	1	0.05	0.0001	(91)	0.15	0.25	0.75	0.25	0.5

which also complies with our abstract framework, as discussed in Remark 3.6.

We choose a constant rate for the death of leaders $\alpha_L = 0.25$, and the following state-dependent rate for the birth of leaders

$$\alpha_F(\mu_t^F, \mu_t^L) = \frac{1}{1 + e^{c_F(\delta_F - \mathcal{D}(\mu_t^F))}}, \tag{91}$$

with $c_F = 1000$, and where $\mathcal{D}(\mu_t^F)$ is the variance of followers' density with respect to the desired configuration $\hat{x}$,

$$\mathcal{D}(\mu_t^F) = \frac{1}{|\sigma_t(F)|} \int_\Omega |\hat{x} - x|^2 d\mu_t^F(x).$$

Hence, we expect the leaders' density to increase when followers are not concentrated around $\hat{x}$, and to vanish as soon as the desired state is approached.

This test case is inspired by applications in pedestrian dynamics, where a part of the total mass of agents (leaders) is used as control variable to improve the evacuation time of a crowd [2, 16]. We remain in a simplified setting: similarly to the previous tests, we solve numerically the evolution of the mean-field interaction dynamics in the one-dimensional domain $\Omega = [-1, 1]$ with zero-flux boundary conditions. For the numerical discretization we select $N = 80$ space grid points, time step $\Delta t = 0.0127$ and final time $T = 15$. We have reported in Table 2 the parameters' choice for the different cases.

Figure 7 shows the evolution of the density $\nu_t(x)$ and the evolution of the followers' and leaders' masses σ_t^F, σ_t^L in the top row. Bottom row shows the evolutions of μ_t^F, μ_t^L: the mass of leaders increases initially since followers are far away from $\hat{x} = 0.5$, as soon as μ_t^F approaches $\hat{x}$, while the density of leaders tends to vanish.

Acknowledgments

This work has been supported by the INdAM-GNCS 2018 project *Numerical methods for multiscale control problems*, and the project *"MIUR Departments of Excellence 2018-2022"*.

A Examples of transition functionals

We prove a simple sufficient condition for α_F and α_L to satisfy (H5).

Proposition A.1. *For $i = 1, \dots, 5$, let $f_i : \mathbb{R}^d \to \mathbb{R}^{m_1}$ be given locally Lipschitz continuous functions and consider a locally Lipschitz function $\alpha : \mathbb{R}^{m_1+m_2+m_3+m_4+m_5} \to \mathbb{R}$.*

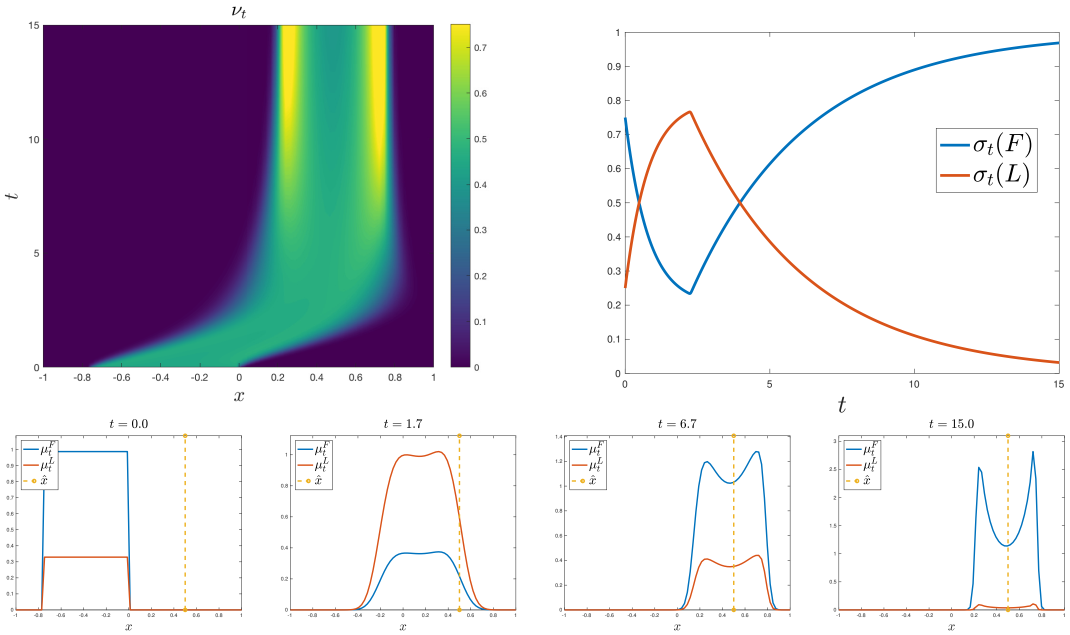

Figure 7: *Test III*. Top line: the left-hand picture shows the total density ν_t, the right-hand picture shows the evolution of the masses σ_t^F, σ_t^L. Bottom line: From left to right we depict the evolution of the leaders' and followers' densities towards the desired position $\hat{x} = 0.5$.

