Interaction Kernel Learning in Multi-Agent Dynamical System

M. Bongini, M. Fornasier, M. Hansen, M. Maggioni

1 Introduction

The continuity equation in physics describes the transport of a conserved quantity. Several quantities, such as mass, energy, momentum, electric charge and opinion (in social dynamics) are conserved, hence a variety of physical and social phenomena can be described using the continuity equation, which can be written in differential form as

$$\frac{\partial \mu}{\partial t} = -\nabla \cdot (H\mu).$$

The function H is the flow velocity vector field and determines the dynamics of the conserved quantity μ.

Due to its ubiquity, it is very easy to find quantities whose evolution can be described by the continuity equation. Once such a quantity has been identified, a crucial problem is then to determine the correct form of the velocity field H, which highly depends on the particular quantity of interest μ. As observers of the dynamics of μ (with which we are not allowed to interact), the exact determination of H has to be done with the only information at our disposal, that is samples of the trajectory of μ.

We are interested in addressing this problem in the context of *convolution-type dynamics*, in which the velocity field has the specific form

$$H = F[a] * \mu,$$

where $F[a](\xi) = a(|\xi|)\xi$ for any $\xi \in \mathbb{R}^d$ and $a : \mathbb{R}_+ \to \mathbb{R}$. The function a is called *interaction kernel* and this form of the continuity equation is often encountered in biology, chemistry and social sciences (see [4, 10, 11, 12]). The interaction kernel a is then the function to be learned from the samples of the trajectories of μ...

2 Preliminaries

The space $\mathcal{P}(\mathbb{R}^n)$ is the set of probability measures which take values on $\mathbb{R}^n$, while the space[1] $\mathcal{P}_p(\mathbb{R}^n)$ is the subset of $\mathcal{P}(\mathbb{R}^n)$ whose elements have finite p-th moment, i.e.,

$$\int_{\mathbb{R}^n} |x|^p d\mu(x) < +\infty.$$

We denote by $\mathcal{P}_c(\mathbb{R}^n)$ the subset of $\mathcal{P}_1(\mathbb{R}^n)$ which consists of all probability measures with compact support.

For any $\mu \in \mathcal{P}(\mathbb{R}^n)$ and any Borel function $f : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, we denote by $f_{\#}\mu \in \mathcal{P}(\mathbb{R}^{n_2})$ the *push-forward of* μ *through* f, defined by

$$f_{\#}\mu(B) := \mu(f^{-1}(B)) \quad \text{for every Borel set } B \text{ of } \mathbb{R}^{n_2}.$$

In particular, if one considers the projection operators π_1 and π_2 defined on the product space $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, for every $\rho \in \mathcal{P}(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ we call *first* (resp., *second*) *marginal* of ρ the probability measure $\pi_{1\#}\rho$ (resp., $\pi_{2\#}\rho$). Given $\mu \in \mathcal{P}(\mathbb{R}^{n_1})$ and $\nu \in \mathcal{P}(\mathbb{R}^{n_2})$, we denote with $\Gamma(\mu, \nu)$ the subset of all probability measures in $\mathcal{P}(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ with first marginal μ and second marginal ν.

On the set $\mathcal{P}_p(\mathbb{R}^n)$ we shall consider the following distance, called the *Wasserstein or Monge-Kantorovich-Rubinstein distance*,

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

If $p = 1$, we have the following equivalent expression for the Wasserstein distance:

$$\mathcal{W}_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^n} \varphi(x) d(\mu - \nu)(x) : \varphi \in \mathrm{Lip}(\mathbb{R}^n),\ \mathrm{Lip}_{\mathbb{R}^n}(\varphi) \le 1 \right\},$$

where $\mathrm{Lip}_{\mathbb{R}^n}(\varphi)$ stands for the Lipschitz constant of φ on $\mathbb{R}^n$. We denote by $\Gamma_o(\mu, \nu)$ the set of optimal plans for which the minimum is attained, i.e.,

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

It is well-known that $\Gamma_o(\mu, \nu)$ is non-empty for every $(\mu, \nu) \in \mathcal{P}_p(\mathbb{R}^n) \times \mathcal{P}_p(\mathbb{R}^n)$, hence the infimum in (1) is actually a minimum. For more details, see e.g. [2, 13].

For any $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ and $f : \mathbb{R}^d \to \mathbb{R}^d$, the notation $f * \mu$ stands for the convolution of f and μ, i.e.,

$$(f * \mu)(x) = \int_{\mathbb{R}^d} f(x - x') d\mu(x');$$

this quantity is well-defined whenever f is continuous and *sublinear*, i.e., there exists a constant $C > 0$ such that $|f(\xi)| \le C(1 + |\xi|)$ for all $\xi \in \mathbb{R}^d$.

[1] We follow the notation of [2].

2.1 The mean-field limit and an existence result

As already stated in the introduction, we are interested in the following *finite time horizon initial value problem*: given $T > 0$ and $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$, consider a curve $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ satisfying

$$\begin{cases} \dfrac{\partial \mu}{\partial t}(t) = -\nabla \cdot ((F[a] * \mu(t))\mu(t)) & \text{for } t \in (0,T], \\ \mu(0) = \mu_0. \end{cases} \tag{2}$$

We consequently give our notion of solution for (2).

Definition 2.1. We say that a map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ is a solution of (2) with initial datum μ_0 if the following hold:

1. μ has uniformly compact support, i.e., there exists $R > 0$ such that $\operatorname{supp}(\mu(t)) \subset B(0,R)$ for every $t \in [0,T]$;

2. μ is continuous with respect to the Wasserstein distance $\mathcal{W}_1$;

3. μ satisfies (2) in the weak sense, i.e., for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R})$ it holds
$$\frac{d}{dt}\int_{\mathbb{R}^d} \phi(x) d\mu(t)(x) = \int_{\mathbb{R}^d} \nabla\phi(x) \cdot (F[a] * \mu(t))(x) d\mu(t)(x).$$

It is well known that system (2) is closely related to the family of ODEs indexed by $N \in \mathbb{N}$ given by

$$\begin{cases} \dot{x}_i^N(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^N F[a](x_i^N(t) - x_j^N(t)) & \text{for } t \in (0,T], \\ x_i^N(0) = x_{0,i}^N, \end{cases} \quad i = 1,\dots,N, \tag{3}$$

which can be more conveniently rewritten as follows

$$\begin{cases} \dot{x}_i^N(t) = (F[a] * \mu^N(t))(x_i^N(t)) & \text{for } t \in (0,T], \\ x_i^N(0) = x_{0,i}^N, \end{cases} \quad i = 1,\dots,N, \tag{4}$$

by means of the *empirical measure* $\mu^N : [0,T] \to \mathcal{P}_c(\mathbb{R}^d)$ defined as

$$\mu^N(t)(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - x_i^N(t)), \quad \text{for every } x \in \mathbb{R}^d. \tag{5}$$

The following preliminary result tells us that solutions to system (3) are also solutions to systems (2), whenever conveniently rewritten.

Proposition 2.2. *Let $N \in \mathbb{N}$ be given. Let $(x_1^N, \dots, x_N^N) : [0,T] \to \mathbb{R}^{dN}$ be the solution of (3) with initial datum $x_0^N \in \mathbb{R}^{dN}$. Then the empirical measure $\mu^N : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ defined as in (5) is a solution of (2) with initial datum $\mu_0^N := \mu^N(0) \in \mathcal{P}_c(\mathbb{R}^d)$.*

Proof. It can be easily proved by arguing exactly as in [7, Lemma 4.3]. $\square$

The well-posedness of system (2) and several crucial properties that it enjoys can be proved as soon as we restrict our attention to interaction kernels belonging to the following *set of admissible kernels*

$$X = \left\{ b : \mathbb{R}_+ \to \mathbb{R} \mid b \in L^\infty(\mathbb{R}_+) \cap W^{1,\infty}_{\mathrm{loc}}(\mathbb{R}_+) \right\}.$$

Notice that, if $a \in X$ then it is weakly differentiable and for every compact set $K \subset \mathbb{R}_+$ its local Lipschitz constant $\mathrm{Lip}_K(a)$ is finite. As a consequence, we are able to state several basic estimates that shall be useful towards an existence and uniqueness result for the solutions of system (3).

