On the Identifiablility of Nonlocal Interaction Kernels in First-Order Systems of Interacting Particles on Riemannian Manifolds

Sui Tang *[1], Malik Tuerkoen†[1], and Hanming Zhou ‡[1]

[1]Department of Mathematics, University of California Santa Barbara

September 24, 2023

Abstract

In this paper, we tackle a critical issue in nonparametric inference for systems of interacting particles on Riemannian manifolds: the identifiability of the interaction functions. Specifically, we define the function spaces on which the interaction kernels can be identified given infinite i.i.d observational derivative data sampled from a distribution. Our methodology involves casting the learning problem as a linear statistical inverse problem using a operator theoretical framework. We prove the well-posedness of inverse problem by establishing the strict positivity of a related integral operator and our analysis allows us to refine the results on specific manifolds such as the sphere and Hyperbolic space. Our findings indicate that a numerically stable procedure exists to recover the interaction kernel from finite (noisy) data, and the estimator will be convergent to the ground truth. This also answers an open question in [MMQZ21] and demonstrate that least square estimators can be statistically optimal in certain scenarios. Finally, our theoretical analysis could be extended to the mean-field case, revealing that the corresponding nonparametric inverse problem is ill-posed in general and necessitates effective regularization techniques.

1 Introduction

Systems of interacting particles are ubiquitous in science and engineering, where the particles may refer to fundamental particles in physics, planets in astronomy, animals in ecology, and cells in biology. These systems exhibit a wide range of collective behaviors at different scales and levels of complexity arising from individual interactions among particles. Understanding and simulating such collective behavior requires the development of effective differential equation models, which is a central subject in applied mathematics.

In many applications, particles may be associated with state variables that are defined on or constrained to move on non-Euclidean spaces. This presents significant difficulties for modeling, analysis, and numerical simulation. Despite these challenges, there has been a growing interest in the last five years in modeling particles moving on various manifolds or surfaces. One such example is a simple first-order system that models the consensus behavior of opinion dynamics on a general Riemannian manifold. This system considers N interacting particles on a Riemannian manifold $(\mathcal{M}, g)$, which evolve according to a general dynamics equation shown in (1):

*Email: suitang@ucsb.edu
†Email: malik@math.ucsb.edu
‡Email:hzhou@math.ucsb.edu

$$\dot{\boldsymbol{x}}_i = \frac{1}{N}\sum_{j=1}^{N}\phi(d(\boldsymbol{x}_i,\boldsymbol{x}_j))w(\boldsymbol{x}_i,\boldsymbol{x}_j), i = 1,\cdots,N. \tag{1}$$

Here, $\boldsymbol{x}_i$ represents the position of the i-th particle on the manifold $\mathcal{M}$. The radial interaction kernel ϕ is a scalar valued function defined over $\mathbb{R}^+$; the distance function $d(\cdot,\cdot)$ with respect to the Riemannian metric g, and influence vector $w(\boldsymbol{x}_i,\boldsymbol{x}_j) \in T_{\boldsymbol{x}_i}\mathcal{M}$ (the tangent space at $\boldsymbol{x}_i$) are all defined in the equation. Other models, such as flocking models in [AHS21, AHPS21], aggregation models in [FZ18, FPP21], and Kuramoto models in [Str00, HKK22], have also been shown to reproduce various qualitative patterns of collective dynamics.

In the field of modelling, one of the greatest challenges is selecting the appropriate governing equations to accurately describe the desired collective behaviour. In the past, this task relied heavily on the expertise of the modeler. However, modern sensor and measurement technologies now allow for the collection of large amounts of high-quality data from a diverse range of systems. As a result, the discovery of interacting particle models that accurately match observational data has become a major area of focus in recent years. This task is often difficult and suffers from the curse of dimensionality, but many systems have low-dimensional structures that enable efficient data-driven methods. Examples of this include the inference of stochastic interacting particle systems in works such as [Kas90, B$^+$11, GSW19, Che21, SKPP21, MB21a, GCL22, DMH22, YCY22], and the learning of radial interaction kernels in [LMT21]. Deterministic interacting particle systems on Euclidean domains have also been studied, with works such as [BFHM17, LZTM19, LMT21, MTZM20] focusing on nonparametric inference methods.

