Inferring Interaction Rules from Observations of Evolutive Systems I: The Variational Approach

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

Abstract

In this paper we are concerned with the learnability of nonlocal interaction kernels for first order systems modeling certain social interactions, from observations of realizations of the dynamics. This paper is the first of a series on learnability of nonlocal interaction kernels and presents a variational approach to the problem. In particular, we assume here that the kernel to be learned is bounded and locally Lipschitz continuous and the initial conditions of the systems are drawn identically and independently at random according to a given initial probability distribution. Then the minimization over a rather arbitrary sequence of (finite dimensional) subspaces of a least square functional measuring the discrepancy from observed trajectories produces uniform approximations to the kernel on compact sets. The convergence result is obtained by combining mean-field limits, transport methods, and a Γ-convergence argument. A crucial condition for the learnability is a certain coercivity property of the least square functional, majoring an L_2-norm discrepancy to the kernel with respect to a probability measure, depending on the given initial probability distribution by suitable push forwards and transport maps. We illustrate the convergence result by a few numerical experiments.

Keywords: interaction kernel learning, first order nonlocal systems, mean-field equations, Γ-convergence

1 Introduction

What are the instinctive individual reactions which make a group of animals forming coordinated movements, for instance a flock of migrating brids or a school of fish? Which biological interactions between cells produce the formation of complex structures, for instance organs? What are the mechanisms which induce certain significant changes in a large amount of players in the financial market? In this paper we are concerned with the "mathematization" of the problem of learning or inferring interaction rules from observations of evolutions. The framework we consider is the one of evolutions driven by gradient descents. The study of gradient flow evolutions to minimize certain energetic landscapes has been the subject of intensive research in the past years [**?**]. Some of the most recent models are aiming at describing time-dependent phenomena also in biology or even in social dynamics, borrowing a leaf from more established and classical models in physics. For instance, starting with the seminal papers of Vicsek et. al. [**?**] and Cucker-Smale [**?**], there has been a flood of models describing consensus or opinion formation, modeling the exchange of information as long-range social interactions (forces) between active agents (particles). However, for the analysis, but even more crucially for the reliable and realistic numerical simulation of such phenomena, one presupposes a complete understanding and determination of the governing energies. Unfortunately, except for physical situations where the calibration of the model can be done by measuring the governing forces rather precisely, for some relevant macroscopical models in physics and most of the models in biology and social sciences the governing energies are far from being precisely determined. In fact, very often in these studies the governing energies are just predetermined to be able to reproduce, at least approximately or qualitatively, some of the macroscopical effects of the observed dynamics, such as the formation of certain patterns, but there has been little or no effort of matching data from real-life cases.

This attitude aiming just at a qualitative description tends however to reduce some of the investigations in this area to beautiful and mathematically interesting toy-cases, which have likely little to do with real-life scenarios. The aim of this paper is providing a mathematical framework for the reliable identification of the governing energies from data obtained by direct observations of corresponding time-dependent evolutions. This is a new kind of inverse problem, beyond more traditionally considered ones, as the forward map is a strongly nonlinear evolution, highly dependent on the probability measure generating the initial conditions. As we aim at a precise quantitative analysis, and to be very concrete, we will attack the learning of the energies for specific models in social dynamics governed by nonlocal interactions.

1.1 General abstract framework

Many time-dependent phenomena in physics, biology, and social sciences can be modelled by a function $x : [0,T] \to \mathcal{H}$, where $\mathcal{H}$ represents the space of states of the physical, biological or social system, which evolves from an initial configuration $x(0) = x_0$ towards a more convenient state or a new equilibrium. The space $\mathcal{H}$ can conveniently be a Banach space or just a metric space. This implicitly assumes that x evolves driven by a minimization process of a potential energy $\mathcal{J} : \mathcal{H} \times [0,T] \to \mathbb{R}$. In this preliminary introduction we consciously avoid specific assumptions on $\mathcal{J}$, as we wish to keep a rather general view. We restrict the presentation to particular cases below.

Inspired by physics for which conservative forces are the derivatives of the potential energies, one can describe the evolution as satisfying a gradient flow inclusion of the type

$$\dot{x}(t) \in -\partial_x \mathcal{J}(x(t), t), \tag{1}$$

where $\partial_x \mathcal{J}(x,t)$ is some notion of differential of $\mathcal{J}$ with respect to x, which might already take into consideration additional constraints which are binding the states to a certain sets.

1.2 Example of gradient flow of nonlocal particle interactions

Assume that $x = (x_1, \dots, x_N) \in \mathcal{H} = \mathbb{R}^{d \times N}$ and that

$$\mathcal{J}_N(x) = \frac{1}{2N} \sum_{i,j=1}^{N} A(|x_i - x_j|),$$

where $A : \mathbb{R}_+ \to \mathbb{R}$ is a suitable nonlinear interaction kernel function, which, for simplicity we assume to be smooth, and $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$. Then, the formal unconstrained gradient flow (1) associated to this energy is written coordinatewise

$$\dot{x}_i(t) = \frac{1}{N} \sum_{j \neq i} \frac{A'(|x_i - x_j|)}{|x_i - x_j|}(x_j - x_i), \quad i = 1, \dots, N. \tag{2}$$

Under suitable assumptions of local Lipschitz continuity and boundedness of

$$a(\cdot) = \frac{A'(|\cdot|)}{|\cdot|}, \tag{3}$$

this evolution is well-posed for any given $x(0) = x_0$ and it is expected to converge for $t \to \infty$ to configurations of the points whose mutual distances are close to local minimizers of the function A, representing steady states of the evolution as well as critical points of $\mathcal{J}_N$.
It is also well-known [**?**] (see also Proposition 2.2 below) that for $N \to \infty$ a mean-field approximation holds, i.e., if the initial conditions $x_i(0)$ are distributed accoding to a compactly supported probability measure $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ for $i = 1, 2, 3, \dots$, the empirical measure $\mu^N(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i(t)}$ weakly converges for $N \to \infty$ to the probability measure valued trajectory $t \to \mu(t)$ satisfying weakly the equation

$$\partial_t \mu(t) = -\nabla \cdot ((F[a] * \mu(t))\mu(t)), \quad \mu(0) = \mu_0. \tag{4}$$

where $F[a](x) = -a(|x|)x, x \in \mathbb{R}^d$. Actually, the differential equation (4) corresponds again to a gradient flow of an energy

$$\mathcal{J}(\mu) = \int_{\mathbb{R}^{d \times d}} A(|x - y|) d\mu(x) d\mu(y),$$

on the metric space $\mathcal{H} = \mathcal{P}_c(\mathbb{R}^d)$ endowed with the so-called Wasserstein distance. Continuity equations of the type (4) with nonlocal interaction kernels are currently the subject of intensive research towards the modeling of the biological and social behavior of microorganisms, animals, humans, etc. We refer to the articles [**?**, **?**] for recent overviews on this subject. Despite the tremendous theoretical success of such research direction in terms of mathematical results, as we shall stress below in more detail, one of the issues which is so far scarcely addressed in the study of models of the type (2) or (4) is their actual applicability. Most of the results are addressing a purely *qualitative analysis* given certain smoothness and asymptotic properties of the kernels A or a at the origin or at infinity, in terms of well-posedness or in terms of asymptotic behavior of the solution for $t \to \infty$. Certainly such results are of great importance, as such interaction functions, if ever they can really describe social dynamics, are likely to differ significantly from well-known models from physics and it is reasonable and legitimate to consider a large variety of classes of such functions. However, a solid mathematical framework which establishes the conditions of "learnability" of the interaction kernels from observations of the dynamics is currently not available and it will be the main subject of this paper.

1.3 Parametric energies and their identifications

Let us now consider an energy $\mathcal{J}[a]$ depending on a parameter function a. As in the example mentioned above, a may be defining a nonlocal interaction kernel as in (3). The parameter function a not only determines the energy, but also the corresponding evolutions driven according to (1), for fixed initial conditions $x(0) = x_0$. (Here we assume that the class of a is such that the evolutions exist and they are essentially well-posed.) The fundamental question to be addressed is: can we recover a with high accuracy given some observations of the realized evolutions? This question is prone to several specifications, for instance, we may want to assume that the initial conditions are generated according to a certain probability distribution or they are chosen deterministically ad hoc to determine at best a, that the observations are complete or incomplete, etc. As one quickly realizes, this is a very broad field to explore with many possible developments. Surprisingly, there are no results in this direction at this level of generality, and very little is done in the specific directions we mentioned in the example above. We may refer for instance to [**?**, **?**] for studies on the inference of social rules in collective behavior.