Lemma 2.3. *Let $a \in X$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$. Then for all $y \in \mathbb{R}^d$ the following hold:*

$$|(F[a] * \mu)(y)| \le \|a\|_{L^\infty(\mathbb{R}_+)} \left(|y| + \int_{\mathbb{R}^d} |x| d\mu(x) \right).$$

Proof. Trivially follows from $a \in L^\infty(\mathbb{R}_+)$. $\square$

Lemma 2.4. *If $a \in X$ then $F[a] \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$.*

Proof. For any compact set $K \subset \mathbb{R}^d$ and for every $x, y \in K$ it holds

$$\begin{aligned}
|F[a](x) - F[a](y)| &= |a(|x|)x - a(|y|)y| \\
&\le |a(|x|)||x-y| + |a(|x|) - a(|y|)||y| \\
&\le (|a(|x|)| + \mathrm{Lip}_K(a)|y|)|x-y|,
\end{aligned}$$

and since $a \in L^\infty(\mathbb{R}_+)$ and $y \in K$, it follows that $F[a]$ is locally Lipschitz with Lipschitz constant depending only on a and K. $\square$

Lemma 2.5. *If $a \in X$ and $\mu \in \mathcal{P}_c(\mathbb{R}^d)$ then $F[a] * \mu \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$.*

Proof. For any compact set $K \subset \mathbb{R}^d$ and for every $x, y \in K$ it holds

$$\begin{aligned}
|(F[a] * \mu)(x) - (F[a] * \mu)(y)| &= \left| \int_{\mathbb{R}^d} a(|x-z|)(x-z) d\mu(z) - \int_{\mathbb{R}^d} a(|y-z|)(y-z) d\mu(z) \right| \\
&\le \int_{\mathbb{R}^d} |a(|x-z|) - a(|y-z|)||x-z| d\mu(z) \\
&\quad + \int_{\mathbb{R}^d} a(|y-z|)|x-y| d\mu(z) \\
&\le \mathrm{Lip}_{\widehat{K}}(a)|x-y| \int_{\mathbb{R}^d} |x-z| d\mu(z) + \|a\|_{L^\infty(\mathbb{R}_+)}|x-y| \\
&\le \left(\mathrm{Lip}_{\widehat{K}}(a)(|x|+1) + \|a\|_{L^\infty(\mathbb{R}_+)} \right) |x-y| \\
&\le \left(C\mathrm{Lip}_{\widehat{K}}(a) + \|a\|_{L^\infty(\mathbb{R}_+)} \right) |x-y|,
\end{aligned}$$

where C is a constant depending on K, and $\widehat{K}$ is a compact set containing both K and $\mathrm{supp}(\mu)$. $\square$

Proposition 2.6. *If $a \in X$ then system* (3) *admits a unique global solution in* $[0, T]$ *for every initial datum* $x_0^N \in \mathbb{R}^{dN}$.

Proof. Rewriting system (3) in the form of (4), from Lemma 2.5 follows trivially that the function $G : \mathbb{R}^{dN} \to \mathbb{R}^{dN}$ defined for every $(x_1, \ldots, x_N) \in \mathbb{R}^{dN}$ as

$$G(x_1, \ldots, x_N) = ((F[a] * \mu^N)(x_1), \ldots, (F[a] * \mu^N)(x_N)),$$

where μ^N is the empirical measure given by (5), satisfies $G \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^{dN})$. The Cauchy-Lipschitz Theorem for ODE systems then yields the desired result. □

Variants of the following result are [7, Lemma 6.7] and [3, Lemma 4.7]

Lemma 2.7. *Let $a \in X$ and let $\mu : [0, T] \to \mathcal{P}_c(\mathbb{R}^d)$ and $\nu : [0, T] \to \mathcal{P}_1(\mathbb{R}^d)$ be two continuous maps with respect to $\mathcal{W}_1$ satisfying*

$$\mathrm{supp}(\mu(t)) \cup \mathrm{supp}(\nu(t)) \subseteq B(0, R), \tag{6}$$

for every $t \in [0, T]$, for some $R > 0$. Then for every $r > 0$ there exists a constant $L_{a,r,R}$ such that

$$\|F[a] * \mu(t) - F[a] * \nu(t)\|_{L^\infty(B(0,r))} \leq L_{a,r,R} \mathcal{W}_1(\mu(t), \nu(t)) \tag{7}$$

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

Proof. Fix $t \in [0, T]$ and take $\pi \in \Gamma_o(\mu(t), \nu(t))$. Since the marginals of π are by definition $\mu(t)$ and $\nu(t)$, it follows

$$\begin{aligned} F[a] * \mu(t)(x) - F[a] * \nu(t)(x) &= \int_{B(0,R)} F[a](x - y) d\mu(t)(y) - \int_{B(0,R)} F[a](x - z) d\nu(t)(z) \\ &= \int_{B(0,R)^2} (F[a](x - y) - F[a](x - z))\, d\pi(y, z) \end{aligned}$$

By using Lemma 2.4 and the hypothesis (6), we have

$$\begin{aligned} \|F[a] * \mu(t) - F[a] * \nu(t)\|_{L^\infty(B(0,r))} &\leq \operatorname*{ess\,sup}_{x \in B(0,r)} \int_{B(0,R)^2} |F[a](x - y) - F[a](x - z)|\, d\pi(y, z) \\ &\leq \mathrm{Lip}_{B(0,R+r)}(F[a]) \int_{B(0,R)^2} |y - z| d\pi(y, z) \\ &= \mathrm{Lip}_{B(0,R+r)}(F[a]) \mathcal{W}_1(\mu(t), \nu(t)), \end{aligned}$$

hence (7) holds with $L_{a,r,R} = \mathrm{Lip}_{B(0,R+r)}(F[a])$. □

The following result provides a strong link between solutions of system (3) and those of system (2), one that in the end enables us to state an existence result for the latter ones.

Proposition 2.8. *Let $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ be given. Let $(\mu_0^N)_{N\in\mathbb{N}} \subset \mathcal{P}_c(\mathbb{R}^d)$ be a sequence of empirical measures of the form*

$$\mu_0^N(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - x_{0,i}^N), \qquad \text{for some } x_{0,i}^N \in \mathrm{supp}(\mu_0) + \overline{B(0,1)}$$

satisfying $\lim_{N\to\infty} \mathcal{W}_1(\mu_0, \mu_0^N) = 0$. *For every $N \in \mathbb{N}$, denote with $\mu^N : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ the curve given by* (5) *where $(x_1^N, \ldots, x_N^N)$ is the unique solution of system* (3).

Then, the sequence $(\mu^N)_{N\in\mathbb{N}}$ converges, up to subsequences, to a solution μ of (2) *with initial datum μ_0. Moreover, there exists $R > 0$ depending only on $T, a,$ and $\mathrm{supp}(\mu_0)$ such that it holds*

$$\mathrm{supp}(\mu^N(t)) \cup \mathrm{supp}(\mu(t)) \subseteq B(0,R), \qquad \text{for every } N \in \mathbb{N} \text{ and } t \in [0,T].$$

Proof. Notice that for every $N \in \mathbb{N}$, by Proposition 2.2, μ^N is the unique solution of (2) with initial datum μ_0^N. We start by fixing $N \in \mathbb{N}$ and estimating the growth of $|x_i^N(t)|^2$ for $i = 1, \ldots, N$. By using Lemma 2.3, we have

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}|x_i^N(t)|^2 &\le \dot{x}_i^N(t) \cdot x_i^N(t) \\
&\le \left|(F[a] * \mu^N(t))(x_i(t))\right| |x_i^N(t)| \\
&\le \|a\|_{L^\infty(\mathbb{R}_+)} \left(|x_i^N(t)| + \frac{1}{N}\sum_{j=1}^N |x_j^N(t)| \right) |x_i^N(t)| \\
&\le 2\|a\|_{L^\infty(\mathbb{R}_+)} \max_{j=1,\ldots,N} |x_j^N(t)||x_i^N(t)| \\
&\le 2\|a\|_{L^\infty(\mathbb{R}_+)} \max_{j=1,\ldots,N} |x_j^N(t)|^2.
\end{aligned}$$

If we denote by $q(t) := \max_{j=1,\ldots,N} |x_j^N(t)|^2$, then the Lipschitz continuity of q implies that q is a.e. differentiable. Stampacchia's Lemma [9, Chapter 2, Lemma A.4] ensures that for a.e. $t \in [0,T]$ there exists $k = 1, \ldots, N$ such that

$$\dot{q}(t) = \frac{d}{dt}|x_k^N(t)|^2 \le 4\|a\|_{L^\infty(\mathbb{R}_+)} q(t).$$

Hence, Gronwall's Lemma and the hypothesis $x_{0,i}^N \in \mathrm{supp}(\mu_0) + \overline{B(0,1)}$ for every $N \in \mathbb{N}$ and $i = 1, \ldots, N$, imply that

$$q(t) \le q(0)e^{4\|a\|_{L^\infty(\mathbb{R}_+)}t} \le C_0 e^{4\|a\|_{L^\infty(\mathbb{R}_+)}t} \text{ for a.e. } t \in [0,T],$$

for some uniform constant C_0 depending only on μ_0. Therefore, the trajectory $\mu^N(\cdot)$ is bounded uniformly in N in a ball $B(0,R) \subset \mathbb{R}^d$, where

$$R = \sqrt{C_0}e^{2\|a\|_{L^\infty(\mathbb{R}_+)}T}. \tag{8}$$

This, in turn, implies that $\mu^N(\cdot)$ is Lipschitz continuous with Lipschitz constant uniform in N, since by the fact that $|x_i^N(t)| \leq R$ for a.e. $t \in [0,T]$, for all $N \in N$ and $i = 1, \ldots, N$, and Lemma 2.3 follows

$$\begin{aligned}|\dot{x}_i^N(t)| &= |(F[a] * \mu^N(t))(x_i^N(t))| \\ &\leq \|a\|_{L^\infty(\mathbb{R}_+)} \left(|x_i^N(t)| + \frac{1}{N}\sum_{j=1}^N |x_j^N(t)| \right) \\ &\leq 2R\|a\|_{L^\infty(\mathbb{R}_+)}.\end{aligned}$$

We have thus found a sequence $(\mu^N)_{N\in\mathbb{N}} \subset \mathcal{C}^0([0,T], \mathcal{P}_1(B(0,R)))$ for which the following holds:

- $(\mu^N)_{N\in\mathbb{N}}$ is equicontinuous and closed, because of the uniform Lipschitz constant $2R\|a\|_{L^\infty(\mathbb{R}_+)}$;
- for every $t \in [0,T]$, the sequence $(\mu^N(t))_{N\in\mathbb{N}}$ is relatively compact in $\mathcal{P}_1(B(0,R))$. This holds because $(\mu^N(t))_{N\in\mathbb{N}}$ is a tight sequence, since $B(0,R)$ is compact, and hence relatively compact due to Prokhorov's Theorem.