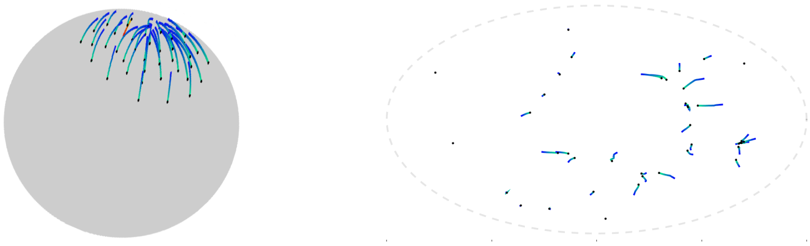

Figure 1: The two examples of systems (1) using code provided in [MMQZ21]. The left models the swarming behaviour of preys and predators on $\mathbb{S}^2$ and the right models the aggregation behaviours of opinions on Poincare disk. The color variations from blue to green denotes the forward time evolutions of particles.

In their work [MMQZ21], the authors propose a least squares algorithm that uses the ℓ^2 ODE residual (1) as the loss function and looks for an approximation of ϕ over piecewise polynomial spaces. This learning scheme exploits the low-dimensional structure of the problem, where the unknown function ϕ is a radial function defined on $\mathbb{R}^+$ regardless of the value of N. However, this approach leads to an inverse problem that requires solving for the identifiability of ϕ, which is essential for analyzing the statistical optimality of least squares estimators.

In this paper, we approach the identifiability of ϕ in (1) from a linear statistical inverse learning problem perspective [FRT21]. That is, the observational data comes in the form of positions of the N particles on the manifold and the (possible noisy) first derivatives of these positions with respect to time, and they follow a joint probability distribution. To simplify notation, we write (1) as

$$\dot{\mathbf{X}}(t) = \mathbf{f}_\phi(\mathbf{X}(t)) \tag{2}$$

where $\mathbf{X} = (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N) \in \mathcal{M}^N$ and $\mathbf{f}_\phi(\boldsymbol{X}) \in T_{\mathbf{X}}\mathcal{M}^N$. We denote $\mathcal{M}^N = \mathcal{M}\times\cdots\times\mathcal{M}$ the product manifold consisting of N copies of $\mathcal{M}$, with $T_{\mathbf{X}}\mathcal{M}^N$ the tangent space of $\mathcal{M}^N$ at $\mathbf{X}$. Let A be a linear operator that maps a function φ to $\mathbf{f}_\varphi$. Our study addresses two fundamental questions:

Question 1. What are function spaces $\mathcal{H}$ and $\mathcal{F}$ such that $A \in \mathcal{B}(\mathcal{H}, \mathcal{F})$ (bounded linear operators mapping $\mathcal{H}$ to $\mathcal{F}$)?

Question 2. Suppose $A \in \mathcal{B}(\mathcal{H}, \mathcal{F})$, when does the linear inverse problem

$$A\varphi = \mathbf{f}_\varphi \tag{3}$$

have a stable solution?

We restrict our attention to $\mathcal{F} = L^2(\mathcal{M}^N; \mu; T_{\mathbf{X}}\mathcal{M}^N)$ where μ is the marginal distribution of positions and $\mathcal{H}$ to be a subspace of $L^2([0, R]; \rho; \mathbb{R})$ with a suitable choice of measure ρ constructed from μ. The use of L^2 spaces in both $\mathcal{H}$ and $\mathcal{F}$ is a natural choice when dealing with the ℓ^2 loss function used in the learning. The space $\mathcal{F}$ is a space of square-integrable vector fields on $\mathcal{M}^N$ with respect to a probability measure μ, which is the marginal distribution of observational position data. The space $\mathcal{H}$ is a space of square-integrable functions supported on a interval $[0, R]$ so that the ODE system (1) is well-posed. For example, if $\mathcal{M}$ is compact, one can choose R to be injective radius of $\mathcal{M}$, see section2 in [MMQZ21] and [AMD17]. The answers to Question 1 and 2 have significant practical implications for designing statistical learning methods for predicting trajectories. In particular, the boundedness of A means that the norm of $\mathcal{H}$ can be used as an effective error metric for estimators, allowing for the production of faithful approximations of the true velocity field. This is essential for accurate trajectory prediction. Additionally, the stability of the minimizer allows for the use of numerically stable procedures for obtaining faithful approximations of the minimizer from a finite amount of noisy data. Such results are crucial in establishing the statistical optimality of computation methods. Our findings provide a positive answer to the conjectured geometric coercivity condition in [MMQZ21] where the measure μ is the distribution of noise-free trajectory data that with randomized initial distributions, but our framework can be applied to any measure μ obtained in more general observation regime, for example, taking the observation noise into account [FRT21].