1.4 The optimal control approach and its drawbacks

Let us introduce an approach, which would perhaps naturally be considered at a first instance and focus for a moment on the gradient flow model (1). Given a certain gradient flow evolution $t \to x[a](t)$ depending on the unknown parameter function a, one might decide to design the recovery of a as an optimal control problem [**?**]: for instance, we may seek for a parameter function $\widehat{a}$ which minimizes

$$\mathcal{E}(\widehat{a}) = \frac{1}{T} \int_0^T \left[\mathrm{dist}_{\mathcal{H}}(x[a](s) - x[\widehat{a}](s))^2 + \mathcal{R}(\widehat{a}) \right] ds, \tag{5}$$

being $t \to x[\widehat{a}](t)$ the solution of gradient flow (1) for $\mathcal{J} = \mathcal{J}[\widehat{a}]$, i.e.,

$$\dot{x}[\widehat{a}](t) \in -\partial_x J(x[\widehat{a}](t), t), \tag{6}$$

and $\mathcal{R}(\cdot)$ is a suitable regularization function, which restricts the possible minimizers of (5) to a specific class. The first fundamental problem one immediately encounters with this formulation is the strongly nonlinear dependency of $t \to x[\widehat{a}](t)$ on $\widehat{a}$, which eventually results in a strong nonconvexity of the functional (5). This also implies that a direct minimization of (5) would most likely risk to end up on suboptimal solutions, and even the computation of a first order optimality condition in terms of Pontryagin's minimum principle would not characterize uniquely the minimal solutions. On top of this essential problem, the numerical implementation of both strategies, direct optimization or solution of the first order optimality conditions, is expected to be computationally infeasible as soon as the underlying discretization dimension grows to meet good approximation accuracies.

1.5 A variational approach towards learning parameter functions in nonlocal energies

Let us consider the framework of the example in Section 1.2. We restrict our attention to interaction kernels a 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. Our goal is to learn a target function $a \in X$ (which we fix throughout the section) from the observation of the dynamics of the empirical measure μ^N that is defined by $\mu^N(t) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i(t)}$, where $x_i(t)$ are solving

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j \neq i} a(|x_i - x_j|)(x_j - x_i), \quad i = 1, \ldots, N. \tag{7}$$

with a as the interaction kernel. We already mentioned that the most immediate approach of formulating the learning of a in terms of an optimal control problem is not efficient and likely will not give optimal solutions due to the strong nonconvexity. We propose instead a rather novel approach which turns out to be both computationally very efficient and produces optimal solutions under reasonable assumptions. As an approximation to a we seek for a minimizer of the following *discrete error functional*

$$\mathcal{E}_N(\widehat{a}) = \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\sum_{j=1}^N \left(\widehat{a}(|x_i(t) - x_j(t)|)(x_i(t) - x_j(t)) - \dot{x}_i(t)\right)\right|^2 dt, \tag{8}$$

among all functions $\widehat{a} \in X$. As we show in Proposition 3.1, if both $a, \widehat{a} \in X$ then there exist a constant $C > 0$ depending on $T, \widehat{a}, \mu_0^N$ and a compact set $K \subset \mathbb{R}_+$ such that

$$\|x[a](t) - x[a](t)\| \leq C(\|\widehat{a}\|_\infty + \mathrm{Lip}_K(\widehat{a}))\sqrt{\mathcal{E}_N(\widehat{a})},$$

for all $t \in [0,T]$. (Here $\|x\| = \frac{1}{N}\sum_{i=1}^N |x_i|$, for $x \in \mathbb{R}^{d \times N}$.) Hence, minimizing $\mathcal{E}_N(\widehat{a})$ with respect to $\widehat{a}$ in turns implies an accurate approximation of the trajectory $t \to x[a](t)$ at finite time as well. Moreover, notice that, contrary to the optimal control approach, the functional $\mathcal{E}_N$ has the remarkable property of being easily computable from the knowledge of x_i and $\dot{x}_i$. Of course, we can consider to approximate the time derivative $\dot{x}_i$ by finite differences $\frac{x_i(t+\delta t) - x_i(t)}{\delta t}$ and we can assume that the data of the problem can be fully defined on observations of $x_i(t)$ for $t \in [0,T]$ for a prescribed finite time horizon $T > 0$. Moreover, being a simple quadratic functional, its minimizers can be efficiently numerically approximated on a finite element space. In particular, given a finite dimensional space $V \subset X$, we consider the minimizer:

$$\widehat{a}_{N,V} = \arg\min_{\widehat{a}\in V} \mathcal{E}_N(\widehat{a}). \tag{9}$$

The fundamental question to be addressed in this paper is

(Q) For which choice of the approximating spaces $V \in \Lambda$ (we assume here that Λ is a countable family of invading subspaces of X) does $\widehat{a}_{N,V} \to a$ for $N \to \infty$ and $V \to X$ and in which topology should this convergence hold?

We show now how we are intending to address this issue in detail by a variational approach. For that let us conveniently rewrite the functional (8) as follows

$$\begin{aligned}\mathcal{E}_N(\widehat{a}) &= \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\sum_{j=1}^N \big(F[\widehat{a}] - F[a]\big)(x_i - x_j)\right|^2 dt \\ &= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \big(F[\widehat{a}] - F[a]\big) * \mu^N(t)\right|^2 d\mu^N(t)(x)dt,\end{aligned} \tag{10}$$

where $F[a](x) = -a(|x|)x, x \in \mathbb{R}^d$. As we are searching for limits of a sequence of minimizers $(\widehat{a}_{N,V})_{N\in\mathbb{N},V\in\Lambda}$, it is natural to consider techniques of Γ-convergence [**?**], whose general aim is establishing the convergence of minimizers for a sequence of equi-coercive functionals Γ-converging to a target functional.

The form (10) of $\mathcal{E}_N$ clearly suggests a specific candidate for a Γ-limit functional:

$$\mathcal{E}(\widehat{a}) = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \big(F[\widehat{a}] - F[a]\big) * \mu(t)\right|^2 d\mu(t)(x)dt, \tag{11}$$

where μ is a weak solution to the mean-field equation (2), as soon as the initial conditions $x_i(0)$ are identically and independently distributed according to a compactly supported probability measure $\mu(0) = \mu_0$. Several issues need to be addressed at this point. The first one is to establish the space where a result of Γ-convergence is supposed to hold. Surprisingly enough we expect that such a space is *not* independent of the initial probability measure μ_0! We clarify this issue immediately, but let us stress that this feature of the problem makes it of particular interest and novelty. We observe that by Jensen inequality

$$\mathcal{E}(\widehat{a}) \leq \frac{1}{T}\int_0^T \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\widehat{a}(|x-y|) - a(|x-y|)|^2 |x-y|^2 d\mu(t)(x)d\mu(t)(y)dt. \tag{12}$$

Using the Euclidean 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 $\varrho : [0,T] \to \mathcal{P}_1(\mathbb{R}_+)$ 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 rewrite the estimate (12),

$$\mathcal{E}(\widehat{a}) \leq \frac{1}{T} \int_0^T \int_{\mathbb{R}_+} \big|\widehat{a}(s) - a(s)\big|^2 d\varrho(t)(s) dt. \tag{13}$$

For every open set $A \subseteq \mathbb{R}_+$ the mapping $t \in [0,T] \mapsto \varrho(t)(A)$ is actually lower semi-continuous, whereas for any compact set A it is upper semi-continuous (see Lemma 3.2). This allows us to define a probability measure $\widehat{\rho}$ on the Borel σ-algebra on $\mathbb{R}_+$: For any open set $A \subseteq \mathbb{R}_+$ we define

$$\widehat{\rho}(A) := \frac{1}{T} \int_0^T \varrho(t)(A) dt, \tag{14}$$

and extend this set function to a probability measure on all Borel sets. Eventually we define

$$\rho(A) := \int_A s^2 d\widehat{\rho}(s),$$

for all $A \subseteq \mathbb{R}_+$ Borel sets. Then one can reformulate (13) as follows

$$\mathcal{E}(\widehat{a}) \leq \int_{\mathbb{R}_+} \big|\widehat{a}(s) - a(s)\big|^2 d\rho(s) = \|\widehat{a} - a\|^2_{L_2(\mathbb{R}_+,\rho)}. \tag{15}$$

Notice that ρ is defined through $\mu(t)$ which in turn is depending on the initial probability measure μ_0. To establish coercivity of the learning problem it is natural to assume that there exists $c_T > 0$ such that the following bound holds

$$c_T \|\widehat{a} - a\|^2_{L_2(\mathbb{R}_+,\rho)} \leq \mathcal{E}(\widehat{a}), \tag{16}$$

for all relevant $\widehat{a} \in X \cap L_2(\mathbb{R}_+, \rho)$. This crucial assumption eventually fixes also the natural space $X \cap L_2(\mathbb{R}_+, \rho)$ for the solutions and, as mentioned above, it depends on the choice of the initial conditions μ_0. In particular the constant $c_T \geq 0$ might not be nondegenerate for all the choices of μ_0 and one has to pick the right initial distribution so that (16) can hold for $c_T > 0$. In Section 3.2 we show that for some specific choices of a and rather arbitrary choices of $\widehat{a} \in X$ one can construct probability measure valued trajectories $t \to \mu(t)$ which allow to validate (16).