Therefore, we can apply the Ascoli-Arzelá Theorem for functions with values in a metric space (see for instance, [8, Chapter 7, Theorem 18]) to infer the existence of a subsequence $(\mu^{N_k})_{k\in\mathbb{N}}$ of $(\mu^N)_{N\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \mathcal{W}_1(\mu^{N_k}(t), \mu(t)) = 0 \quad \text{uniformly for a.e. } t \in [0,T], \tag{9}$$

for some $\mu \in \mathcal{C}^0([0,T], \mathcal{P}_1(B(0,R)))$ with Lipschitz constant bounded by $2R\|a\|_{L^\infty(\mathbb{R}_+)}$. The hypothesis $\lim_{N\to\infty} \mathcal{W}_1(\mu_0^N, \mu_0) = 0$ now obviously implies $\mu(0) = \mu_0$.

We are now left with verifying that this curve μ is a solution of (2). For all $t \in [0,T]$ and for all $\varphi \in \mathcal{C}_c^1(\mathbb{R}^d;\mathbb{R})$, since it holds

$$\frac{d}{dt}\langle \varphi, \mu^N(t)\rangle = \frac{1}{N}\frac{d}{dt}\sum_{i=1}^N \varphi(x_i^N(t)) = \frac{1}{N}\sum_{i=1}^N \nabla\varphi(x_i^N(t)) \cdot \dot{x}_i^N(t),$$

by directly applying the substitution $\dot{x}_i^N(t) = (F[a] * \mu^N(t))(x_i^N(t))$, we have

$$\langle \varphi, \mu^N(t) - \mu^N(0)\rangle = \int_0^t \left[\int_{\mathbb{R}^d} \nabla\varphi(x) \cdot (F[a] * \mu^N(s))(x) d\mu^N(s)(x) \right] ds.$$

By Lemma 2.7, the inequality (9), and the compact support of $\varphi \in \mathcal{C}_c^1(\mathbb{R}^d;\mathbb{R})$, follows

$$\lim_{N\to\infty} \|\nabla\varphi \cdot (F[a] * \mu^N(t) - F[a] * \mu(t))\|_{L^\infty(\mathbb{R}^d)} = 0 \quad \text{uniformly for a.e. } t \in [0,T].$$

If we denote with $\mathcal{L}^1\llcorner_{[0,t]}$ the Lebesgue measure on the time interval $[0,t]$, since the product measures $\frac{1}{t}\mu^N(s) \times \mathcal{L}^1\llcorner_{[0,t]}$ converge in $\mathcal{P}_1([0,t] \times \mathbb{R}^d)$ to $\frac{1}{t}\mu(s) \times \mathcal{L}^1\llcorner_{[0,t]}$, we finally get from the dominated convergence theorem that

$$\lim_{N\to\infty} \int_0^t \int_{\mathbb{R}^d} \nabla\phi(x) \cdot (F[a]*\mu^N(s))(x) d\mu^N(s)(x) ds$$
$$= \int_0^t \int_{\mathbb{R}^d} \nabla\phi(x) \cdot (F[a] * \mu(s))(x) d\mu(s)(x) ds,$$

which proves that μ is a solution of (2) with initial datum μ_0. $\square$

2.2 The transport map and a uniqueness result

Another way for building a solution of system (2) is by means of the so-called *transport map*, i.e., the function describing the evolution in time of the initial measure μ_0. The transport map can be constructed by considering the following one-agent version of system (4),

$$\begin{cases} \dot{\xi}(t) = (F[a] * \mu(t))(\xi(t)) & \text{for } t \in (0,T], \\ \xi(0) = \xi_0, \end{cases} \tag{10}$$

where ξ is a mapping from $[0,T]$ to $\mathbb{R}^d$ and $a \in X$. Here $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ is a continuous map with respect to the Wasserstein distance $\mathcal{W}_1$ satisfying $\mu(0) = \mu_0$ and $\mathrm{supp}(\mu(t)) \subseteq B(0,R)$, where R is given by (8) from the choice of T, a and μ_0.

For the reader's convenience we start by briefly recalling some general, well-known results about solutions to Carathéodory differential equations. We fix a domain $\Omega \subset \mathbb{R}^d$, a Carathéodory function $g\colon [0,T] \times \Omega \to \mathbb{R}^d$, and $0 < \tau \leq T$. A function $y\colon [0,\tau] \to \Omega$ is called a solution of the Carathéodory differential equation

$$\dot{y}(t) = g(t, y(t)) \tag{11}$$

on $[0,\tau]$ if and only if y is absolutely continuous and (11) is satisfied a.e. in $[0,\tau]$. The following existence result holds.

Theorem 2.9. *Fix $T > 0$ and $y_0 \in \mathbb{R}^d$. Suppose that there exists a compact subset Ω of $\mathbb{R}^d$ such that $y_0 \in \mathrm{int}(\Omega)$ and there exists $m_\Omega \in L^1([0,T])$ for which it holds*

$$|g(t,y)| \leq m_\Omega(t), \tag{12}$$

for a.e. $t \in [0,T]$ and for all $y \in \Omega$. Then there exists a $\tau > 0$ and a solution $y(t)$ of (11) defined on the interval $[0,\tau]$ which satisfies $y(0) = y_0$. If there exists $C > 0$ such that the function g also satisfies the condition

$$|g(t,y)| \leq C(1 + |y|), \tag{13}$$

for a.e. $t \in [0,T]$ *and every* $y \in \Omega$*, and it holds* $B(0,R) \subseteq \Omega$*, for* $R > |y_0| + CTe^{CT}$*, then the local solution* $y(t)$ *of* (11) *which satisfies* $y(0) = y_0$ *can be extended to the whole interval* $[0,T]$*. Moreover, for every* $t \in [0,T]$*, any solution satisfies*

$$|y(t)| \leq \Big(|y_0| + Ct\Big)\, e^{Ct}. \tag{14}$$

Proof. Since $y_0 \in \mathrm{int}(\Omega)$, we can consider a ball $B(y_0, r) \subset \Omega$. The classical result [5, Chapter 1, Theorem 1] and (12) yield the existence of a local solution defined on an interval $[0,\tau]$ and taking values in $B(y_0, r)$.

If (13) holds, any solution of (11) with initial datum y_0 satisfies

$$|y(t)| \leq |y_0| + Ct + C\int_0^t |y(s)|\, ds$$

for every $t \in [0,\tau]$, therefore (14) follows from Gronwall's inequality. In particular the graph of a solution $y(t)$ cannot reach the boundary of $[0,T] \times B(0, |y_0| + CTe^{CT})$ unless $\tau = T$, therefore the continuation of the local solution to a global one on $[0,T]$ follows, for instance, from [5, Chapter 1, Theorem 4]. □

Gronwall's inequality easily gives us the following results on continuous dependence on the initial data.

Lemma 2.10. *Let* g_1 *and* $g_2 \colon [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ *be Carathéodory functions both satisfying* (13) *for the same constant* $C > 0$*. Let* $r > 0$ *and define*

$$\rho_{r,C,T} := \Big(r + CT\Big)\, e^{CT}\, .$$

Assume in addition that there exists a constant $L > 0$ *satisfying*

$$|g_1(t, y_1) - g_1(t, y_2)| \leq L|y_1 - y_2|$$

for every $t \in [0,T]$ *and every* y_1*,* y_2 *such that* $|y_i| \leq \rho_{r,C,T}$*,* $i = 1,2$*. Then, if* $\dot{y}_1(t) = g_1(t, y_1(t))$*,* $\dot{y}_2(t) = g_2(t, y_2(t))$*,* $|y_1(0)| \leq r$ *and* $|y_2(0)| \leq r$*, one has*

$$|y_1(t) - y_2(t)| \leq e^{Lt}\left(|y_1(0) - y_2(0)| + \int_0^t \|g_1(s,\cdot) - g_2(s,\cdot)\|_{L^\infty(B(0,\rho_{r,C,T}))}\, ds\right) \tag{15}$$

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

Proof. We can bound $|y_1(t) - y_2(t)|$ from above as follows:

$$\begin{aligned} |y_1(t) - y_2(t)| &\leq |y_1(0) - y_2(0)| + \int_0^t |\dot{y}_1(s) - \dot{y}_2(s)| ds \\ &= |y_1(0) - y_2(0)| \\ &\quad + \int_0^t |g_1(s, y_1(s)) - g_1(s, y_2(s)) + g_1(s, y_2(s)) - g_2(s, y_2(s))| ds \end{aligned}$$

$$
\begin{aligned}
&\leq |y_1(0) - y_2(0)| + \int_0^t \|g_1(s,\cdot) - g_2(s,\cdot)\|_{L^\infty(B(0,\rho_{r,C,T}))}\, ds \\
&\quad + L \int_0^t |y_1(s) - y_2(s)| ds.
\end{aligned}
$$

Since the function $\alpha(t) = |y_1(0) - y_2(0)| + \int_0^t \|g_1(s,\cdot) - g_2(s,\cdot)\|_{L^\infty(B(0,\rho_{r,C,T}))}\, ds$ is increasing, an application of Gronwall's inequality gives (15), as desired. $\square$

Proposition 2.11. *Fix $T > 0$, $a \in X$, $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$, $\xi_0 \in \mathbb{R}^d$ and let $R > 0$ be given by Proposition 2.8 from the choice of T, a and μ_0. For every map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ which is continuous with respect to $\mathcal{W}_1$ such that*

$$
\operatorname{supp}(\mu(t)) \subseteq B(0,R) \quad \text{for every } t \in [0,T],
$$

there exists a unique solution of system (10) *with initial value μ_0 defined on the whole interval* $[0,T]$.