Related work and contribution Ordinary differential equations (ODEs) are a well-established tool for modeling dynamic processes with broad applications in various scientific fields. However, the explicit forms of these equations often remain elusive. The impressive advancement in data science and machine learning prompts researchers to devise methods originated from machine learning for data-driven discovery of dynamics. These examples include symbolic regression (e.g.[SL09]), sparse regression/optimization (e.g.[BPK16, STW18, MB21b]), kernel and Bayesian methods ([RPK17, HYM$^+$18, YWK21]), deep learning (e.g.[QWX19, DGYZ22]. The identifiability analysis is the first step in determining unknown parameters in ODE models. Previous work has mainly focused on parameter estimation. For instance, the identifiability analysis from single trajectory data for linear dynamical systems with linear parameters is performed in [SRS14], and later generalized to affine dynamical systems in [DRS20]. For nonlinear ODEs arising from biomedical applications, one can refer to the survey in [MXPW11] for identifiability analysis methodologies for nonlinear ODE models and references therein.

Our study focuses on the application of data-driven modeling to complex particle systems on manifolds, and aims to conduct identifiability analysis motivated by the usage of nonparametric statistical learning methods. Specifically, we extend previous work on identifying interaction kernels in Euclidean spaces, as presented in [LZTM19, LMT21, LLM$^+$21], to the more challenging manifold case. We explore how the geometry of the interacting domain affect the learning process. We adopt newly developed linear statistical inverse problem framework in [FRT21], which allows for the analysis of data distributions from a broad range of observation regimes, in contrast to the previous work that only considered noise-free observations. Our work makes the following contributions:

- We find the solutions of **Question** 1 and **Question** 2 on general manifolds, thereby providing a framework that includes the previous results as special cases.
- Using our analysis, we refine the generic results on specific manifolds, such as the sphere and Hyperbolic space, providing a more nuanced understanding of the behavior of these systems.
- Additionally, we investigate cases where the stability conditions may fail, deepening our understanding of the limitations and potential pitfalls of our approach.

Considering the manifold case has significant implications for various fields such as physics, biology, and social sciences, where systems can be modeled as interacting particles or agents on a manifold. Our approach provides a computational foundation for data-driven discovery of these systems, leading to new insights and discoveries in these fields.

2 Notation and Preliminaries

Notation: Let ν be a Borel positive measure on a domain $\mathbb{D}_1$. We use $L^2(\mathbb{D}_1;\nu;\mathbb{D}_2)$ to denote the set of $L^2(\nu)$-integrable vector-valued functions that map $\mathbb{D}_1$ to $\mathbb{D}_2$.

Let $\mathcal{S}_1$ be a measurable subset of $\mathbb{R}^D$, then the restriction of the measure ρ on $\mathcal{S}_1$, denoted by $\rho \llcorner \mathcal{S}_1$, is defined as $\rho \llcorner \mathcal{S}_1(\mathcal{S}_2) = \rho(\mathcal{S}_1 \cap \mathcal{S}_2)$ for any measurable subset $\mathcal{S}_2$ of $\mathbb{R}^D$. For two Borel positive measures ρ_1, ρ_2 defined on $\mathbb{R}^D$, ρ_1 is said to be absolutely continuous with respect to ρ_2, denoted by $\rho_1 \ll \rho_2$, if $\rho_1(\mathcal{S}) = 0$ for every set $\rho_2(\mathcal{S}) = 0$, $\mathcal{S} \subset \mathbb{R}^D$. ρ_1 and ρ_2 are called equivalent iff $\rho_1 \ll \rho_2$ and $\rho_2 \ll \rho_1$.