Notice also that (16) implies that a is the unique minimizer of $\mathcal{E}$ in $X \cap L_2(\mathbb{R}_+, \rho)$, which is an important condition for the learnability via variational limit. Let us now state the main result of this paper, Theorem 4.2, on the learnability problem by means of a variational approach: Fix

$$M \geq \|a\|_{L_\infty(K)} + \|a'\|_{L_\infty(K)},$$

and define the set

$$X_{M,K} = \left\{ b \in W^1_\infty(K) : \|b\|_{L_\infty(K)} + \|b'\|_{L_\infty(K)} \le M \right\}$$

for a suitable interval $K = [0, 2R]$, for $R > 0$ large enough for $\mathrm{supp}(\rho) \subset K$. Additionally for every $N \in \mathbb{N}$, let V_N be a closed subset of $X_{M,K}$ w.r.t. the uniform convergence on K with the following *uniform approximation property*: for all $b \in X_{M,K}$ there exists a sequence $(b_N)_{N\in\mathbb{N}}$ converging uniformly to b on K and such that $b_N \in V_N$ for every $N \in \mathbb{N}$. Then the minimizers

$$\widehat{a}_N \in \arg\min_{\widehat{a}\in V_N} \mathcal{E}_N(\widehat{a}). \tag{17}$$

converge uniformly to some continuous function $\widehat{a} \in X_{M,K}$ such that $\mathcal{E}(\widehat{a}) = 0$. If we additionally assumed the coercivity condition (16), then we could eventually conclude that $\widehat{a} = a$ in $L_2(\mathbb{R}_+, \rho)$. This would be our first instance of an answer to question (Q). The proof of such a statement is addressed by exploiting the uniform relative compactness of $X_{M,K}$ and the fact that uniform convergence matches well with the weak convergence of the measures $\mu_N \rightharpoonup \mu$.

1.6 Numerical implementation of the variational approach

The strength of the result from the variational approach followed in Section 1.5 is the total arbitrariness of the sequence V_N except for the assumed *uniform approximation property* and that the result is to hold - deterministically - with uniform convergence (hence a very strong convergence). Nevertheless such a great result may hide a horrible truth: the promise of an efficient and easy learning of a by solving (17) might have been badly broken! In fact the spaces V_N are supposed to be picked as subsets of $X_{M,K}$ and one presupposes to have knowledge of $M \ge \|a\|_{L_\infty(K)} + \|a'\|_{L_\infty(K)}$. First of all this means that we need to know a priori a lot about the solution itself. Secondly, the finite dimensional optimization (17) is not anymore a simple *unconstrained* least squares (as claimed in (9)), but hiddenly it requires a uniform bound on both the solution and its gradient! It is clearly very interesting to understand how severe these two drawbacks are. In particular, what happens if one picks initially a wrong possible estimate for the constant $M > 0$? Can one choose M very large for N moderately small and then tune it down to the "right" level adaptively, for N growing? How can one implement efficiently such a numerical optimization (17) with L_∞ constraints? We address these aspects in Section 5 with several numerical experiments ... etc

We conclude this introduction by mentioning that the variational approach is based on a compactness argument and, as a consequence, it does not provide any rate of convergence. This is another significant drawback of this technique.

Although we expect these mentioned drawbacks to be only "mildly severe", one may wonder anyway whether we can relax some aspects of the approximation strategy developed in Section 1.5 which can lead to a more practical and efficient algorithmic solution with guaranteed rates of convergence. In our follow-up paper [**?**] we are following the approach developed by DeVore et al. in [**?**, **?**] towards universal algorithms for learning regression functions from independent samples drawn according to an unknown probability distribution. To the price of a weaker approximation to hold only with high probability and of a constrained choice of the approximating spaces V_N, we obtain that for every $\beta > 0$

$$\mathcal{P}_{\mu_0}\Big(\|a - \widehat{a}_{N,V_N}\|_{L_2(\mathbb{R}_+,\rho)} > (c_3\|a\|_\infty + |a|_{\mathcal{A}^s_\mu})\big(\tfrac{\log N}{N^3}\big)^{\frac{s}{2s+1}} \Big) \leq c_4 N^{-\beta}\,, \tag{18}$$

if c_3 is chosen sufficiently large (depending on β), where $\mathcal{A}^s_\mu$ is a suitable functional class indicating how efficiently a can be approximated by piecewise polynomial functions in $L_2(\mathbb{R}_+, \rho)$. Let us now present the details of the variational approach.

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 $\mu \in \mathcal{P}(\mathbb{R}^{n_1)}$ 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*,

$$\mathcal{W}_p^p(\mu, \nu) = \inf\left\{ \int_{\mathbb{R}^{2n}} |x - y|^p d\rho(x, y) : \rho \in \Gamma(\mu, \nu) \right\}. \tag{19}$$

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

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

[1]We follow the notation of [**?**].

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 (19) is actually a minimum. For more details, see e.g. [**?**, **?**].

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.,

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

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

2.1 The mean-field limit equation and existence of solutions

As already stated in the introduction, our learning approach is based on the following underlying *finite time horizon initial value problem*: given $T > 0$ and $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$, consider a probability measure valued trajectory $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ satisfying

$$\begin{cases} \dfrac{\partial \mu}{\partial t}(t) = -\nabla \cdot ((F[a] * \mu(t))\mu(t)) & \text{for } t \in (0,T], \\ \mu(0) = \mu_0. \end{cases} \tag{20}$$

We consequently give our notion of solution for (20).

Definition 2.1. We say that a map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ is a solution of (20) with initial datum μ_0 if the following hold:

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

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

3. μ satisfies (20) in the weak sense, i.e., for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R})$ it holds
$$\frac{d}{dt}\int_{\mathbb{R}^d} \phi(x) d\mu(t)(x) = \int_{\mathbb{R}^d} \nabla\phi(x) \cdot (F[a] * \mu(t))(x) d\mu(t)(x).$$

As we recall in a moment, system (20) is closely related to the family of ODEs indexed by $N \in \mathbb{N}$ given by

$$\begin{cases} \dot{x}_i^N(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^N F[a](x_i^N(t) - x_j^N(t)) & \text{for } t \in (0,T], \\ x_i^N(0) = x_{0,i}^N, \end{cases} \qquad i = 1,\dots,N, \tag{21}$$

which can be more conveniently rewritten as follows

$$\begin{cases} \dot{x}_i^N(t) = (F[a] * \mu^N(t))(x_i^N(t)) & \text{for } t \in (0,T], \\ x_i^N(0) = x_{0,i}^N, \end{cases} \qquad i = 1,\dots,N, \tag{22}$$

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

$$\mu^N(t) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i^N(t)}. \tag{23}$$

In the followingas we restrict our attention to interaction kernels belonging to the following *set of admissible kernels*

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

The well-posedness of (22) is rather standard under the assumption $a \in X$ and for reader's convenience we report its proof in Appendix 6.3. The well-posedness of system (20) and several crucial properties that it enjoys can be proved as soon as $a \in X$ as well. We may refer to [**?**] for by now standard results on existence and uniqueness of solutions for (20) and to [**?**] for generalizations in case of interaction kernels not belonging to the class X. In the following we report the main results, whose proofs are collected in the Appendix for the sake of self-containedness and to allow explicit reference to constants.

Proposition 2.2. *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*

$$\mu_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{x_{0,i}^N}, \qquad \textit{for some } x_{0,i}^N \in \operatorname{supp}(\mu_0) + \overline{B(0,1)}$$

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 (23) where $(x_1^N, \dots, x_N^N)$ is the unique solution of system (21).

Then, the sequence $(\mu^N)_{N\in\mathbb{N}}$ converges, up to subsequences, to a solution μ of (20) with initial datum μ_0. Moreover, there exists $R > 0$ depending only on $T, a,$ and $\operatorname{supp}(\mu_0)$ such that it holds

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

A proof of this standard result is reported in Appendix 6.2 together with the necessary technical lemmas in Appendix 6.1.

2.2 The transport map and uniqueness of mean-field solutions

Another way for building a solution of equation (20) 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 (22),

$$\begin{cases} \dot{\xi}(t) = (F[a] * \mu(t))(\xi(t)) & \text{for } t \in (0,T], \\ \xi(0) = \xi_0, \end{cases} \tag{24}$$

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 (48) from the choice of T, a and μ_0.

We consider now 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 (24) with initial datum ξ_0. The by now well-known result [**?**, Theorem 3.10] shows that the solution of (20) with initial value μ_0 is the unique fixed-point of the *push-foward map*

$$\Gamma[\mu](t) := (\mathcal{T}_t^\mu)_\# \mu_0. \tag{25}$$

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

Proposition 2.3. *$\mathcal{T}_t^\mu$ 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 6.8 and the inequality (55) trivially implies the following stability estimate

$$\left|\mathcal{T}_t^\mu(x_0) - \mathcal{T}_t^\mu(x_1)\right| \leq e^{T \operatorname{Lip}_{B(0,R)}(F[a])}|x_0 - x_1|, \quad \text{for } |x_i| \leq R\,, \quad i = 0,1\,. \tag{26}$$

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

In view of the uniqueness of the solutions to the ODE (24), 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 result shows uniqueness and continuous dependence on the initial data for (20). Although it is a well-known result, for the sake of self-containedness and to have explicit reference to constants, we report a proof of it in Appendix 6.4.

Theorem 2.4. *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* (20)*, for $\mu(0) = \mu_0$ and $\nu(0) = \nu_0$ respectively. Consider $R > 0$ such that*

$$\mathrm{supp}(\mu(t)) \cup \mathrm{supp}(\nu(t)) \subseteq B(0,R) \tag{27}$$

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

$$\mathcal{W}_1(\mu(t),\nu(t)) \leq \overline{C}\,\mathcal{W}_1(\mu_0,\nu_0) \tag{28}$$

for every $t \in [0,T]$. In particular, equi-compactly supported solutions of (20) *are uniquely determined by the initial datum.*

3 The learning problem for the kernel function

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 of μ^N that stems from system (21) with a as interaction kernel, μ_0^N as initial datum and T as finite time horizon.