Proof. By Lemma 2.3 follows that, for any compact set $K \subset \mathbb{R}^d$ containing ξ_0, there exists a function $m_K \in L^1([0,T])$ for which the function $g(t,y) = (F[a] * \mu(t))(y)$ satisfies (12). Moreover, for fixed t this function is locally Lipschitz continuous, as follows from Lemma 2.5, thus $g(t,y) = (F[a] * \mu(t))(y)$ is a Carathéodory function.

From the hypothesis that the support of μ is contained in $B(0,R)$ and Lemma 2.3, follows the existence of a constant C depending on T, a and μ_0 such that

$$
|(F[a] * \mu(t))(y)| \leq C(1 + |y|)
$$

holds for every $y \in \mathbb{R}^d$ and for every $t \in [0,T]$. Hence $F[a] * \mu(t)$ is sublinear and (13) holds. By considering a sufficiently large compact set K containing ξ_0, Theorem 2.9 guarantees the existence of a solution of system (10) defined on $[0,T]$.

To establish uniqueness notice that, from Lemma 2.4, for every compact subset $K \in \mathbb{R}^d$ and any $x, y \in K$, it holds

$$
\begin{aligned}
|(F[a] * \mu(t))(x) - (F[a] * \mu(t))(y)| &\leq \left| \int_{\mathbb{R}^d} F[a](x-z) d\mu(t)(z) - \int_{\mathbb{R}^d} F[a](y-z) d\mu(t)(z) \right| \\
&\leq \int_{\mathbb{R}^d} |F[a](x-z) - F[a](y-z)|\, d\mu(t)(z) \\
&\leq \operatorname{Lip}_{\widehat{K}}(F[a]) |x-y|,
\end{aligned} \tag{16}
$$

where $\widehat{K}$ is a compact set containing both K and $B(0,R)$. Hence, uniqueness follows from (16) and Lemma 2.10 by taking $g_1 = g_2$, $y_1(0) = y_2(0)$ and $r = |y_1(0)|$. $\square$

We can therefore consider the family of flow maps $\mathcal{T}_t^\mu : \mathbb{R}^d \to \mathbb{R}^d$, indexed by $t \in [0,T]$ and the choice of the mapping μ, defined by

$$
\mathcal{T}_t^\mu(\xi_0) = \xi(t),
$$

where $\xi : [0,T] \to \mathbb{R}^d$ is the unique solution of (10) with initial datum ξ_0. The classical result [3, Theorem 3.10] shows that the solution of (2) with initial value μ_0 is the unique fixed-point of the *push-foward map*

$$\Gamma[\mu](t) := (\mathcal{T}_t^\mu)_\# \mu_0. \tag{17}$$

A first, basic property of the transport map is proved in the following

Proposition 2.12. *$\mathcal{T}_t^\mu$ is a locally bi-Lipschitz map, i.e. it is a bijective locally Lipschitz map, with locally Lipschitz inverse.*

Proof. The choice $r = R$ in Lemma 2.10 and the inequality (16) trivially implies the following stability estimate

$$\left|\mathcal{T}_t^\mu(x_0) - \mathcal{T}_t^\mu(x_1)\right| \le e^{T \operatorname{Lip}_{B(0,R)}(F[a])}|x_0 - x_1|, \quad \text{for } |x_i| \le R\,, \quad i = 0,1\,. \tag{18}$$

i.e., $\mathcal{T}_t^\mu$ is locally Lipschitz.

In view of the uniqueness of the solutions to the ODE (10), it is furthermore clear that, for any $t_0 \in [0,T]$, the inverse of $\mathcal{T}_{t_0}^\mu$ is given by the transport map associated to the backward-in-time ODE

$$\begin{cases} \dot\xi(t) = (F[a] * \mu(t))(\xi(t)) & \text{for } t \in [0,t_0), \\ \xi(t_0) = \xi_0. \end{cases}$$

However, this problem in turn can be cast into the form of an usual IVP simply by considering the reverse trajectory $\nu_t = \mu_{t_0 - t}$. Then $y(t) = \xi(t_0 - t)$ solves

$$\begin{cases} \dot y(t) = -\big(F[a] * \nu(t)\big)(y(t)) & \text{for } t \in (0,t_0], \\ y(0) = \xi(t_0). \end{cases}$$

The corresponding stability estimate for this problem then yields that the inverse of $\mathcal{T}_t^\mu$ is indeed locally Lipschitz too (with the same local Lipschitz constant). $\square$

The following Lemma and (15) are the main ingredients of the forthcoming result on continuous dependance on initial data.

Lemma 2.13. *Let $\mathcal{T}_1$ and $\mathcal{T}_2 \colon \mathbb{R}^n \to \mathbb{R}^n$ be two bounded Borel measurable functions. Then, for every $\mu \in \mathcal{P}_1(\mathbb{R}^n)$ one has*

$$\mathcal{W}_1((\mathcal{T}_1)_\#\mu, (\mathcal{T}_2)_\#\mu) \le \|\mathcal{T}_1 - \mathcal{T}_2\|_{L^\infty(\operatorname{supp}\mu)}.$$

If in addition $\mathcal{T}_1$ is locally Lipschitz continuous, and μ, $\nu \in \mathcal{P}_1(\mathbb{R}^n)$ are both compactly supported on a ball $B(0,r)$ of $\mathbb{R}^n$, then

$$\mathcal{W}_1((\mathcal{T}_1)_\#\mu, (\mathcal{T}_1)_\#\nu) \le \operatorname{Lip}_{B(0,r)}(E_1)\mathcal{W}_1(\mu,\nu).$$

Proof. See [3, Lemma 3.11] and [3, Lemma 3.13]. $\square$

Theorem 2.14. *Fix $T > 0$ and let $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ and $\nu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ be two equi-compactly supported solutions of* (2). *Let $\mu_0 := \mu(0)$ and $\nu_0 := \nu(0)$. Consider $R > 0$ such that*

$$supp(\mu(t)) \cup \mathrm{supp}(\nu(t)) \subseteq B(0,R) \tag{19}$$

for every $t \in [0,T]$. Then, there exist a positive constant $\overline{C}$ depending only on T, a, r, and R such that

$$\mathcal{W}_1(\mu(t), \nu(t)) \leq \overline{C}\, \mathcal{W}_1(\mu_0, \nu_0) \tag{20}$$

for every $t \in [0,T]$. In particular, equi-compactly supported solutions of (2) *are uniquely determined by the initial datum.*

Proof. Let $\mathcal{T}_t^\mu$ and $\mathcal{T}_t^\nu$ be the flow maps associated to system (10) with measure μ and ν, respectively. By (17), the triangle inequality, Lemma 2.7, Lemma 2.13 and (18) we have for every $t \in [0,T]$

$$\begin{aligned}
\mathcal{W}_1(\mu(t), \nu(t)) &= \mathcal{W}_1((\mathcal{T}_t^\mu)_\# \mu_0, (\mathcal{T}_t^\nu)_\# \nu_0) \\
&\leq \mathcal{W}_1((\mathcal{T}_t^\mu)_\# \mu_0, (\mathcal{T}_t^\mu)_\# \nu_0) + \mathcal{W}_1((\mathcal{T}_t^\mu)_\# \nu_0, (\mathcal{T}_t^\nu)_\# \nu_0) \\
&\leq e^{T \,\mathrm{Lip}_{B(0,R)}(F[a])} \mathcal{W}_1(\mu_0, \nu_0) + \|\mathcal{T}_t^\mu - \mathcal{T}_t^\nu\|_{L^\infty(B(0,R))}.
\end{aligned} \tag{21}$$

Using (15) with $y_1(0) = y_2(0)$ we get

$$\|\mathcal{T}_t^\mu - \mathcal{T}_t^\nu\|_{L^\infty(B(0,r))} \leq e^{t \,\mathrm{Lip}_{B(0,R)}(F[a])} \int_0^t \|F[a] * \mu(s) - F[a] * \nu(s)\|_{L^\infty(B(0,R))}\, ds. \tag{22}$$

Combining (21) and (22) with Lemma 2.7, we have

$$\mathcal{W}_1(\mu(t), \nu(t)) \leq e^{T \,\mathrm{Lip}_{B(0,R)}(F[a])} \left(\mathcal{W}_1(\mu_0, \nu_0) + L_{a,R,R} \int_0^t \mathcal{W}_1(\mu(s), \nu(s))\, ds \right)$$

for every $t \in [0,T]$, where $L_{a,R,R}$ is the constant from Lemma 2.7. Gronwall's inequality now gives

$$\mathcal{W}_1(\mu(t), \nu(t)) \leq e^{T \,\mathrm{Lip}_{B(0,R)}(F[a]) + L_{a,R,R}} \mathcal{W}_1(\mu_0, \nu_0),$$

which is exactly (20) with $\overline{C} = e^{T \,\mathrm{Lip}_{B(0,R)}(F[a]) + L_{a,R,R}}$.