Let $\mathcal{H}_1, \mathcal{H}_2$ be Hilbert spaces. We use $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ to denote the inner product over $\mathcal{H}_1$, and still use $\langle \cdot, \cdot \rangle$ to denote the inner product on the Euclidean space. We denote by $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ the set of bounded linear operators mapping $\mathcal{H}_1$ to $\mathcal{H}_2$. Let $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$, we use $\|A\|$ to denote its operator norm.

2.1 Preliminaries in Geometry

We will use several geometric tools. For the convenience of the reader, we state them here. More details of these can be found in various books such as [Car92] and [GHL04], and we refer the reader to those for more details.

We let $(\mathcal{M}, g)$ denote a smooth, connected, and complete n-dimensional Riemannian manifold with Riemannian metric g. For any point $\boldsymbol{x} \in \mathcal{M}$, we let $T_{\boldsymbol{x}}\mathcal{M}$ be the tangent space at $\boldsymbol{x}$ and let

$$\exp_{\boldsymbol{x}} : T_{\boldsymbol{x}}\mathcal{M} \to \mathcal{M}$$

denote the *exponential map*. If $\gamma_{\boldsymbol{x},\mathbf{v}}$ is the geodesic starting at $\boldsymbol{x}$ with initial velocity $\mathbf{v}$, then $\exp_{\boldsymbol{x}}(\mathbf{v}) = \gamma_{\boldsymbol{x},\mathbf{v}}(1)$. Let $I(\boldsymbol{x}, \mathbf{v}) > 0$ be the first time when the geodesic $s \mapsto \exp_{\boldsymbol{x}}(s\mathbf{v})$ stops to be minimizing. Define

$$\mathcal{D}(\boldsymbol{x}) := \{t\mathbf{v} \in T_{\boldsymbol{x}}\mathcal{M} : \mathbf{v} \in \mathcal{S}^{n-1},\ 0 \le t < I(\boldsymbol{x}, \mathbf{v})\},$$

where $\mathcal{S}^{n-1}$ is the unit sphere in the tangent space with respect to the metric $g(\boldsymbol{x})$, then $\exp_{\boldsymbol{x}}$ is a diffeomorphism from $\mathcal{D}(\boldsymbol{x})$ onto its image, and

$$d(\boldsymbol{x}, \mathbf{p}) = |\exp_{\boldsymbol{x}}^{-1}(\mathbf{p})|$$

for any $\mathbf{p} \in \exp_{\boldsymbol{x}}(\mathcal{D}(\boldsymbol{x}))$. Recall that $d(\boldsymbol{x}, \mathbf{p})$ is the Riemannian distance between $\boldsymbol{x}$ and $\mathbf{p}$, which is obtained by minimizing the length of all piecewise C^1 curves connecting $\boldsymbol{x}$ and $\mathbf{p}$. Let $\mathcal{CL}(\boldsymbol{x}) := \exp_x(\partial\mathcal{D}(\boldsymbol{x})) \subset \mathcal{M}$ be the *cut locus* of $\boldsymbol{x}$, it is known that for connected complete Riemannian manifolds

$$\mathcal{M} = \exp_{\boldsymbol{x}}(\mathcal{D}(\boldsymbol{x})) \cup \mathcal{CL}(\boldsymbol{x})$$

for any $\boldsymbol{x} \in \mathcal{M}$, and $\mathcal{CL}(\boldsymbol{x})$ has zero measure, see, e.g. [[GHL04], page 158, Corolarry 3.77]. We denote $\mathrm{Inj}(\boldsymbol{x}) := d(\boldsymbol{x}, \mathcal{CL}(\boldsymbol{x})) > 0$ to be the *injectivity radius* at $\boldsymbol{x} \in \mathcal{M}$ (note that if $\mathcal{CL}(\boldsymbol{x}) = \emptyset$, $\mathrm{Inj}(\boldsymbol{x}) = \infty$), and let $\mathrm{Inj}(\mathcal{M}) = \inf_{\boldsymbol{x}\in\mathcal{M}} \mathrm{Inj}(\boldsymbol{x})$ be the injectivity radius of $\mathcal{M}$. If $\mathcal{M}$ is compact, then $\mathrm{Inj}(\mathcal{M})$ is strictly positive.