A very reasonable and intuitive strategy would be to pick $\widehat{a} \approx a$ among those functions in X which would give rise to a dynamics close to μ^N, i.e., choose $\widehat{a} \in X$ as a minimizer of the following *discrete error functional*

$$\mathcal{E}_N(\widehat{a}) = \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\sum_{j=1}^N \left(\widehat{a}(|x_i^N(t) - x_j^N(t)|)(x_i^N(t) - x_j^N(t)) - \dot{x}_i^N(t)\right)\right|^2 dt. \tag{29}$$

Proposition 3.1. *If $a, \widehat{a} \in X$ then there exist a constant $C > 0$ depending on $T, \widehat{a}$ and μ_0^N and a compact set $K \subset \mathbb{R}_+$ such that*

$$\|x[a](t) - x[\widehat{a}](t)\| \leq C(\|\widehat{a}\|_\infty + \mathrm{Lip}_K(\widehat{a}))\sqrt{\mathcal{E}_N(\widehat{a})}, \tag{30}$$

for all $t \in [0,T]$, and $x[a], x[\widehat{a}]$ are the solutions to (21) *for the interaction kernels a and $\widehat{a}$ respectively. (Here $\|x\| = \frac{1}{N}\sum_{i=1}^N |x_i|$, for $x \in \mathbb{R}^{d\times N}$.)*

Proof. Let us denote $x = x[a]$ and $\widehat{x} = x[\widehat{a}]$ and we estimate by Jensen or Hölder inequalities

$$\|x(t) - \widehat{x}(t)\|^2 = \left\|\int_0^t (\dot{x}(s) - \dot{\widehat{x}}(s))ds\right\|^2 \leq t\int_0^t \|\dot{x}(s) - \dot{\widehat{x}}(s)\|^2 ds$$

$$= \quad t\int_0^t \frac{1}{N}\sum_{i=1}^N \left|(F[a]*\mu^N(x_i) - F[\widehat{a}]*\widehat{\mu}^N(\widehat{x}_i))\right|^2 ds$$

$$
\begin{aligned}
&\leq \quad 2t\int_0^t \left[\frac{1}{N}\sum_{i=1}^N \left|(F[a]-F[\widehat{a}])*\mu^N(x_i))\right|^2 + \left|\frac{1}{N}\sum_{j=1}^N \widehat{a}(|x_i-x_j|)((\widehat{x}_j-x_j)+(x_i-\widehat{x}_i))\right.\right.\\
&+ \quad \left.\left.\widehat{a}(|\widehat{x}_i-\widehat{x}_j|)((\widehat{x}_j-\widehat{x}_i)-\widehat{a}(|x_i-x_j|)((\widehat{x}_j-\widehat{x}_i)\right|^2\right]ds\\
&\leq \quad 2T^2\mathcal{E}_N(\widehat{a})+\int_0^t 8T(\|\widehat{a}\|^2_{L_\infty(K)}+(R\operatorname{Lip}_K(\widehat{a}))^2)\|x(s)-\widehat{x}(s)\|^2ds,
\end{aligned}
$$

for $K=[0,2R]$ and $R>0$ is as in Proposition 2.2 for a substituted by $\widehat{a}$. An application of Gronwall's inequality yields the estimate

$$
\|x(t)-\widehat{x}(t)\|^2 \leq 2T^2(8T(\|\widehat{a}\|^2_{L_\infty(K)}+(R\operatorname{Lip}_K(\widehat{a}))^2))\mathcal{E}_N(\widehat{a})t \leq 16T^2(\|\widehat{a}\|^2_{L_\infty(K)}+(R\operatorname{Lip}_K(\widehat{a}))^2))\mathcal{E}_N(\widehat{a})
$$

or, more simply

$$
\|x[a](t)-x[\widehat{a}](t)\| \leq C(\|\widehat{a}\|_\infty+\operatorname{Lip}_K(\widehat{a}))\sqrt{\mathcal{E}_N(\widehat{a})}.
$$

□

As already mentioned in the introduction, for $N\to\infty$ a natural mean-field approximation to the learning problem can be stated in terms of the following functional,

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

where $\mu(t)$ is a weak solution to (20).

We clarify in the following in detail how the coercivity assumption (16) arises.

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

trivially holds and it follows

$$
\mathcal{E}(\widehat{a}) \leq \frac{1}{T}\int_0^T\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\left|\widehat{a}(|x-y|)-a(|x-y|)|x-y|\right|d\mu(t)(y)\right)^2 d\mu(t)(x)dt.
$$

Now observe that for any $\nu\in\mathcal{P}_1(\mathbb{R}^d)$ and by Hölder's inequality we have

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

Hence, $\mathcal{E}$ can be bounded from above as

$$
\mathcal{E}(\widehat{a}) \leq \frac{1}{T}\int_0^T\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left|\widehat{a}(|x-y|)-a(|x-y|)\right|^2|x-y|^2 d\mu(t)(y)d\mu(t)(x)dt.
$$

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 rewrite the estimate as follows

$$\mathcal{E}(\widehat{a}) \leq \frac{1}{T}\int_0^T \int_{\mathbb{R}_+} \big|\widehat{a}(s) - a(s)\big|^2 s^2 d\varrho(t)(s)dt. \tag{31}$$

In order to rigorously derive (16), we need to explore finer properties of the family of measures $(\varrho(t))_{t\in[0,T]}$.

3.1 The measure ρ

Lemma 3.2. *For every open set $A \subseteq \mathbb{R}_+$ the mapping $t \in [0,T] \mapsto \varrho(t)(A)$ 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. [**?**, Proposition 1.62]. ☐

Lemma 3.2 justifies the following

Definition 3.3. The probability measure $\widehat{\rho}$ on the Borel σ-algebra on $\mathbb{R}_+$ is defined for any Borel set $A \subseteq \mathbb{R}_+$ as follows

$$\widehat{\rho}(A) := \frac{1}{T}\int_0^T \varrho(t)(A)dt. \tag{32}$$

Notice that Lemma 3.2 shows that (32) is well-defined only for sets A that are open or compact in $\mathbb{R}_+$. This directly implies that $\widehat{\rho}$ 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 3.2 we infer that for any Borel set A it holds

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

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

The measure $\widehat{\rho}$ has a deep relationship with our learning process: it tells us which regions of $\mathbb{R}_+$ (the set of distances) were 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 $\widehat{\rho}$ w.r.t. the Lebesgue measure on $\mathbb{R}_+$.

Lemma 3.4. *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 2.3 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 (48), since $\mu(t)(B \setminus B(0, R)) = 0$ for all $t \in [0, T]$ by Proposition 2.2. 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 $B_1, B_2, \ldots$ (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 $\mathcal{L}_d((\mathcal{T}_t^\mu)^{-1}(A)) = 0$. □

Lemma 3.5. *Let μ_0 be absolutely continuous w.r.t. $\mathcal{L}_d$. Then, for all $t \in [0,T]$, the measures $\varrho(t)$ and $\widehat{\rho}$ are absolutely continuous w.r.t. $\mathcal{L}_{1 \llcorner \mathbb{R}_+}$.*

Proof. Fix $t \in [0,T]$. By Lemma 3.4 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 $\widehat{\rho}$ now follows immediately from the one of $\varrho(t)$ for every t and its definition as an integral average (32). □

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 $\widehat{\rho}$.

Lemma 3.6. *The measure $\widehat{\rho}$ is finite and has compact support.*

Proof. To show that ρ is finite, we compute

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

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

$$K = \{|x-y| : x, y \in B(0,R)\} = [0, 2R],$$

where R is given by (48). By construction we have also $\operatorname{supp} \widehat{\rho} \subseteq K$. $\square$

Remark 1. While absolute continuity of μ_0 implies the same for $\widehat{\rho}$, the situation is different for purely atomic measures μ_0^N. On the one hand, also $\mu^N(t)$ is then purely atomic for every t, and this remains true for $\varrho^N(t) = d_{\#}(\mu^N(t) \oplus \mu^N(t))$. However, due to the averaging (32) involved in the definition of $\widehat{\rho}^N = \frac{1}{T}\int_0^T \varrho^N(t)dt$, 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 assumption

With the measure $\widehat{\rho}$ at disposal we define

$$\rho(A) = \int_A s^2 d\widehat{\rho}(s), \tag{33}$$

for all Borel sets $A \subset \mathbb{R}_+$. By means of ρ, we can continue equivalently to estimate $\mathcal{E}$ from (31) as follows,

$$\begin{aligned} \mathcal{E}(\widehat{a}) &\leq \frac{1}{T}\int_0^T \int_{\mathbb{R}_+} \big|\widehat{a}(s) - a(s)\big|^2 s^2 d\varrho(t)(s)dt \\ &= \int_{\mathbb{R}_+} \big|\widehat{a}(s) - a(s)\big|^2 d\rho(s) \\ &= \|\widehat{a} - a\|^2_{L_2(\mathbb{R}_+,\rho)}. \end{aligned} \tag{34}$$

Equation (34) thus suggests the following additional condition to impose in order to ensure that a is the unique minimizer of $\mathcal{E}$: we assume that there exists a constant $c_T > 0$ such that

$$\mathcal{E}(\widehat{a}) \geq c_T \|\widehat{a} - a\|^2_{L_2(\mathbb{R}_+,\rho)}. \tag{35}$$

In the following we clarify that this assumption is in principle verifiable, at least for certain classes of dynamical systems, although we do not furnish a complete characterization of them yet. In particular, we show examples of certain a which generates very special trajectories $t \to \mu(t)$ for which the coercivity condition (35) holds.