Then the map

$$\alpha(\mu,\eta) = \alpha\left(\int_{\mathbb{R}^d}(f_1 * \mu)\,\mathrm{d}\mu, \int_{\mathbb{R}^d}(f_2 * \eta)\,\mathrm{d}\mu, \int_{\mathbb{R}^d}(f_3 * \eta)\,\mathrm{d}\eta, \int_{\mathbb{R}^d} f_4\,\mathrm{d}\mu, \int_{\mathbb{R}^d} f_5\,\mathrm{d}\eta\right)$$

satisfies Assumption (H5).

Proof. By possibly arguing componentwise on the f_i's we can only consider the case $m_1 = m_2 = m_3 = m_4 = m_5 = 1$. For all μ, η satisfying $\mu(\mathbb{R}^d)$, $\eta(\mathbb{R}^d) \leq M$ and with support contained in $B(0,R)$, we clearly have

$$\left|\int_{\mathbb{R}^d}(f_1 * \mu)\,\mathrm{d}\mu\right| \leq M^2\mathrm{max}_{B(0,2R)}|f_1|, \quad \left|\int_{\mathbb{R}^d}(f_2 * \mu)\,\mathrm{d}\eta\right| \leq M^2\mathrm{max}_{B(0,2R)}|f_2|,$$
$$\left|\int_{\mathbb{R}^d}(f_3 * \eta)\,\mathrm{d}\eta\right| \leq M^2\mathrm{max}_{B(0,2R)}|f_3|, \quad \left|\int_{\mathbb{R}^d} f_4\,\mathrm{d}\mu\right| \leq M\mathrm{max}_{B(0,R)}|f_4|$$
$$\left|\int_{\mathbb{R}^d} f_5\,\mathrm{d}\eta\right| \leq M\mathrm{max}_{B(0,R)}|f_5|\,.$$

With this hypothesis, since the function α is Lipschitz, it only suffices to show that the functions

$$(\mu,\eta) \mapsto \int_{\mathbb{R}^d}(f_1 * \mu)\,\mathrm{d}\mu, \quad (\mu,\eta) \mapsto \int_{\mathbb{R}^d}(f_2 * \mu)\,\mathrm{d}\eta$$
$$(\mu,\eta) \mapsto \int_{\mathbb{R}^d}(f_3 * \eta)\,\mathrm{d}\eta, \quad (\mu,\eta) \mapsto \int_{\mathbb{R}^d} f_4\,\mathrm{d}\mu, \qquad (\mu,\eta) \mapsto \int_{\mathbb{R}^d} f_5\,\mathrm{d}\eta$$

satisfy (H5). We only discuss the second case, since the proof in the other cases is similar.

Denote with $\tilde{f}_2$ the function defined by $\tilde{f}_2(x) = f_2(-x)$. Whenever μ has support contained in $B(0,R)$ and satisfies $\mu(\mathbb{R}^d) \leq M$ we clearly have

$$\sup_{x\in B(0,R)} |f_2 * \mu|(x) \leq M \sup_{x\in B(0,2R)} |f_2(x)|, \quad \sup_{x\in B(0,R)} |\tilde{f}_2 * \mu|(x) \leq M \sup_{x\in B(0,2R)} |f_2(x)|\,. \tag{92}$$

A direct computation also shows that, if we denote with Lip_R the Lipschitz constant on a ball of radius R, it holds

$$\mathrm{Lip}_R(f_2 * \mu) \leq M\mathrm{Lip}_{2R}(f_2), \quad \mathrm{Lip}_R(\tilde{f}_2 * \mu) \leq M\mathrm{Lip}_{2R}(f_2)\,. \tag{93}$$