It is easy to check that $\mathcal{D}(\boldsymbol{x})$ is an open star-shaped subset of $T_{\boldsymbol{x}}\mathcal{M}$, so $(\mathcal{D}(\boldsymbol{x}), \exp_{\boldsymbol{x}})$ is a smooth chart and one has that for any Borel measurable function f on $\mathcal{M}$

$$\int_{\mathcal{M}} f\, dV_{\mathcal{M}} = \int_{\mathcal{D}(\boldsymbol{x})} (f \circ \exp_{\boldsymbol{x}}) \sqrt{\det(G_{\boldsymbol{x}})}\, d\lambda^n, \tag{4}$$

where $dV_{\mathcal{M}}$ is the Riemannian measure on the manifold $(\mathcal{M}, g)$, $G_{\boldsymbol{x}} := \exp_{\boldsymbol{x}}^* g$ is the pullback of g by $\exp_{\boldsymbol{x}}$, and λ^n denotes the n dimensional Lebesgue measure on $\mathbb{R}^n$. Notice that given coordinates $\{u^1, \cdots, u^n\}$ of $\mathcal{D}(\boldsymbol{x})$, for any $u \in \mathcal{D}(\boldsymbol{x})$

$$(\exp_{\boldsymbol{x}}^* g)_u(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}) = g_{\exp_{\boldsymbol{x}}(u)}(d\exp_{\boldsymbol{x}}\big|_u(\frac{\partial}{\partial u^i}), d\exp_{\boldsymbol{x}}\big|_u(\frac{\partial}{\partial u^j})),$$

where $d\exp_{\boldsymbol{x}}\big|_u : T_u\mathcal{D}(\boldsymbol{x}) \to T_{\exp_{\boldsymbol{x}}(u)}\mathcal{M}$ is the differential of the exponential map.

As an explicit example to illustrate these concepts, let us consider the sphere $\mathcal{M} = \mathbb{S}^n \subset \mathbb{R}^{n+1}$, where one finds that

$$T_{\boldsymbol{x}}\mathbb{S}^n = \{\mathbf{v} \in \mathbb{R}^{n+1} : \langle \boldsymbol{x}, \mathbf{v} \rangle = 0\} =: \boldsymbol{x}^{\perp},$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^{n+1}$. We choose the round metric on $\mathbb{S}^n$ which is the pullback of the Euclidean metric of $\mathbb{R}^{n+1}$. In this metric, one can compute the distance between two points as follows for any $\boldsymbol{x}, \mathbf{y} \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$, we have

$$d_{\mathbb{S}^n}(\mathbf{x}, \mathbf{y}) = \arccos(\langle \boldsymbol{x}, \mathbf{y} \rangle). \tag{5}$$

Note that for this metric the exponential chart can be computed explicitly, being

$$\exp_{\boldsymbol{x}}(\mathbf{v}) = \cos(\|\mathbf{v}\|)\mathbf{x} + \sin(\|\mathbf{v}\|)\frac{\mathbf{v}}{\|\mathbf{v}\|} \quad \text{for } \mathbf{v} \in T_{\boldsymbol{x}}\mathbb{S}^n = \boldsymbol{x}^{\perp}. \tag{6}$$

The cut domain is given by $\mathcal{D}(\boldsymbol{x}) = B_{\pi}(\mathbf{0})$ for any $\boldsymbol{x} \in \mathbb{S}^n$, i.e., the ball of radius π centered at origin. Writing the metric in polar coordinates $g_{\mathbb{S}^n} = dr^2 + \sin^2(r) g_{\mathbb{S}^{n-1}}$, we thus we get for any measurable f on $\mathbb{S}^n$ that

$$\int_{\mathbb{S}^n} f\, dV_{\mathbb{S}^n} = \int_0^{\pi} \int_{\mathbb{S}^{n-1}} f(\exp_{\mathbf{x}}(t\mathbf{v})) \sin^{n-1} d\sigma(\mathbf{v})\, dt, \tag{7}$$

where $\mathbb{S}^{n-1}$ denotes the $n-1$ dimensional unit sphere and $d\sigma$ denotes the induced surface measure.