We start with the simple case of two particles, i.e., $N = 2$, for which no specific assumptions on $a, \widehat{a}$ are required to verify (35) other than their boundedness in 0. It is also convenient to write $\mathcal{K}(r) = (a(r) - \widehat{a}(r))r$ so that the coercivity condition in this case can be formulated as

$$\frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\sum_{j=1}^N \mathcal{K}(|x_i - x_j|)\frac{x_i - x_j}{|x_i - x_j|}\right|^2 dt \geq \frac{c_T}{N^2 T}\int_0^T \sum_{i=1}^N\sum_{j=1}^N |\mathcal{K}(|x_i - x_j|)|^2 dt. \tag{36}$$

Now, let us observe more closely the integrand on the left-hand-side and for $\widehat{i} \neq i$, $i, \widehat{i} \in \{1, 2\}$ we obtain

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^2 \left| \frac{1}{2}\sum_{j=1}^2 \mathcal{K}(|x_i - x_j|)\frac{x_i - x_j}{|x_i - x_j|}\right|^2 &= \frac{1}{2}\sum_{i=1}^2 \left| \frac{1}{2}\sum_{j\neq i}^2 \mathcal{K}(|x_i - x_j|)\frac{x_i - x_j}{|x_i - x_j|}\right|^2 \\
&= \frac{1}{4}\sum_{i=1}^2 \left| \mathcal{K}(|x_i - x_{\widehat{i}}|)\frac{x_i - x_{\widehat{i}}}{|x_i - x_{\widehat{i}}|}\right|^2 \\
&= \frac{1}{4}\sum_{i=1}^2 \left| \mathcal{K}(|x_i - x_{\widehat{i}}|)\right|^2 \\
&= \frac{1}{4}\sum_{i=1}^2\sum_{j=1}^2 |\mathcal{K}(|x_i - x_j|)|^2.
\end{aligned}$$

Integrating over time the latter equality yields (36) for $N = 2$ with an actual equivalence and $c_T = 1$. Notice that here we have not made any specific assumptions on the trajectories $t \to x_i(t)$. However, the case of two particles might be perceived too simple and we may want to extend our arguments to multiple particles. Let us then consider the case of $N = 3$ particles. Already in this simple case the angles between particles may be rather arbitrary and analyzing the many possible cases is an involved exercise. For this reason we assume that $d = 2$ and that at a certain time t the particles are disposed precisely at the vertexes of a equilateral triangle of edge length r. We also assume that $\mathcal{K}$ gets its maximal value precisely at r, hence

$$\frac{1}{9}\sum_{i=1}^3\sum_{j=1}^3 |\mathcal{K}(|x_i - x_j|)|^2 dt \leq \|\mathcal{K}\|_\infty^2 = \mathcal{K}(r)^2.$$

Notice that it holds also

$$\frac{1}{9T}\int_0^T \sum_{i=1}^3\sum_{j=1}^3 |\mathcal{K}(|x_i - x_j|)|^2 dt \leq \|\mathcal{K}\|_\infty^2 = \mathcal{K}(r)^2. \tag{37}$$

A direct computation in this case shows that

$$\frac{1}{3}\sum_{i=1}^3 \left| \frac{1}{3}\sum_{j=1}^3 \mathcal{K}(|x_i - x_j|)\frac{x_i - x_j}{|x_i - x_j|}\right|^2 = \frac{1}{3}\mathcal{K}(r)^2.$$

and therefore

$$\frac{1}{3}\sum_{i=1}^{3}\left|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(|x_i-x_j|)\frac{x_i-x_j}{|x_i-x_j|}\right|^2 \geq \frac{1}{18}\sum_{i=1}^{3}\sum_{j=1}^{3}|\mathcal{K}(|x_i-x_j|)|^2.$$

Unfortunately the assumption that $\mathcal{K}$ achieves its maximum at r does not allow us yet to conclude by a simple integration over time the coercivity condition as we did for the case of two particles. In order to extend the validity of the inequality to arbitrary functions taking maxima at other points, we need to integrate over time by assuming now that the particles are vertexes of equilateral triangles with time dependent edge length, say from $r=0$ growing in time up to $r=2R>0$. This will allow the trajectories to explore any possible distance within a given interval and to capture the maximal value of any kernel. More precisely, let us now assume that $\mathcal{K}$ is an arbitrary bounded continuous function, achieving its maximum value over $[0,2R]$, say at $r_0\in(0,2R)$ and we can assume that this is obtained corresponding to the time t_0 when the particles forms precisely the triangle of side length r_0. Now we need to make a stronger assumption on $\widehat{a}$, i.e., we require $\widehat{a}$ to be coming from a class of equi-continuous functions, for instance functions which are Lipschitz continuous with uniform Lipschitz constant. (Actually, later we will restrict our attention to functions in $X_{M,K}$ which is precisely of this type!) Under this equi-continuity assumption, there exist $\varepsilon>0$ and a constant $c_{T,\varepsilon}>0$ independent of $\mathcal{K}$ (but perhaps depending only on its modulus of continuity) such that

$$\begin{aligned}\frac{1}{T}\int_0^T\frac{1}{3}\sum_{i=1}^{3}\left|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(|x_i-x_j|)\frac{x_i-x_j}{|x_i-x_j|}\right|^2 dt &\geq \frac{1}{T}\int_{t_0-\varepsilon}^{t_0+\varepsilon}\frac{1}{3}\sum_{i=1}^{3}\left|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(|x_i-x_j|)\frac{x_i-x_j}{|x_i-x_j|}\right|^2\\ &\geq \frac{c_{T,\varepsilon}}{3}\mathcal{K}(r_0)\geq\frac{c_{T,\varepsilon}}{18T}\int_0^T\sum_{i=1}^{3}\sum_{j=1}^{3}|\mathcal{K}(|x_i-x_j|)|^2dt.\end{aligned}$$

In the latter inequality we used (37). Hence, also in this case, one can construct examples for which the coercivity assumption is verifiable. Actually this construction can be extended to any group of N particles disposed on the vertexes of regular polygons. As an example of how one should proceed, let us consider the case of $N=4$ particles disposed instantanously at the vertexes of a square of side length $\sqrt{2}r>0$. In this case one directly verfies that

$$\frac{1}{4}\sum_{i=1}^{4}\left|\frac{1}{4}\sum_{j=1}^{4}\mathcal{K}(|x_i-x_j|)\frac{x_i-x_j}{|x_i-x_j|}\right|^2=\frac{1}{16}(\mathcal{K}(2r)^2+2\sqrt{2}\mathcal{K}(2r)\mathcal{K}(\sqrt{2}r)+2\mathcal{K}(\sqrt{2}r)^2). \quad (38)$$

Let us assume that the maximal value of $\mathcal{K}$ is attained precisely at $2r$. Then the expression (38) is always bounded below by

$$\frac{1}{4}\sum_{i=1}^{4}\left|\frac{1}{4}\sum_{j=1}^{4}\mathcal{K}(|x_i-x_j|)\frac{x_i-x_j}{|x_i-x_j|}\right|^2\geq\frac{1}{16}(\mathcal{K}(2r)^2-2\sqrt{2}\mathcal{K}(2r)^2+2\mathcal{K}(\sqrt{2}r)^2)=\frac{3-2\sqrt{2}}{16}\mathcal{K}(\sqrt{2}r)^2.$$

Hence, also in this case, we can apply the continuity argument above to eventually show the coercivity condition. Similar procedures can be followed for any $N \geq 5$. However, as $N \to \infty$ one can show numerically that the lower bound vanishes quite rapidly, making it impossible, perhaps not suriprisingly, to conclude the coercivity condition for the uniform distribution over the circle.