Take now (μ_1,η_1) and (μ_2,η_2) positive measures satisfying (23) and (24). Use (92) and (93), toghether with the Kantorovich-Rubinstein duality, we have

$$\left|\int_{\mathbb{R}^d}(f_2 * \mu_1)\,\mathrm{d}\eta_1 - \int_{\mathbb{R}^d}(f_2 * \mu_2)\,\mathrm{d}\eta_2\right| = \left|\int_{B(0,R)}(f_2 * \mu_1)\,\mathrm{d}\eta_1 - \int_{B(0,R)}(f_2 * \mu_2)\,\mathrm{d}\eta_2\right|$$
$$= \left|\int_{B(0,R)}(f_2 * \mu_1)\,\mathrm{d}(\eta_1 - \eta_2) + \int_{B(0,R)}(\tilde{f}_2 * \eta_2)\,\mathrm{d}(\mu_1 - \mu_2)\right|$$
$$\leq C_{M,R}(\mathcal{W}_g\,(\mu_1,\mu_2) + \mathcal{W}_g\,(\eta_1,\eta_2)),$$

with $C_{M,R} = M(\sup_{x\in B(0,2R)}|f_2(x)| + \mathrm{Lip}_{2R}(f_2))$. This concludes the proof. □

Example A.2. The statement above is clearly still valid if α only depends on some of the variables indicated above. In some applications (as for instance in [27]) the transition rate α_i behaves countercyclically with respect to the mass of μ_t^i: whenever $|\mu_t^i|$ is below a certain threshold $\varepsilon > 0$, the function α_i increases in order to restore $|\mu_t^i|$ to higher levels. To model this phenomenon, let χ_ε be a mollification of the function

$$\overline{\chi}_\varepsilon(x) = \begin{cases} 1 & \text{if } x \leq \varepsilon \\ \overline{c} & \text{otherwise,} \end{cases}$$

with $0 \leq \overline{c} < 1$. Then, by Proposition A.1 (with $f_i = 0$ for $i = 1, \ldots, 4$ and $f_5 \equiv 1$) the function $\alpha(\mu, \eta) := \chi_\varepsilon(|\eta|)$ satisfies Assumption (H5).

Also quotients of functions of the type considered in Proposition A.1 are easily seen to comply with Assumption (H5), provided that the denominator is bounded away from zero. For instance, for a given scalar-valued $f : \mathbb{R}^d \to \mathbb{R}$ and $g(\lambda) = ((1-\lambda) \wedge \epsilon)^2$, where $\epsilon > 0$ is a fixed threshold, one can consider a function of the type

$$\alpha(\mu, \eta) := \alpha\left(\frac{\int_{\mathbb{R}^d} (f * \mu)\, \mathrm{d}\mu}{g(|\eta|)}\right). \tag{94}$$

If $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, $\sigma \in \mathcal{P}_1(\{F, L\})$ and we set $\mu := \sigma(L)\nu, \eta := \sigma(F)\nu$, then the above function reduces to

$$\alpha(\nu, \sigma) = \alpha\left(\frac{\sigma(L)^2 \int_{\mathbb{R}^d} (f * \nu)\, \mathrm{d}\nu}{(\sigma(L) \wedge \epsilon)^2}\right)$$

which, as long as $\sigma(L) \geq \epsilon$, coincides with $\alpha(\int_{\mathbb{R}^d} (f * \nu)\, \mathrm{d}\nu)$ and only takes into account the total distribution ν of the two populations.

B Finite-volume scheme for mean-field leader-follower dynamics

We introduce a finite-volume scheme for the discretization of the mean-field system (1) in one-space dimension. Hence we consider a constant discretization step $\Delta x > 0$, and we define $x_\ell = \ell \Delta x$ with $\ell \in \mathbb{Z}$, and the cells $C_\ell = [x_{\ell-1/2}, x_{\ell+1/2}]$, with $x_\ell \pm 1/2 = x_\ell \pm \Delta x/2$, over which we define the averages

$$\mu_\ell^i(t) = \frac{1}{\Delta x} \int_{x_{\ell-1/2}}^{x_{\ell+1/2}} \mu^i(x,t) dx, \qquad i \in \{F, L\},$$

where we used the notation $\mu^i(x,t)$ for the measure $\mu_t^i(x)$. In the same spirt we define the numerical fluxes as follows

$$\mathcal{F}_{\ell+1/2}^i = \mathcal{K}_{\ell+1/2}[\mu^F, \mu^L] \mu_{\ell+1/2}^i, \qquad i \in \{F, L\}.$$