The other simple case is the hyperbolic space, $\mathbb{H}^n = \{\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n : x_n > 0\}$ with the usual metric $g_{\mathbb{H}^n} = \frac{1}{x_n^2} g_{\mathrm{Euclid}}$. By the Hadamard theorem, we have that $\mathcal{CL}(\mathbf{p}) = \emptyset$ for any $\mathbf{p} \in \mathbb{H}^n$, and that $\mathcal{D}(\mathbf{p}) = \mathbb{R}^n$. The equation (7) for $\mathbb{H}^n$ is the same except one has to change the area of integration and replaces sin by sinh :

$$\int_{\mathbb{H}^n} f \, dV_{\mathbb{H}^n} = \int_0^\infty \int_{\mathbb{S}^{n-1}} f(\exp_{\mathbf{x}}(t\mathbf{v}))\sinh^{n-1} d\sigma(\mathbf{v})\, dt. \tag{8}$$

2.2 Interactions on manifold

Following the work in [MMQZ21, AMD17], we choose $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ to be the geodesic distance of any two points $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{M}$. Notice that we can connect $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ with a minimizing geodesic which is unique if they are not cut point of each other. In view of that we define the influence vector as

$$w(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} d(\boldsymbol{x}_i, \boldsymbol{x}_j)\dot{\gamma}_{\boldsymbol{x}_i,\boldsymbol{x}_j}(0) & \text{if } \boldsymbol{x}_i \neq \boldsymbol{x}_j, \text{ and } \boldsymbol{x}_i \notin \mathcal{CL}(\boldsymbol{x}_j) \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $\gamma_{\boldsymbol{x}_i,\boldsymbol{x}_j} : [0, d(\boldsymbol{x}_i, \boldsymbol{x}_j)] \to \mathcal{M}$ denotes the unique normalized minimizing geodesic connecting $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. So $\dot{\gamma}_{\boldsymbol{x}_i,\boldsymbol{x}_j}(0)$ is the unit tangent vector to the geodesic $\gamma_{\boldsymbol{x}_i,\boldsymbol{x}_j}$ at $\boldsymbol{x}_i$. In particular, $w(\boldsymbol{x}, \boldsymbol{y}) = \exp_{\boldsymbol{x}}^{-1}(\boldsymbol{y})$ for $\boldsymbol{y} \notin \mathcal{CL}(\boldsymbol{x})$.

Remark 1. *We add several comments on the first-order models.*

- *It is sometimes conventional to consider the unit norm influence vector by moving the term $d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ in (9) to the interaction part, so the interaction kernel becomes $\phi(\cdot)\cdot$. The analysis developed in this paper can be equivalently applied to this case as well by slight modification on the function spaces.*

- *When $\mathcal{M} = \mathbb{R}^n$ or $\mathbb{S}^n$, we can calculate the influence vector as follows*

$$w(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} \Pi_{T_{\boldsymbol{x}_i}\mathcal{M}}(\boldsymbol{x}_j - \boldsymbol{x}_i) & \text{if } \Pi_{T_{\boldsymbol{x}_i}\mathcal{M}}(\boldsymbol{x}_j - \boldsymbol{x}_i) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

  where $\Pi_{T_{\mathbf{x}_i}\mathcal{M}}$ denotes the orthogonal projection onto the tangent space $T_{\boldsymbol{x}_i}\mathcal{M}$. In particular, when $\mathcal{M} = \mathbb{R}^n$, one has that the vectors of influence are simply

$$w(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_j - \boldsymbol{x}_i.$$

  This recovers the classical first order models from [MT14].