Consider now two solutions of (2) with the same initial datum μ_0. Since, from Proposition 2.8 they both satisfy (19) for the given *a priori known* R given by (8), then (20) guarantees they both describe the same curve in $\mathcal{P}_1(\mathbb{R}^d)$. This concludes the proof. □

3 The error functional E

As already explained in the introduction, our goal is to learn a target function $a \in X$ (which we fix throughout the section) from the observation of the dynamics μ that stems from system (2) with a as interaction kernel, μ_0 as initial datum and T as finite time horizon.

A very reasonable and intuitive strategy would be to pick a among those functions in X which would give rise to a dynamics similar to μ. In light of the close bond between μ and the empirical measures μ^N which are solutions to (3) (see Proposition 2.8), let us address the same problem for μ^N, first. In this case we deal with the discrete system (4) and, by assumption, we know the trajectories of the agents $x_i^N(\cdot)$ *and* their velocities $\dot{x}_i^N(\cdot)$ (or, at the very least, an approximation of them as difference quotients). The function a should then be chosen as the function $\widehat{a} \in X$ minimizing the following *discrete error functional*

$$E_N(\widehat{a}) = \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\sum_{j=1}^N \left(\widehat{a}(|x_i^N(t) - x_j^N(t)|)(x_i^N(t) - x_j^N(t)) - \dot{x}_i^N(t)\right)\right|^2 dt, \quad (23)$$

among all functions $\widehat{a} \in X$. Notice that the functional E_N has the remarkable property of being easily computable from the knowledge of x_i^N and $\dot{x}_i^N$.

To link the discrete error functional (23) to the problem of learning a for μ, notice that, by means of the empirical measure μ^N, E_N can be rewritten as follows

$$\begin{aligned} E_N(\widehat{a}) &= \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\sum_{j=1}^N \big(F[\widehat{a}] - F[a]\big)(x_i - x_j)\right|^2 dt \\ &= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \big(F[\widehat{a}] - F[a]\big) * \mu^N(t)\right|^2 d\mu^N(t)(x)dt, \end{aligned}$$

This form of E_N clearly gives a specific candidate for the error functional to be minimized in the case of μ, that is

$$E(\widehat{a}) = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \big(F[\widehat{a}] - F[a]\big) * \mu(t)\right|^2 d\mu(t)(x)dt.$$

Hence, we can now formulate a tentative strategy for the learning of a given the dynamics of μ (with initial value $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$):

1. choose $N \in \mathbb{N}$ *sufficiently large* and draw N i.i.d. initial values $x_{0,1}^N, \dots, x_{0,N}^N \sim \mu_0$. By defining $\mu_0^N(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - x_{0,i}^N)$ for every $x \in \mathbb{R}^d$, it is well-known from [6, Lemma 3.3] that it holds

$$\lim_{N\to\infty} \mathcal{W}_1(\mu_0^N, \mu_0) = 0; \quad (24)$$

2. compute μ^N, the solution of system (2) with initial datum μ_0^N. By (20) and (24) then follows

$$\lim_{N\to\infty} \mathcal{W}_1(\mu^N(t), \mu(t)) = 0 \qquad \text{for every } t \in [0, T]; \quad (25)$$

3. compute a minimizer $\widehat{a}_N$ for E_N among all functions in a proper subset $V_N \subset X$ where this computation is feasible. The convergence of trajectories (25) and the condition that $V_N \hookrightarrow X$ for $N \to +\infty$ shall guarantee that $\widehat{a}_N$ is a *sufficiently good* approximation of a.

In the next sections we shall make point 3. of the list above precise and state under which conditions on the function a and the error functional E the above program can be successful. In order to do this, let us first look at E more closely and see whether additional coercivity assumptions are required in order to ensure that a is the unique minimizer of E.

Since by Proposition 2.8 follows that $\operatorname{supp}(\mu(t)) \subseteq B(0, R)$, where R is given by (8), whenever $x, y \in \mathbb{R}^d$ are drawn from the probability distribution $\mu(t)$ the estimate

$$|F[\widehat{a}](x-y) - F[a](x-y)| \leq 2R|\widehat{a}(|x-y|) - a(|x-y|)|$$

holds. It is then clearly true that

$$E(\widehat{a}) \leq \frac{4R^2}{T} \int_0^T \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \big|\widehat{a}(|x-y|) - a(|x-y|)\big| d\mu(t)(y) \right)^2 d\mu(t)(x) dt.$$

Now observe that for any $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ the following estimate holds by Hölder's inequality

$$\int_{\mathbb{R}^d} |f(x)| d\nu(x) \leq \left(\int_{\mathbb{R}^d} |f(x)|^2 d\nu(x) \right)^{1/2}.$$

Hence, E can be bounded from above as

$$E(\widehat{a}) \leq \frac{4R^2}{T} \int_0^T \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \big|\widehat{a}(|x-y|) - a(|x-y|)\big|^2 d\mu(t)(y) d\mu(t)(x) dt.$$

Using the distance map

$$d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+ , \qquad (x, y) \mapsto d(x, y) = |x-y| ,$$

we define by push-forward the probability measure-valued mapping $\varrho : [0, T] \to \mathcal{P}_1(\mathbb{R}_+)$ defined for every Borel set $A \subset \mathbb{R}_+$ as

$$\varrho(t)(A) = (\mu(t) \otimes \mu(t))\big(d^{-1}(A)\big).$$

With the introduction of ϱ we can further estimate

$$E(\widehat{a}) \leq \frac{4R^2}{T} \int_0^T \int_{\mathbb{R}_+} \big|\widehat{a}(s) - a(s)\big|^2 d\varrho(t)(s) dt. \tag{26}$$

The idea, now, would be to find a proper measure ρ for which we can state the equivalence

$$E(\widehat{a}) \sim \|\widehat{a} - a\|_{L^2(\mathbb{R}_+, \rho)},$$

which would actually imply that minimizing E is actually equivalent to find (an $L^2(\mathbb{R}_+, \rho)$-equivalent of) the interaction kernel a. But, in order to go on with the estimate, we need to focus our attention on the properties of the family of measures $(\varrho(t))_{t \in [0,T]}$.

3.1 The measure ρ

Lemma 3.1. *For every open set $A \subseteq \mathbb{R}_+$ the mapping $t \in [0,R] \mapsto \varrho(t)(A)$ is lower semi-continuous, whereas for any compact set A it is upper semi-continuous.*

Proof. As a first step we show that for every given sequence $(t_n)_{n\in\mathbb{N}}$ converging to $t \in [0,T]$ we have the weak convergence $\varrho(t_n) \rightharpoonup \varrho(t)$ for $n \to +\infty$. For this, in turn we first prove the weak convergence of the product measure $\mu(t_n) \otimes \mu(t_n) \rightharpoonup \mu(t) \otimes \mu(t)$.

It is a basic property of the space $\mathcal{C}(\mathbb{R}^d \times \mathbb{R}^d)$ that it coincides with the inductive tensor product $\mathcal{C}(\mathbb{R}^d) \otimes_\varepsilon \mathcal{C}(\mathbb{R}^d)$. In particular, functions of the form $h = \sum_{j=1}^J f_j \otimes g_j$ with $f_j, g_j \in \mathcal{C}(\mathbb{R}^d)$, for $j = 1, \ldots, J$ and $J \in \mathbb{N}$, are a dense subspace of $\mathcal{C}(\mathbb{R}^{2d})$. Hence, to prove the weak convergence of measures on $\mathbb{R}^{2d}$, we can restrict the proof to functions of this form. Due to linearity of integrals, this can be further reduced to simple tensor products of the form $h = f \otimes g$.