In what follows we will consider an upwinding scheme, where the convolutional operator $\mathcal{K}(\mu^F, \mu^L) := (K^F * \mu^F + K^L * \mu^L)(x_{\ell+1/2})$ is evaluated at the interfaces according to quadrature formula, and the densities $\mu^i_{\ell+1/2}$ are defined as follows

$$\mu^i_{\ell+1/2} = \begin{cases} \mu^i_{\ell+1} & \text{if } \mathcal{K}_{\ell+1/2} < 0, \\ \mu^i_{\ell} & \text{otherwise.} \end{cases}$$

The sources terms are computed by averaging the transition rates α_F, α_L, as follows

$$\mathcal{A}^i_\ell(t) = \frac{1}{\Delta x}\int_{x_{\ell-1/2}}^{x_{\ell+1/2}} \alpha_i(\mu^F, \mu^L, t)\mu^i(x)dx, \qquad i \in \{F, L\}.$$

We employ a first-order time marching scheme to compute the solution $\mu^i_\ell(t)$ over the time grid $0 = t_0, \dots, t_{N_t} = T$, with fixed time step $\Delta t = t_{n+1} - t_n$. Moreover we used a splitting technique to evaluate separately the contribution by the non-linear transport and the source terms. Thus the full discrete scheme reads

$$\begin{cases} \mu^{F,\star}_\ell = \mu^{F,n}_\ell - \frac{\Delta t}{\Delta x}\left(\mathcal{F}^F_{\ell+1/2} - \mathcal{F}^F_{\ell-1/2}\right), \\ \mu^{L,\star}_\ell = \mu^{L,n}_\ell - \frac{\Delta t}{\Delta x}\left(\mathcal{F}^L_{\ell+1/2} - \mathcal{F}^L_{\ell-1/2}\right), \end{cases}$$
$$\begin{cases} \mu^{F,n+1}_\ell = \mu^{F,\star}_\ell - \Delta t\left(\mathcal{A}^{F,\star}_\ell - \mathcal{A}^{L,\star}_\ell\right), \\ \mu^{L,n+1}_\ell = \mu^{L,\star}_\ell + \Delta t\left(\mathcal{A}^{F,\star}_\ell - \mathcal{A}^{L,\star}_\ell\right). \end{cases}$$

References

[1] S. Ahn, H.-O. Bae, S.-Y. Ha, Y. Kim, and H. Lim. Application of flocking mechanism to the modeling of stochastic volatility. *Math. Models Methods Appl. Sci.*, 23(9):1603–1628, 2013.

[2] G. Albi, M. Bongini, E. Cristiani, and D. Kalise. Invisible control of self-organizing agents leaving unknown environments. *SIAM J. Appl. Math.*, 76(4):1683–1710, 2016.

[3] G. Albi, Y.-P. Choi, M. Fornasier, and D. Kalise. Mean field control hierarchy. *Applied Mathematics & Optimization*, 76(1):93–135, 2017.

[4] G. Albi, L. Pareschi, and M. Zanella. Boltzmann-type control of opinion consensus through leaders. *Phil. Trans. R. Soc. A*, 372(2028):20140138, 2014.

[5] G. Albi, L. Pareschi, and M. Zanella. Opinion dynamics over complex networks: Kinetic modelling and numerical methods. *Kinetic & Related Models*, 10(1):1–32, 2017.

[6] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.*, 61(1):18–53, 2008.

[7] 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, 2008.

[8] P. Bak. *How nature works: the science of self-organized criticality.* Springer Science & Business Media, 2013.

[9] M. Ballerini, N. Cabibbo, R. Candelier, et al. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *P. Natl. Acad. Sci. USA*, 105(4):1232–1237, 2008.

[10] R. Bellman. *Dynamic programming.* Princeton University Press, 1957.

[11] N. Bellomo and C. Dogbe. On the modeling of traffic and crowds: a survey of models, speculations, and perspectives. *SIAM Rev.*, 53(3):409–463, 2011.

[12] M. Bongini and G. Buttazzo. Optimal control problems in transport dynamics. *Math. Models Methods Appl. Sci.*, 27(3):427–451, 2017.

[13] M. Bongini and M. Fornasier. Sparse control of multiagent systems. In *Active Particles, Volume 1*, pages 259–298. Springer, 2017.

[14] M. Bongini, M. Fornasier, M. Hansen, and M. Maggioni. Inferring interaction rules from observations of evolutive systems I: The variational approach. *Math. Models. Meth. Appl. Sci.*, 27(05):909–951, 2017.