At this point one may complain that all the examples presented so far are based on discrete measures $\mu^N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ supported on particles lying on the vertexes of polytopes. However, one can consider an approximated identity g_ε for which $g_\varepsilon \to \delta_0$ for $\varepsilon \to 0$ and the regularized probability measure

$$\mu_\varepsilon(x) = g_\varepsilon * \mu^N(x) = \frac{1}{N}\sum_{i=1}^N g_\varepsilon(x - x_i).$$

This diffuse measure approximates μ^N in the sense that $\mathcal{W}_1(\mu_\varepsilon, \mu^N) \to 0$ for $\varepsilon \to 0$, hence in particular intergrals against Lipschitz functions can be well-approximated, i.e.,

$$\left| \int_{\mathbb{R}^d} \varphi(x) d\mu^N(x) - \int_{\mathbb{R}^d} \varphi(x) d\mu_\varepsilon(x) \right| \leq \mathrm{Lip}(\varphi)\mathcal{W}_1(\mu_\varepsilon, \mu^N).$$

Under the additional assumption that $\mathrm{Lip}_K(\widehat{a}) \sim \|\widehat{a}\|_{L_\infty(K)}$ (and this is true whenever $\widehat{a}$ is a piecewise polynomial function over a finite partition of $\mathbb{R}_+$) one can extend the validity of the coercivity condition for μ^N (36) to μ_ε as follows

$$\frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \mathcal{K}(|x-y|)\frac{y-x}{|y-x|} d\mu_\varepsilon(x) \right|^2 d\mu_\varepsilon(y) dt \geq \frac{c_{T,\varepsilon}}{T}\int_0^T \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\mathcal{K}(|x-y|)|^2 d\mu_\varepsilon(x) d\mu_\varepsilon(y) dt,$$

for a constant $c_{T,\varepsilon} > 0$ for $\varepsilon > 0$ small enough.

We hope that these examples convince the reader that the coercivity condition is verifiable in many cases and from now on we assume it without further concerns. We conclude this section by presenting two simple implications of the coercivity condition. Notice that, as an easy consequence of Lemma 3.6, it holds

$$\|a\|^2_{L_2(\mathbb{R}_+,\rho)} = \int_{\mathbb{R}_+} \left|a(s)\right|^2 d\rho(s) \leq \mathrm{diam}(\mathrm{supp}(\rho))^2 \|a\|^2_{L_\infty(\mathrm{supp}(\rho))}, \tag{39}$$

and hence, $X \subseteq L_2(\mathbb{R}_+, \rho)$.

Proposition 3.7. *Assume $a \in X$ and that* (35) *holds. Then any minimizer of $\mathcal{E}$ in X coincides ρ-a.e. with a.*

Proof. Notice that $\mathcal{E}(a) = 0$, and since $\mathcal{E}(\widehat{a}) \geq 0$ for all $\widehat{a} \in X$ this implies that a is a minimizer of $\mathcal{E}$. Now suppose that $\mathcal{E}(\widehat{a}) = 0$ for some $\widehat{a} \in X$. By (35) we obtain that $\widehat{a} = a$ in $L_2(\mathbb{R}_+, \rho)$, and therefore they coincide ρ-almost everywhere. □

3.3 Existence of minimizers of $\mathcal{E}_N$

The following proposition, which is a straightforward consequence of Ascoli-Arzelá Theorem, indicates the right ambient space where to state an existence result for the minimizers of $\mathcal{E}_N$.

Proposition 3.8. *Fix $M > 0$ and $K = [0, 2R] \subset \mathbb{R}_+$ for any $R > 0$. Define the set*

$$X_{M,K} = \left\{ b \in W^{1,\infty}(K) : \|b\|_{L_\infty(K)} + \|b'\|_{L_\infty(K)} \leq M \right\}.$$

The space $X_{M,K}$ is relatively compact with respect to the uniform convergence on K.

Proof. Consider $(\widehat{a}_n)_{n\in\mathbb{N}} \subset X_{M,K}$. The Fundamental Theorem of Calculus (which is applicable for functions in $W^{1,\infty}$, see [**?**, Theorem 2.8]) tells us that, for any $r_1, r_2 \in K$ it holds

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

This implies

$$\left|a_n(r_1) - a_n(r_2)\right| \leq \int_{r_1}^{r_2} |a_n'(r)|dr \leq \|a_n'\|_{L_\infty(K)}|r_2 - r_1|.$$

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. Hence from the Ascoli-Arzelá Theorem we can deduce the existence of a subsequence (which we do not relabel) converging uniformly on K to some $\widehat{a} \in X_{M,K}$, proving the statement. $\square$

Proposition 3.9. *Assume $a \in X$. Fix $M > 0$ and $K = [0, 2R] \subset \mathbb{R}_+$ for $R > 0$ as in Proposition 2.2. Let V be a closed subset of $X_{M,K}$ w.r.t. the uniform convergence. Then, the optimization problem*

$$\textit{minimize } \mathcal{E}_N(\widehat{a}) \textit{ among all } \widehat{a} \in V$$

admits a solution.

Proof. In light of the fact that $\inf \mathcal{E}_N \geq 0$, we can consider a minimizing sequence $(\widehat{a}_n)_{n\in\mathbb{N}} \subset V$, i.e., it holds $\lim_{n\to\infty} \mathcal{E}_N(\widehat{a}_n) = \inf_V \mathcal{E}_N$. By Proposition 3.8 there exists a subsequence of $(\widehat{a}_n)_{n\in\mathbb{N}}$ (which we do not relabel) converging uniformly on K to a function $\widehat{a} \in V$ (since V is closed). We now show that $\lim_{n\to\infty} \mathcal{E}_N(\widehat{a}_n) = \mathcal{E}_N(\widehat{a})$, from which it follows that $\mathcal{E}_N$ 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 K and the compactness of K 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 (48)). Moreover, we have the uniform bound

$$\begin{aligned}\sup_{x,y\in B(0,R)} |F[\widehat{a}_n](x-y) - F[a](x-y)| &= \sup_{x,y\in B(0,R)} |\widehat{a}_n(|x-y|) - a(|x-y|)||x-y| \\ &\leq 2R \sup_{r\in K} |\widehat{a}_n(r) - a(r)| \\ &\leq 2R(M + \|a\|_{L_\infty(K)}).\end{aligned} \tag{40}$$

As the measures $\mu^N(t)$ are compactly supported in $B(0,R)$ uniformly in time, the boundedness (40) allows us to apply three times the dominated convergence theorem to yield

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

which proves the statement.

$\square$

4 Γ-convergence of $\mathcal{E}_N$ to $\mathcal{E}$

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 $\mathcal{E}_N$ converge to minimizers of $\mathcal{E}$ by Γ-convergence.

Definition 4.1. Let $M > 0$ and $K = [0, 2R]$ interval in $\mathbb{R}_+$ be given . We say that a family of closed subsets $V_N \subset X_{M,K}$, $N \in \mathbb{N}$ has the *uniform approximation property* in $L_\infty(K)$ if for all $b \in X_{M,K}$ there exists a sequence $(b_N)_{N\in\mathbb{N}}$ converging uniformly to b on K and such that $b_N \in V_N$ for every $N \in \mathbb{N}$.

Theorem 4.2. *Assume $a \in X$, fix $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and let $K = [0, 2R]$ be an interval in $\mathbb{R}_+$ with $R > 0$ as in Proposition 2.2. Set*

$$M \geq \|a\|_{L_\infty(K)} + \|a'\|_{L_\infty(K)}.$$

For every $N \in \mathbb{N}$, let $x^N_{0,1}, \dots, x^N_{0,N}$ be i.i. μ_0-distributed and define $\mathcal{E}_N$ as in (29) *for the solution μ^N of system* (20) *with initial datum*

$$\mu^N_0 = \frac{1}{N}\sum_{i=1}^N \delta_{x^N_{0,i}}, \quad x \in \mathbb{R}^d.$$

For $N \in \mathbb{N}$, let $V_N \subset X_{M,K}$ be a sequence of subsets with the uniform approximation property as in Definition 4.1 and consider

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

Then the sequence $(\widehat{a}_N)_{N\in\mathbb{N}}$ converges uniformly on K to some continuous function $\widehat{a} \in X_{M,K}$ such that $\mathcal{E}(\widehat{a}) = 0$. If we additionally assume the coercivity condition (35), *then it holds $\widehat{a} = a$ in $L_2(\mathbb{R}_+, \rho)$.*

We start with a technical lemma.

Lemma 4.3. *Under the assumptions of Theorem 4.2, let* $(b_N)_{N\in\mathbb{N}} \subset X_{M,K}$ *be a sequence of continuous functions uniformly converging to a function* $b \in X_{M,K}$ *on* $K = [0, 2R]$ *with* $R > 0$ *as in* (48). *Then it holds*

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

Proof. From [**?**, Lemma 3.3] it follows $\mathcal{W}_1(\mu_0, \mu_0^N) \to 0$ for $N \to \infty$. Hence, from (28) we have that $W_1(\mu(t), \mu^N(t)) \to 0$ for $N \to \infty$, uniformly for $t \in [0, T]$.

$$\begin{aligned}|(F[a] - F[b])(x - y') - &(F[a] - F[b])(x - y)| \\ &\leq [2R(\mathrm{Lip}_K(b) + \mathrm{Lip}_K(b)) + \|a\|_{L_\infty(K)} + \|b\|_{L_\infty(K)}]|y - y'|,\end{aligned}$$

implying the uniform Lipschitz continuity of $(F[a] - F[b])(x - \cdot)$ with respect to $x \in B(0, R)$. For every $\varepsilon > 0$ one can find $N_0(\varepsilon)$ such that, for all $N \geq N_0(\varepsilon)$ we have

$$\sup_{x,y\in B(0,R)} |F[b_N](x - y) - F[b](x - y)| \leq 2R\|b_N - b\|_{L_\infty(K)} \leq \varepsilon/2,$$

as well as

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

uniformly with respect to $t \in [0, T]$ and $x \in B(0, R)$. The first estimate follows from the uniform convergence of the b_N, while the second one follows from the uniform Lipschitz continuity of $(F[a]-F[b])(x-\cdot)$ with respect to $x \in B(0, R)$ and the uniform Wasserstein convergence of $\mu^N(t)$ to $\mu(t)$ with respect to $t \in [0, T]$. Hence for $N \geq N_0(\varepsilon)$ we obtain