For such tensor products we can directly apply Fubini's Theorem and the weak convergence $\mu(t_n) \rightharpoonup \mu(t)$ (which is a consequence of the continuity of μ w.r.t. the Wasserstein metric $\mathcal{W}_1$), and find

$$\int_{\mathbb{R}^{2d}} f \otimes g \, d(\mu(t_n) \otimes \mu(t_n)) = \int_{\mathbb{R}^d} f d\mu(t_n) \cdot \int_{\mathbb{R}^d} g d\mu(t_n) \overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} f d\mu(t) \cdot \int_{\mathbb{R}^d} g d\mu(t, .$$

This implies the claimed weak convergence $\varrho(t_n) \rightharpoonup \varrho(t)$, since for any function $f \in \mathcal{C}(\mathbb{R}_+)$ we have that the continuity of d implies continuity of $f \circ d$, and hence

$$\begin{aligned}\int_{\mathbb{R}_+} f \, d\varrho(t_n) &= \int_{\mathbb{R}^{2d}} (f \circ d)(x,y) d(\mu(t_n) \otimes \mu(t_n))(x,y) \\ &\overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^{2d}} (f \circ d)(x,y) d(\mu(t) \otimes \mu(t))(x,y) = \int_{\mathbb{R}_+} f \, d\varrho(t).\end{aligned}$$

The claim now follows from general results for weakly* convergent sequences of Radon measures, see e.g. [1, Proposition 1.62]. □

Lemma 3.1 justifies the following

Definition 3.2. The probability measure ρ on the Borel σ-algebra on $\mathbb{R}_+$ is defined for any Borel set $A \subseteq \mathbb{R}_+$ as follows

$$\rho(A) := \frac{1}{T} \int_0^T \varrho(t)(A) dt. \tag{27}$$

Notice that Lemma 3.1 shows that (27) is well-defined only for sets A that are open or compact in $\mathbb{R}_+$. This directly implies that ρ can be extended to any Borel set A, since both families of sets provide a basis for the Borel σ-algebra on $\mathbb{R}_+$. Notice that, in addition, from the semicontinuity properties shown in Lemma 3.1 we infer that for any Borel set A it holds

$$\rho(A) = \begin{cases} \sup\{\rho(F) : F \subseteq A, \ F \text{ compact}\}, \\ \inf\{\rho(G) : A \subseteq G, \ G \text{ open}\}, \end{cases}$$

which shows that ρ is a regular measure on $\mathbb{R}_+$.

The measure ρ has a deep relationship with our learning process: it tells us which regions of $\mathbb{R}_+$ (the set of distances) where actually explored in the entire dynamics of the system, and hence where we can expect our learning process to be successful, since these are the zones where we do have information to reconstruct the function a.

We now proceed to show the absolute continuity of ρ w.r.t. the Lebesgue measure on $\mathbb{R}_+$.

Lemma 3.3. *Let μ_0 be absolutely continuous w.r.t. the d-dimensional Lebesgue measure $\mathcal{L}^d$. Then, for every $t \in [0,T]$, also the measures $\mu(t)$ are absolutely continuous w.r.t. $\mathcal{L}^d$.*

Proof. Let a Lebesgue null-set $A \subset \mathbb{R}^d$ be given. Put $B = (\mathcal{T}_t^\mu)^{-1}(A)$, the image of A under the inverse of the transport map $(\mathcal{T}_t^\mu)^{-1}$, which by Proposition 2.12 is a locally Lipschitz map. The claim now follows from showing $\mathcal{L}^d(B) = 0$, since by assumption we have $\mu_0(B) = 0$, which by definition gives us

$$0 = \mu_0(B) = \mu_0\big((\mathcal{T}_t^\mu)^{-1}(A)\big) = \mu(t)(A)\,.$$

Moreover, we can reduce this further to consider only $B \cap B(0,R)$ with R as in (8), since $\mu(t)(B \setminus B(0,R)) = 0$ for all $t \in [0,T]$ by Proposition 2.8. Hence we no longer need to distinguish between local and global Lipschitz maps.

It thus remains to show that the image of a Lebesgue null-set under a Lipschitz map is again a Lebesgue null-set. To see this, recall that a measurable set A has Lebesgue measure zero if and only if for every $\varepsilon > 0$ there exists a family of balls $B_1, B_2, \dots$ (or, equivalently, of cubes) such that

$$A \subset \bigcup_n B_n \qquad \text{and} \qquad \sum_n \mathcal{L}^d(B_n) < \varepsilon\,.$$

Let L be the Lipschitz constant of $(\mathcal{T}_t^\mu)^{-1}$, and $\operatorname{diam}(B_n)$ the diameter. Then clearly the image of B_n under $(\mathcal{T}_t^\mu)^{-1}$ is contained in a ball of diameter at most $L\operatorname{diam}(B_n)$. Denote those balls by $\widetilde{B}_n$. Then it immediately follows

$$(\mathcal{T}_t^\mu)^{-1}(A) \subset \bigcup_n \widetilde{B}_n \qquad \text{as well as} \qquad \sum_n \mathcal{L}^d(\widetilde{B}_n) = L^d \sum_n \mathcal{L}^d(B_n) < L^d\varepsilon\,.$$

Thus we have found a cover for $(\mathcal{T}_t^\mu)^{-1}(A)$ whose measure is bounded from above by (a multiple of) ε, which finally yields $\mathcal{L}^d((\mathcal{T}_t^\mu)^{-1}(A)) = 0$. $\square$

Lemma 3.4. *Let μ_0 be absolutely continuous w.r.t. $\mathcal{L}^d$. Then, for all $t \in [0,T]$, the measures $\varrho(t)$ and ρ are absolutely continuous w.r.t. $\mathcal{L}^1 \llcorner_{\mathbb{R}_+}$.*

Proof. Fix $t \in [0,T]$. By Lemma 3.3 we already know that $\mu(t)$ is absolutely continuous w.r.t. $\mathcal{L}^d$. This immediately implies that $\mu(t) \otimes \mu(t)$ is absolutely continuous w.r.t. $\mathcal{L}^{2d}$.

It hence remains to show that $d_{\#}\mathcal{L}^{2d}$ is absolutely continuous w.r.t. $\mathcal{L}^1 \llcorner \mathbb{R}_+$, where d is the distance function.

Let $A \subset \mathbb{R}_+$ be a Lebesgue null-set, and put $B = d^{-1}(A) \subset \mathbb{R}^{2d}$. Moreover, we denote by $B_x = \{y \in \mathbb{R}^d : |x-y| \in A\}$. Then clearly $B_{x+z} = z + B_x$. Moreover, using Fubini's Theorem we obtain

$$\mathcal{L}^{2d}(B) = \int_{\mathbb{R}^d} \mathcal{L}^d(B_x) d\mathcal{L}^d(x) \,.$$

It thus remains to show that $\mathcal{L}^d(B_x) = 0$ for one single $x \in \mathbb{R}^d$ (and thus for all, due to translation invariance of $\mathcal{L}^d$). However, to calculate $\mathcal{L}^d(B_0)$, we can pass to polar coordinates, and once again using Fubini's Theorem we obtain

$$\mathcal{L}^d(B_x) = \int_{\mathbb{R}^d} \chi_{B_0}(y) d\mathcal{L}^d(y) = \int_{S^d} \int_{\mathbb{R}_+} \chi_A(r) dr d\omega = \Omega_d \mathcal{L}^1(A) = 0 \,,$$

where Ω_d is the surface measure of the unit sphere S_d. This proves the absolute continuity of $\varrho(t)$, since

$$\mathcal{L}^1(A) = 0 \Longrightarrow \mathcal{L}^{2d}(d^{-1}(A)) \Longrightarrow (\mu(t) \otimes \mu(t))(d^{-1}(A)) = 0 \iff \varrho(t)(A) = 0 \,.$$

The absolute continuity of ρ now follows immediately from the one of $\varrho(t)$ for every t and its definition as an integral average (27). □

As an easy consequence that the dynamics of our system has support uniformly bounded in time, we get the following crucial properties of the measure ρ.

Lemma 3.5. *The measure ρ is finite and has compact support.*

Proof. To show that ρ is finite, we compute

$$\begin{aligned} \rho(\mathbb{R}_+) &= \frac{1}{T} \int_0^T \varrho(t)(\mathbb{R}_+) dt \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y| d\mu(t)(x) d\mu(t)(y) dt \\ &< +\infty, \end{aligned}$$

since the distance function is continuous and the support of μ is uniformly bounded in time.

Now, notice that the supports of the measures $\varrho(t)$ are the subsets of

$$K = d(B(0,R), B(0,R)) = \{|x-y| : x, y \in B(0,R)\},$$

where R is given by (8). Due to the continuity of d, this set K is a compact subset of $\mathbb{R}_+$, and we then obtain $\operatorname{supp} \rho \subseteq K$. □

Remark 1. While absolute continuity of μ_0 implies the same for ρ, the situation is different for purely atomic measures μ_0. On the one hand, also $\mu(t)$ is then purely atomic for every t, and this remains true for $\varrho(t)$. However, due to the averaging (27) involved in the definition of ρ, it generally cannot be atomic. For example, we obtain

$$\frac{1}{T}\int_0^T \delta(t)dt = \frac{1}{T}\mathcal{L}^1 \llcorner_{[0,T]} ,$$

as becomes immediately clear when integrating a continuous function against those kind of measures.