[15] M. Bongini, M. Fornasier, F. Rossi, and F. Solombrino. Mean-field Pontryagin maximum principle. *J. Optim. Theory Appl.*, 175(1):1–38, 2017.

[16] M. Burger, M. Di Francesco, P. A. Markowich, and M.-T. Wolfram. Mean field games with nonlinear mobilities in pedestrian dynamics. *Discrete & Continuous Dynamical Systems-B*, 19(5):1311–1333, 2014.

[17] J. A. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: mean-field limit and wasserstein distances. In *Collective dynamics from bacteria to crowds*, pages 1–46. Springer, 2014.

[18] J. A. Carrillo, Y.-P. Choi, and S. P. Perez. A review on attractive–repulsive hydrodynamics for consensus in collective behavior. In *Active Particles, Volume 1*, pages 173–228. Springer, 2017.

[19] J. A. Carrillo, S. Fagioli, F. Santambrogio, and M. Schmidtchen. Splitting schemes and segregation in reaction cross-diffusion systems. *SIAM Journal on Mathematical Analysis*, 50(5):5695–5718, 2018.

[20] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.

[21] M. Cirant. Multi-population mean field games systems with Neumann boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 103(5):1294–1315, 2015.

[22] S. Cordier, L. Pareschi, and G. Toscani. On a kinetic model for a simple market economy. *J. Stat. Phys.*, 120(1-2):253–277, 2005.

[23] E. Cristiani, B. Piccoli, and A. Tosin. *Multiscale modeling of pedestrian dynamics*, volume 12. Springer, 2014.

[24] R. Dobrushin. Vlasov equations. *Funct. Anal. Appl.*, 13(2):115–123, 1979.

[25] M. Dorigo and C. Blum. Ant colony optimization theory: A survey. *Theoretical computer science*, 344(2-3):243–278, 2005.

[26] R. M. Dudley. *Real Analysis and Probability.* Chapman and Hall/CRC, 1989.

[27] 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.

[28] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic Publishers, 1988.

[29] M. Fornasier, B. Piccoli, and F. Rossi. Mean-field sparse optimal control. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130400, 2014.

[30] M. Fornasier and F. Solombrino. Mean-field optimal control. *ESAIM Control Optim. Calc. Var.*, 20(4):1123–1152, 2014.

[31] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.

[32] A. Garroni, P. van Meurs, M. A. Peletier, and L. Scardia. Convergence and non-convergence of many-particle evolutions with multiple signs. *arXiv preprint arXiv:1810.04934*, 2018.

[33] F. Golse. The mean-field limit for the dynamics of large particle systems. In *Journées "Équations aux dérivées partielles", Forges-les-Eaux, France, 2 au 6 juin 2003. Exposés Nos. I-XV*, pages 1–47. Nantes: Université de Nantes, 2003.

[34] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simulat.*, 5(3), 2002.

[35] H. Huang, J.-G. Liu, and J. Lu. Learning interacting particle systems: Diffusion parameter estimation for aggregation equations. *Mathematical Models and Methods in Applied Sciences*, to appear, 2018.

[36] J. Kennedy. Particle swarm optimization. In *Encyclopedia of machine learning*, pages 760–766. Springer, 2011.

[37] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11-12):1117–1164, 2016.

[38] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[39] M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves. *SIAM J. Math. Anal.*, 48(4):2869–2911, 2016.

[40] H. P. McKean. Propagation of chaos for a class of non-linear parabolic equations. *Lecture Series in Differential Equations, Catholic University, Washington D.C.*, 7:41–57, 1967.

[41] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[42] B. Piccoli and F. Rossi. Transport equation with nonlocal velocity in Wasserstein spaces: Convergence of numerical schemes. *Acta Appl. Math.*, 124(1):73–105, 2013.

[43] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.*, 211(1):335–358, 2014.

[44] B. Piccoli and F. Rossi. On properties of the generalized Wasserstein distance. *Arch. Ration. Mech. Anal.*, 222(3):1339–1365, 2016.

[45] B. Piccoli and F. Rossi. Measure-theoretic models for crowd dynamics. In *Crowd Dynamics Volume 1 - Theory, Models, and Safety Problems.* N. Bellomo and L. Gibelli Eds, Birkhauser, to appear.

[46] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[47] M.-N. Thai. Birth and death process in mean field type interaction. *Bernoulli*, to appear, 2018.

[48] T. Vicsek and A. Zafeiris. Collective motion. *Phys. Rep.*, 517(3):71–140, 2012.

[49] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.