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

which implies that, for every $t \in [0, T]$ and $x \in B(0, R)$, it holds

$$\lim_{N\to\infty} \left| \int_{\mathbb{R}^d} \big(F[b_N] - F[a]\big)(x - y) d\mu^N(t)(y) \right|^2 = \left| \int_{\mathbb{R}^d} \big(F[b] - F[a]\big)(x - y) d\mu(t)(y) \right|^2. \tag{41}$$

Denote

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

and we estimate

$$
\begin{aligned}
|G_N(t) - G(t)| \leq & \left| \int_{\mathbb{R}^d} H(t,x)d\mu^N(t)(x) - \int_{\mathbb{R}^d} H(t,x)d\mu(t)(x) \right| \\
& + \int_{\mathbb{R}^d} |H_N(t,x) - H(t,x)|\, d\mu(t)(x). \qquad (42)
\end{aligned}
$$

We now prove that the function H is Lipschitz continuous with respect to $x \in B(0,R)$ uniformly in $t \in [0,T]$. To do so, we write H as

$$
H(t,x) = g\left(\int_{\mathbb{R}^d} f(x,y)d\mu(t)(y) \right),
$$

where g is a differentiable function in $\mathbb{R}^d$ and $f(\cdot,y)$ is a Lipschitz continuous function uniformly in y with values in $\mathbb{R}^d$ (in this case $f(\cdot,y) = (F[a]-F[b])(\cdot - y)$). Since the measure $\mu(t)$ has support contained in $B(0,R)$, it follows

$$
\begin{aligned}
\left| \int_{\mathbb{R}^d} f(x,y)d\mu(t)(y) \right| &\leq \int_{\mathbb{R}^d} |f(x,y)|d\mu(t)(y) \\
&\leq \sup_{x,y\in B(0,R)} |f(x,y)| \\
&=: S < +\infty.
\end{aligned}
$$

Therefore, for every $x, x' \in B(0,R)$ we can argue

$$
\begin{aligned}
\left| H(t,x) - H(t,x') \right| &\leq \mathrm{Lip}_{B(0,S)}(g) \left| \int_{\mathbb{R}^d} f(x,y)d\mu(t)(y) - \int_{\mathbb{R}^d} f(x',y)d\mu(t)(y) \right| \\
&\leq \mathrm{Lip}_{B(0,S)}(g) \int_{\mathbb{R}^d} \left| f(x,y) - f(x',y) \right| d\mu(t)(y) \\
&\leq \mathrm{Lip}_{B(0,S)}(g)\ \mathrm{Lip}_{B(0,R)}(f)\ |x - x'|,
\end{aligned}
$$

from which follows the Lipschitz continuity of $H(t,x)$ with respect to $x \in B(0,R)$ uniformly in $t \in [0,T]$.

From this uniform Lipschitz continuity and the uniform Wasserstein convergence of $\mu^N(t)$ to $\mu(t)$ with respect to $t \in [0,T]$, it follows that for every $\varepsilon > 0$ we can find $N_0(\varepsilon)$ such that for all $N \geq N_0(\varepsilon)$ it holds

$$\left| \int_{\mathbb{R}^d} H(t,x) d\mu^N(t)(x) - \int_{\mathbb{R}^d} H(t,x) d\mu(t)(x) \right| \leq \frac{\varepsilon}{2}, \tag{43}$$

uniformly with respect to $t \in [0,T]$. From (41) it follows also that for all $N \geq N_0(\varepsilon)$ we have

$$|H_N(t,x) - H(t,x)| \leq \frac{\varepsilon}{2}, \tag{44}$$

uniformly with respect to $t \in [0,T]$ and $x \in B(0,R)$. A combination of (42) with (43) and (44) yields $|G_N(t) - G(t)| \leq \varepsilon$ uniformly in $t \in [0,T]$. Hence

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

holds uniformly in $t \in [0,T]$.

To eventually show that $\lim_{N\to\infty} \mathcal{E}_N(b_N) = \mathcal{E}(b)$, we simply note that uniform convergence of G_N to G implies

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

□

Proof of Theorem 4.2. The sequence of minimizers $(\widehat{a}_N)_{N\in\mathbb{N}}$ is by definition a subset of $X_{M,K}$, hence by Proposition 3.8 it admits a subsequence $(\widehat{a}_{N_k})_{k\in\mathbb{N}}$ uniformly converging to a function $\widehat{a} \in X_{M,K}$.

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

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

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

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

We can therefore conclude the fundamental estimate

$$\mathcal{E}(b) \geq \mathcal{E}(\widehat{a}), \tag{45}$$

which holds for every $b \in X_{M,K}$. In particular, (45) applies to $b = a \in X_{M,K}$ (by the particular choice of M), which finally implies

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

showing that $\widehat{a}$ is also a minimizer of $\mathcal{E}$. In case (35) holds, by Proposition 3.7 it follows $\widehat{a} = a$ in $L_2(\mathbb{R}_+, \rho)$. □

5 Numerical experiments

In order to convince the reader of the usefulness of Theorem 4.2, we report below in Figure 1 one (of several we did) preliminary numerical experiment indicating the successful recovery by an adaptive algorithm based on (9) of a potential function a in a first order model of the type (2). More specifically, we performed successive applications of (9) based on an adaptive refinement of a mesh on the positive real line, defining approximating spaces V of continuous piecewise linear functions. The adaptation is lead by a posteriori error estimator based on local comparison of two successive iterations.

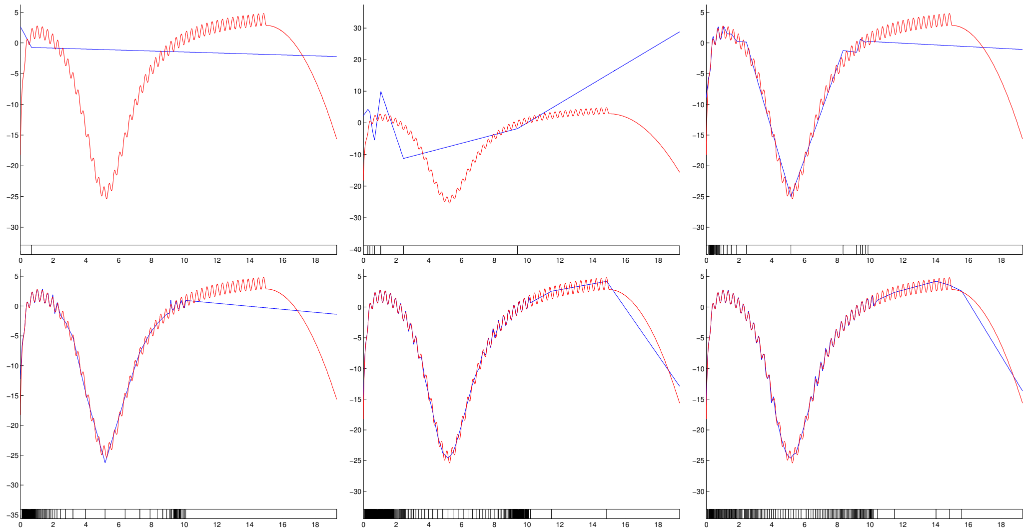

Figure 1: First preliminary numerical experiments indicating the successful recovery by an adaptive algorithm based on (9) of a potential function a in a first order model of the type (2). The potential a to be recovered is displayed in red color and it's the strongly oscillating function. The blue function is the piecewise linear approximant computed at each successive iteration after adaptation of the underlying mesh, scketched on the bottom of the figures.

Despite the highly oscillatory nature of the parameter function a, the algorithm performs an excellent approximation, providing also a sort of "numerical homogenization" in those locations of the positive real line, where not enough data are provided by the evolution.

6 Appendix

6.1 Technical lemmas for the mean-field limit

The following preliminary result tells us that solutions to system (21) are also solutions to systems (20), whenever conveniently rewritten.

Proposition 6.1. *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 (21) 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 (23) is a solution of (20) with initial datum $\mu_0 = \mu^N(0) \in \mathcal{P}_c(\mathbb{R}^d)$.*

Proof. It can be easily proved by arguing exactly as in [**?**, Lemma 4.3]. □

We are able to state several basic estimates that shall be useful towards an existence and uniqueness result for the solutions of system (21).