3.2 Coercivity assumptions and the existence of minimizers of E_N

By means of ρ, we can continue to estimate E from (26) as follows,

$$\begin{aligned} E(\widehat{a}) &\leq \frac{4R^2}{T}\int_0^T \int_{\mathbb{R}_+} \left|\widehat{a}(s) - a(s)\right|^2 d\varrho(t)(s)dt \\ &= 4R^2 \int_{\mathbb{R}_+} \left|\widehat{a}(s) - a(s)\right|^2 d\rho(s) \\ &= 4R^2 \|\widehat{a} - a\|^2_{L^2(\mathbb{R}_+,\rho)}. \end{aligned} \tag{28}$$

Equation (28) thus suggests the following additional condition to impose in order to ensure that a is the unique minimizer of E: we assume that there exists a constant $c > 0$ such that

$$E(\widehat{a}) \geq c\|\widehat{a} - a\|^2_{L^2(\mathbb{R}_+,\rho)}. \tag{29}$$

Notice that, as an easy consequence of Lemma 3.5, it holds

$$\|a\|^2_{L^2(\mathbb{R}_+,\rho)} = \int_{\mathbb{R}_+} \left|a(s)\right|^2 d\rho(s) \leq \|a\|_{L^\infty(\mathrm{supp}(\rho))}, \tag{30}$$

and hence, $X \subseteq L^2(\mathbb{R}_+, \rho)$. Thanks to this fact and to (29) we are able to prove the following result.

Proposition 3.6. *Assume that* (29) *holds. Then a is the unique minimizer of E among all functions in X.*

Proof. Notice that $E(a) = 0$, and since $E(\widehat{a}) \geq 0$ for all $\widehat{a} \in X$ this implies that a is a minimizer of E. Now suppose that $E(\widehat{a}) = 0$ for some $\widehat{a} \in X$. By (29) we obtain that $\widehat{a} = a$ in $L^2(\mathbb{R}_+, \rho)$, and by (30) follows that $\widehat{a} = a$ also in X. □

The following easy fact tells us the right ambient space where to state an existence result for the minimizers of E_N.

Proposition 3.7. *Fix $M > 0$ and define the set*

$$X_M = \left\{ b : \mathbb{R}_+ \to \mathbb{R}_+ : \begin{array}{c} b \text{ is continuous, weakly differentiable and} \\ \|b\|_{L^\infty(\mathbb{R}_+)} + \|b'\|_{L^\infty(\mathrm{supp}(\rho))} \le M \end{array} \right\}.$$

Then $X_M \subset X$ and it is relatively compact.

Proof. Notice that $X_M \subset X$ follows from the fact that ρ has compact support. Now, consider $(\widehat{a}_n)_{n\in\mathbb{N}} \subset X_M$. The Fundamental Theorem of Calculus (which is applicable for functions in $W^{1,p}$, see [1, Theorem 2.8]) tells us that, for any $r_1, r_2 \in \mathrm{supp}(\rho)$ it holds

$$a_n(r_1) - a_n(r_2) = \int_{r_1}^{r_2} a_n'(r)dr.$$

This implies

$$\big|a_n(r_1) - a_n(r_2)\big| \le \int_{r_1}^{r_2} |a_n'(r)|dr \le \|a_n'\|_{L^p(\mathrm{supp}(\rho))}|r_2 - r_1|.$$

In particular, the functions $\widehat{a}_n$ are all Lipschitz continuous with Lipschitz constant uniformly bounded by M, which in turn implies equi-continuity. They are moreover pointwise uniformly equibounded, since for every $r \in \mathrm{supp}(\rho)$ it holds

$$|\widehat{a}_n(r)| \le \|\widehat{a}_n\|_{L^\infty(\mathrm{supp}(\rho))} \le \|\widehat{a}_n\|_{L^\infty(\mathbb{R}_+)} \le M.$$

Hence from the Ascoli-Arzelá Theorem we can deduce the existence of a subsequence (which we do not relabel) converging uniformly on $\mathrm{supp}(\rho)$ to some $\widehat{a} \in X_M$, proving the statement. □

Proposition 3.8. *Fix $M > 0$ and let V be a closed subset of X_M w.r.t. the uniform convergence on* $\mathrm{supp}(\rho)$. *Then, the minimization problem*

$$\textit{minimize } E_N(\widehat{a}) \textit{ among all } \widehat{a} \in V$$

admits a solution.

Proof. In light of the fact that $\inf E_N \ge 0$, we can consider a minimizing sequence $(\widehat{a}_n)_{n\in\mathbb{N}} \subset V$, i.e., it holds $\lim_{n\to\infty} E_N(\widehat{a}_n) = \inf_V E_N$. By Proposition 3.7 there exists a subsequence of $(\widehat{a}_n)_{n\in\mathbb{N}}$ (which we do not relabel) converging uniformly on $\mathrm{supp}(\rho)$ to a function $\widehat{a} \in V$ (since V is closed). We now show that $\lim_{n\to\infty} E_N(\widehat{a}_n) = E_N(\widehat{a})$, from which shall follow the fact that E_N attains its minimum in V.

As a first step, notice that the uniform convergence of $(\widehat{a}_n)_{n\in\mathbb{N}}$ to $\widehat{a}$ on $\mathrm{supp}(\rho)$ and the compactness of $\mathrm{supp}(\rho)$, imply that the functionals $F[\widehat{a}_n](x-y)$ converge uniformly to $F[\widehat{a}](x-y)$ on $B(0,R) \times B(0,R)$ (where R is as in (8)). Furthermore, the fact that the measures $\mu^N(t)$ are compactly supported in $B(0,R)$ uniformly in time implies that

$$\begin{aligned} \sup_{x,y\in B(0,R)} |F[\widehat{a}_n](x-y) - F[a](x-y)| &= \sup_{x,y\in B(0,R)} |\widehat{a}_n(|x-y|) - a(|x-y|)||x-y| \\ &\le 2R \sup_{r\in\mathrm{supp}(\rho)} |\widehat{a}_n(r) - a(r)| \qquad (31) \\ &\le 2R(M + \|a\|_{L^\infty(\mathrm{supp}(\rho))}). \end{aligned}$$

Hence, we can apply three times the dominated convergence theorem to get

$$\begin{aligned}
\lim_{n\to\infty} E_N(\widehat{a}_n) &= \lim_{n\to\infty} \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \left(F[\widehat{a}_n](x-y) - F[a](x-y)\right) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x)dt \\
&= \frac{1}{T}\int_0^T \lim_{n\to\infty} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \left(F[\widehat{a}_n](x-y) - F[a](x-y)\right) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x)dt \\
&= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \lim_{n\to\infty} \int_{\mathbb{R}^d} \left(F[\widehat{a}_n](x-y) - F[a](x-y)\right) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x)dt \\
&= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \left(F[\widehat{a}](x-y) - F[a](x-y)\right) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x)dt \\
&= E_N(\widehat{a}),
\end{aligned}$$

which proves the statement.

□

Remark 2. The requirement to search for the minimizer inside the set X_M is indeed very strong, since it implies that one not only needs to know an upper bound for the target function a but also of its derivative. It is however true that such upper bounds need not be sharp, as they are only required as part of a compactness argument. Moreover, in real-life applications, these quantities may be preliminary computed thanks to a statistical analysis of the trajectories of the system under study.

4 Γ-convergence of the E_N to E

At the end of the last section, we have seen which are the main ingredients to ensure that our error functional E has the target interaction kernel a as unique minimizer (the coercivity assumption (29)), and for the well-posedness of the minimization problem of the functionals E_N (the knowledge of an upper bound for $\|a\|_{L^\infty(\mathbb{R}_+)}$ and $\|a'\|_{L^\infty(\mathrm{supp}(\rho))}$).

The last ingredient we now introduce is the key property that a family of approximation spaces V_N must possess in order to ensure that the minimizers of the functionals E_N are sufficiently close to those of E.

Definition 4.1. Let $M > 0$ be given. For every $N \in \mathbb{N}$, let V_N be a closed subset of X_M w.r.t. the convergence in $L^\infty(\mathrm{supp}(\rho))$ with the following property: for all $b \in X_M$ there exists a sequence $(b_N)_{N\in\mathbb{N}}$ converging uniformly to b on $\mathrm{supp}(\rho)$ and such that $b_N \in V_N$ for every $N \in \mathbb{N}$.

We now pass to provide a link between the minimizers of the functionals $(E_N)_{N\in\mathbb{N}}$ and those of E.

Theorem 4.2. *Assume $a \in X$, fix $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and set*

$$M \geq \|a\|_{L^\infty(\mathbb{R}_+)} + \|a'\|_{L^\infty(\mathrm{supp}(\rho))}.$$

For every $N \in \mathbb{N}$, let $x_{0,1}^N, \dots, x_{0,N}^N$ be i.i. μ_0-distributed and define E_N as in (23) *for the solution μ^N of system* (2) *with initial datum*

$$\mu_0^N(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - x_{0,i}^N), \text{ for every } x \in \mathbb{R}^d.$$

For every $N \in \mathbb{N}$, let $V_N \subset X_M$ be a sequence of subspaces satisfying Definition 4.1 and consider

$$\widehat{a}_N \in \arg\min_{\widehat{a}\in V_N} E_N(\widehat{a}).$$

Then the sequence $(\widehat{a}_N)_{N\in\mathbb{N}}$ converges uniformly to some continuous function $\widehat{a} \in X_M$ such that $E(\widehat{a}) = 0$. If we additionally assume the coercivity condition (29)*, then it holds $\widehat{a} = a$.*

We start with a technical lemma.