Lemma 6.2. *Let $a \in X$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$. Then for all $y \in \mathbb{R}^d$ the following hold:*

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

Proof. Trivially follows from $a \in L_\infty(\mathbb{R}_+)$. □

Lemma 6.3. *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

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

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. □

Lemma 6.4. *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

$$\begin{aligned}
|(F[a] * \mu)(x) - (F[a] * \mu)(y)| &= \left| \int_{\mathbb{R}^d} a(|x-z|)(x-z) d\mu(z) - \int_{\mathbb{R}^d} a(|y-z|)(y-z) d\mu(z) \right| \\
&\leq \int_{\mathbb{R}^d} |a(|x-z|) - a(|y-z|)||x-z| d\mu(z) \\
&\quad + \int_{\mathbb{R}^d} a(|y-z|)|x-y| d\mu(z)
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathrm{Lip}_{\widehat{K}}(a)|x-y| \int_{\mathbb{R}^d} |x-z| d\mu(z) + \|a\|_{L_\infty(\mathbb{R}_+)}|x-y| \\
&\leq \left(\mathrm{Lip}_{\widehat{K}}(a)(|x|+1) + \|a\|_{L_\infty(\mathbb{R}_+)}\right)|x-y| \\
&\leq \left(C\mathrm{Lip}_{\widehat{K}}(a) + \|a\|_{L_\infty(\mathbb{R}_+)}\right)|x-y|,
\end{aligned}
$$

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

Proposition 6.5. *If $a \in X$ then system (21) admits a unique global solution in $[0,T]$ for every initial datum $x_0^N \in \mathbb{R}^{dN}$.*

Proof. Rewriting system (21) in the form of (22), from Lemma 6.4 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 (23), 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 [**?**, Lemma 6.7] and [**?**, Lemma 4.7]

Lemma 6.6. *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*

$$
\mathrm{supp}(\mu(t)) \cup \mathrm{supp}(\nu(t)) \subseteq B(0,R), \tag{46}
$$

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

$$
\|F[a] * \mu(t) - F[a] * \nu(t)\|_{L_\infty(B(0,r))} \leq L_{a,r,R}\mathcal{W}_1(\mu(t), \nu(t)) \tag{47}
$$

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

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

By using Lemma 6.3 and the hypothesis (46), 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 (47) holds with $L_{a,r,R} = \mathrm{Lip}_{B(0,R+r)}(F[a])$. □

6.2 Proof of Proposition 2.2

Notice that for every $N \in \mathbb{N}$, by Proposition 6.1, μ^N is the unique solution of (20) 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 6.2, we have

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

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 [**?**, Chapter 2, Lemma A.4] ensures that for a.e. $t \in [0, T]$ there exists $k = 1, \ldots, N$ such that

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

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

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

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

$$R = \sqrt{C_0} e^{2\|a\|_{L_\infty(\mathbb{R}_+)}T}. \tag{48}$$

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 6.2 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, [**?**, 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

$$\lim_{k\to\infty}\mathcal{W}_1(\mu^{N_k}(t),\mu(t))=0\quad\text{uniformly for a.e. } t\in[0,T], \tag{49}$$

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 (20). 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 6.6, the inequality (49), 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

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

which proves that μ is a solution of (20) with initial datum μ_0.

6.3 Existence and uniqueness of solutions for (24)

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 $0<\tau\le T$. A function $y\colon[0,\tau]\to\Omega$ is called a solution of the Carathéodory differential equation

$$\dot{y}(t)=g(t,y(t)) \tag{50}$$

on $[0,\tau]$ if and only if y is absolutely continuous and (50) is satisfied a.e. in $[0,\tau]$. The following existence result holds.

Theorem 6.7. *Fix $T>0$ and $y_0\in\mathbb{R}^d$. Suppose that there exists a compact subset Ω of $\mathbb{R}^d$ such that $y_0\in\operatorname{int}(\Omega)$ and there exists $m_\Omega\in L_1([0,T])$ for which it holds*

$$|g(t,y)|\le m_\Omega(t), \tag{51}$$

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 (50) 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

$$|g(t,y)|\le C(1+|y|), \tag{52}$$

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 (50) 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

$$|y(t)|\le\Big(|y_0|+Ct\Big)\,e^{Ct}. \tag{53}$$

Proof. Since $y_0\in\operatorname{int}(\Omega)$, we can consider a ball $B(y_0,r)\subset\Omega$. The classical result [**?**, Chapter 1, Theorem 1] and (51) yield the existence of a local solution defined on an interval $[0,\tau]$ and taking values in $B(y_0,r)$.

If (52) holds, any solution of (50) with initial datum y_0 satisfies

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

for every $t\in[0,\tau]$, therefore (53) 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 [**?**, Chapter 1, Theorem 4]. $\square$

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

Lemma 6.8. *Let g_1 and $g_2\colon[0,T]\times\mathbb{R}^n\to\mathbb{R}^n$ be Carathéodory functions both satisfying (52) 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

$$|y_1(t) - y_2(t)| \leq e^{Lt} \left(|y_1(0) - y_2(0)| + \int_0^t \|g_1(s, \cdot) - g_2(s, \cdot)\|_{L_\infty(B(0,\rho_{r,C,T}))} \, ds \right) \tag{54}$$

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

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

$$\begin{aligned}
|y_1(t) - y_2(t)| &\leq |y_1(0) - y_2(0)| + \int_0^t |\dot{y}_1(s) - \dot{y}_2(s)| ds \\
&= |y_1(0) - y_2(0)| \\
&\quad + \int_0^t |g_1(s, y_1(s)) - g_1(s, y_2(s)) + g_1(s, y_2(s)) - g_2(s, y_2(s))| ds \\
&\leq |y_1(0) - y_2(0)| + \int_0^t \|g_1(s, \cdot) - g_2(s, \cdot)\|_{L_\infty(B(0,\rho_{r,C,T}))} \, ds \\
&\quad + L \int_0^t |y_1(s) - y_2(s)| ds.
\end{aligned}$$

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 (54), as desired. □

Proposition 6.9. *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 2.2 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 (24) with initial value μ_0 defined on the whole interval $[0, T]$.

Proof. By Lemma 6.2 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 (51). Moreover, for fixed t this function is locally Lipschitz continuous, as follows from Lemma 6.4, 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 6.2, 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 (52) holds. By considering a sufficiently large compact set K containing ξ_0, Theorem 6.7 guarantees the existence of a solution of system (24) defined on $[0, T]$.

To establish uniqueness notice that, from Lemma 6.3, for every compact subset $K \in \mathbb{R}^d$ and any $x, y \in K$, it holds

$$
\begin{aligned}
|(F[a] * \mu(t))(x) - (F[a] * \mu(t))(y)| &\leq \left| \int_{\mathbb{R}^d} F[a](x-z) d\mu(t)(z) - \int_{\mathbb{R}^d} F[a](y-z) d\mu(t)(z) \right| \\
&\leq \int_{\mathbb{R}^d} |F[a](x-z) - F[a](y-z)| \, d\mu(t)(z) \\
&\leq \mathrm{Lip}_{\widehat{K}}(F[a]) |x-y|,
\end{aligned}
\tag{55}
$$

where $\widehat{K}$ is a compact set containing both K and $B(0, R)$. Hence, uniqueness follows from (55) and Lemma 6.8 by taking $g_1 = g_2$, $y_1(0) = y_2(0)$ and $r = |y_1(0)|$. □

6.4 Continuous dependence on the initial data

The following Lemma and (54) are the main ingredients of the proof of Theorem 2.4 on continuous dependance on initial data.

Lemma 6.10. *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*

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

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

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

Proof. See [**?**, Lemma 3.11] and [**?**, Lemma 3.13]. □

We can now prove Theorem 2.4.

Proof of Theorem 2.4. Let $\mathcal{T}_t^\mu$ and $\mathcal{T}_t^\nu$ be the flow maps associated to system (24) with measure μ and ν, respectively. By (25), the triangle inequality, Lemma 6.6, Lemma 6.10 and (26) we have for every $t \in [0, T]$

$$
\begin{aligned}
\mathcal{W}_1(\mu(t), \nu(t)) &= \mathcal{W}_1((\mathcal{T}_t^\mu)_\# \mu_0, (\mathcal{T}_t^\nu)_\# \nu_0) \\
&\leq \mathcal{W}_1((\mathcal{T}_t^\mu)_\# \mu_0, (\mathcal{T}_t^\mu)_\# \nu_0) + \mathcal{W}_1((\mathcal{T}_t^\mu)_\# \nu_0, (\mathcal{T}_t^\nu)_\# \nu_0) \\
&\leq e^{T \, \mathrm{Lip}_{B(0,R)}(F[a])} \mathcal{W}_1(\mu_0, \nu_0) + \|\mathcal{T}_t^\mu - \mathcal{T}_t^\nu\|_{L_\infty(B(0,R))}.
\end{aligned}
\tag{56}
$$

Using (54) with $y_1(0) = y_2(0)$ we get

$$\|\mathcal{T}_t^\mu - \mathcal{T}_t^\nu\|_{L_\infty(B(0,r))} \leq e^{t \, \mathrm{Lip}_{B(0,R)}(F[a])} \int_0^t \|F[a] * \mu(s) - F[a] * \nu(s)\|_{L_\infty(B(0,R))} \, ds. \tag{57}$$

Combining (56) and (57) with Lemma 6.6, 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 6.6. 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 (28) with $\overline{C} = e^{T\,\mathrm{Lip}_{B(0,R)}(F[a])+L_{a,R,R}}$.

Consider now two solutions of (20) with the same initial datum μ_0. Since, from Proposition 2.2 they both satisfy (27) for the given *a priori known* R given by (48), then (28) guarantees they both describe the same curve in $\mathcal{P}_1(\mathbb{R}^d)$. This concludes the proof. $\square$