Lemma 4.3. *Let $(\widehat{a}_N)_{N\in\mathbb{N}} \subset L^2(\mathbb{R}_+, \rho)$ be a sequence of continuous functions uniformly converging to a function $\widehat{a}$. Let the functionals E_N be defined as in Theorem 4.2. Then it holds*

$$\lim_{N\to\infty} E_N(\widehat{a}_N) = E(\widehat{a}).$$

Proof. As already noticed before, from the hypothesis of Theorem 4.2 and [6, Lemma 3.3] it follows $\mathcal{W}_1(\mu_0, \mu_0^N) \to 0$ for $N \to \infty$. Hence, from (20) we have that $W_1(\mu(t), \mu^N(t)) \to 0$ for $N \to \infty$, uniformly for a.e. $t \in [0, T]$, which in particular implies the weak convergence of the sequence of measures $(\mu^N(t))_{N\in\mathbb{N}}$ towards $\mu(t)$ for a.e. $t \in [0, T]$. Combining the uniform convergence of the sequence $(a_N)_{N\in\mathbb{N}}$ and the weak convergence of $\mu^N(t)$ we see that the limit

$$\lim_{N\to\infty}\left|\int_{\mathbb{R}^d}\big(F[\widehat{a}_N] - F[a]\big)(x-y)d\mu^N(t)(y)\right|.$$

exists for a.e. $t \in [0, T]$ and for every $x \in \mathbb{R}^d$: indeed, for every $\varepsilon > 0$ you can find $N_0(\varepsilon)$ such that, for all $N \geq N_0(\varepsilon)$ we have

$$\sup_{x,y\in B(0,R)} |F[\widehat{a}_N](x-y) - F[\widehat{a}](x-y)| \leq 2R\|\widehat{a}_N - \widehat{a}\|_{L_\infty(\mathrm{supp}(\rho))} \leq \varepsilon/2,$$

as well as

$$\left|\int_{\mathbb{R}^d}\big(F[\widehat{a}] - F[a]\big)(x-y)d\mu^N(t)(y) - \int_{\mathbb{R}^d}\big(F[\widehat{a}] - F[a]\big)(x-y)d\mu(t)(y)\right| \leq \varepsilon/2.$$

The first estimate follows from (31) and the uniform convergence of the $\widehat{a}_N$, while the second one follows from the continuity of $F[a]$ and $F[\widehat{a}]$ (coming from the continuity of

a and $\widehat{a}$, uniform limit of continuous functions) and the uniform weak convergence of $\mu^N(t)$. Hence for $N \geq N_0(\varepsilon)$ we obtain

$$\begin{aligned}
&\left| \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[a]\big)(x-y) d\mu^N(t)(y) \right| - \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu(t)(y) \right| \right| \\
&\leq \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[a]\big)(x-y) d\mu^N(t)(y) - \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu(t)(y) \right| \\
&\leq \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[\widehat{a}]\big)(x-y) d\mu^N(t)(y) \right| \\
&\qquad + \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu^N(t)(y) - \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu(t)(y) \right| \\
&\leq 2R \|\widehat{a}_N - \widehat{a}\|_{L_\infty(\mathrm{supp}(\rho))} \int_{\mathbb{R}^d} d\mu^N(t)(y) + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

which implies that, for every $t \in [0,T]$ and $x \in \mathbb{R}^d$, it holds

$$\lim_{N\to\infty} \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[a]\big)(x-y) d\mu^N(t)(y) \right|^2 = \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu(t)(y) \right|^2. \tag{32}$$

Notice that in (32) we added a square and that the limit is clearly uniform in t and x.

For a.e. $t \in [0,T]$, we now pass to compute the limit

$$\lim_{N\to\infty} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[a]\big)(x-y) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x).$$

For the sake of compactness we set

$$\begin{aligned}
H_N(t,x) &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[a]\big)(x-y) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x), \\
G_N(t) &= \int_{\mathbb{R}^d} H_N(t,x) d\mu^N(t)(x), \\
H(t,x) &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu(t)(y) \right|^2 d\mu(t)(x), \\
G(t) &= \int_{\mathbb{R}^d} H(t,x) d\mu(t)(x),
\end{aligned}$$

and we estimate

$$|G_N(t) - G(t)| \leq \left| \int_{\mathbb{R}^d} H_N(t,x) d\mu^N(t)(x) - \int_{\mathbb{R}^d} H_N(t,x) d\mu(t)(x) \right|$$

$$+ \int_{\mathbb{R}^d} |H_N(t,x) - H(t,x)| \, d\mu(t)(x).$$

From the fact that the H_N are continuous and bounded, and that the measures $\mu^N(t)$ are weakly converging to $\mu(t)$, it follows that for every $\varepsilon > 0$ we can find $N_0(\varepsilon)$ such that for all $N \geq N_0(\varepsilon)$ it holds

$$\left| \int_{\mathbb{R}^d} H_N(t,x) d\mu^N(t)(x) - \int_{\mathbb{R}^d} H_N(t,x) d\mu(t)(x) \right| \leq \frac{\varepsilon}{2}.$$

From (32) also follows that for all $N \geq N_0(\varepsilon)$ we have

$$|H_N(t,x) - H(t,x)| \leq \frac{\varepsilon}{2},$$

which yields $|G_N(t) - G(t)| \leq \varepsilon$, and thus

$$\lim_{N\to\infty} \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}_N] - F[a]\big)(x-y) d\mu^N(t)(y) \right|^2 d\mu^N(t)(x) =$$

$$\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \big(F[\widehat{a}] - F[a]\big)(x-y) d\mu(t)(y) \right|^2 d\mu(t)(x).$$

To prove that $\lim_{N\to\infty} E_N(\widehat{a}_N) = E(\widehat{a})$, we are simply left to show that

$$\lim_{N\to\infty} \frac{1}{T} \int_0^T G_N(t) dt = \frac{1}{T} \int_0^T G(t) dt.$$

But this follows easily from the dominated convergence theorem and the fact that T is finite, since we can bound the functions G_N uniformly from above using (31) as

$$\begin{aligned} |G_N(t)| &\leq \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \big|(F[\widehat{a}_N] - F[a])(x-y)\big| d\mu^N(t)(y) \right)^2 d\mu^N(t)(x) \\ &\leq 4R^2 \|\widehat{a}_N - a\|^2_{L^\infty(\mathrm{supp}(\rho))} \\ &\leq 4R^2 (B + \|a\|_{L^\infty(\mathrm{supp}(\rho))})^2. \end{aligned}$$

Here B is an appropriate bound for $\|\widehat{a}_N\|^2_{L^\infty(\mathrm{supp}(\rho))}$, which exists since $(\widehat{a}_N)_{N\in\mathbb{N}}$ is a uniformly convergent sequence (and thus bounded). $\square$

Proof of Theorem 4.2. The sequence of minimizers $(\widehat{a}_N)_{N\in\mathbb{N}}$ is by definition a subset of X_M, hence by Proposition 3.7 it admits a subsequence $(\widehat{a}_{N_k})_{k\in\mathbb{N}}$ uniformly converging to a function $\widehat{a} \in X_M$.

To show the optimality of $\widehat{a}$, let $b \in X_M$ be given. By Definition 4.1, we can find a sequence $(b_N)_{N\in\mathbb{N}}$ converging uniformly to b on $\mathrm{supp}(\rho)$ such that $b_N \in V_N$ for every $N \in \mathbb{N}$. Hence, by Lemma 4.3, it holds

$$E(b) = \lim_{N\to\infty} E_N(b_N).$$

Now, by the optimality of $\widehat{a}_{N_k}$ and again by Lemma 4.3, follows that

$$E(b) = \lim_{N\to\infty} E_N(b_N) = \lim_{k\to\infty} E_{N_k}(b_{N_k}) \geq \lim_{k\to\infty} E_{N_k}(\widehat{a}_{N_k}) = E(\widehat{a})\,.$$

We can therefore conclude the fundamental estimate

$$E(b) \geq E(\widehat{a}), \tag{33}$$

which holds for every $b \in X_M$. In particular, (33) applies to $b = a \in X_M$ (by the particular choice of M), which finally implies

$$0 = E(a) \geq E(\widehat{a}) \geq 0 \Longrightarrow E(\widehat{a}) = 0,$$

showing that $\widehat{a}$ is a minimizer of E. In case (29) holds, by Proposition 3.6 follows $\widehat{a} = a$, as desired. □

References

[1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.*, volume 254. Clarendon Press Oxford, 2000.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

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

[4] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5):852–862, 2007.

[5] A. Filippov. *Differential equations with discontinuous right-hand sides.* Kluwer Academic Publishers, 1988.

[6] M. Fornasier and J.-C. Hütter. Consistency of probability measure quantization by means of power repulsion-attraction potentials. Submitted, 2015.

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

[8] J. L. Kelley. *General topology.* Springer-Verlag, 1955.

[9] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and their Applications.* Academic Press, New York, NY, 1980.

[10] J. E. Lennard-Jones. On the determination of molecular fields. *Proc. R. Soc. Lond. A*, 106(738):463—477, 1924.

[11] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4):25–34, 1987.

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

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