- *In [AMD17], the well-posedness of (1) is examined for compact manifolds. The well-posedness for the mean-field equation of (1) is examined in [FZ18] on two dimensional sphere and two dimensional hyperbolic spaces. It is established that if the support of the interaction kernel is smaller than the injective radius, then (1) is well-posed. However, it is worth noting that our analysis does not necessitate the manifold being compact.*

We let $\mathcal{M}^N = \mathcal{M} \times \mathcal{M} \times \cdots \times \mathcal{M}$ be the canonical product of Riemannian manifolds with product Riemannian metric given by $g_{\mathcal{M}}^N$. Denote by

$$\boldsymbol{X} := \begin{bmatrix} \vdots \\ \boldsymbol{x}_i \\ \vdots \end{bmatrix} \in \mathcal{M}^N, \tag{10}$$

the tangent space at $\boldsymbol{X}$ is defined by $T_{\mathbf{X}}\mathcal{M}^N = T_{\boldsymbol{x}_1}\mathcal{M} \times \cdots \times T_{\boldsymbol{x}_N}\mathcal{M}$ such that

$$\left\langle \begin{bmatrix} \vdots \\ \mathbf{u}_i \\ \vdots \end{bmatrix}, \begin{bmatrix} \vdots \\ \mathbf{v}_i \\ \vdots \end{bmatrix} \right\rangle_{T_{\mathbf{X}}\mathcal{M}^N} := \frac{1}{N}\sum_{i=1}^{N} g\left(\mathbf{u}_i, \mathbf{v}_i\right)$$

which is the weighted product metric induced by the metric g.

2.3 Preliminaries on probability

We say that N random vectors $X_1, \cdots, X_N$ is **conditional i.i.d** if there exists a random vector $\theta \sim \pi(\theta)$ such that $X_1|\theta, \cdots, X_N|\theta$ are independent and identically distributed. Then

$$p(X_1, \cdots, X_N) = \int p(X_1|\theta)p(X_2|\theta)\cdots p(X_N|\theta)\pi(\theta)d\theta,$$

It is also called a **mixture** of i.i.d. For example, $X_1, \cdots, X_N$ is conditional i.i.d when they are from an infinite exchangeable sequence, which is a consequence of the celebrated de Finetti's representation theorem. It also includes N i.i.d random vectors as a special example.

3 On the Boundedness of A

We begin by defining a measure ρ on the Borel σ-algebra of $\mathbb{R}^+$, which addresses Question 1. Given the measure μ defined on $\mathcal{M}^N$, which we throughout the paper, if not specified otherwise, assume to be absolutely continues with respect to the volume formthe measure ρ yields the function space $\mathcal{H} = L^2([0,R]; \rho \llcorner [0,R]; \mathbb{R})$. We will show that the operator A is bounded from $\mathcal{H}$ to $\mathcal{F}$.

Definition 1. *Let $\mathbf{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N) \in \mathcal{M}^N$ and let $B \subset \mathbb{R}^+$ be a Borel set. Then, we define the positive measure ρ as follows:*

$$\rho(B) = \frac{1}{N(N-1)}\sum_{i\neq j}\int_{P_{ij}^{-1}(B)} P_{ij}^2(\mathbf{X})d\mu(\mathbf{X}),$$

where the function $P_{ij}(\mathbf{X}) := d(\boldsymbol{x}_i, \boldsymbol{x}_j)$ maps $\mathcal{M}^N$ to $\mathbb{R}^+$. In other words, $\rho(B)$ is the average of measures that are push-forwards of μ by the distance map P_{ij}^2.

The measure ρ serves to capture the structural information of the governing equations (1) by quantifying the extent to which regions of $\mathbb{R}^+$ are explored through samples from $\mathbf{f}_\phi(\boldsymbol{X})$ at the population level. By virtue of being the average of measures that are push-forwards of μ by the distance map P_{ij}^2, ρ inherits the finer properties of μ. For instance, if μ is a Borel regular measure on $\mathcal{M}^N$, then ρ is also Borel regular on $\mathbb{R}^+$. Given our assumption that μ is absolutely continuous with respect to the volume measure $V_{\mathcal{M}^N}$ of $\mathcal{M}^N$, it follows that ρ enjoys the same property, as succinctly outlined below:

Lemma 3.1. *The measure ρ in Definition 1 is absolute continuous with respect to the Lebesgue measure of $\mathbb{R}_+$.*

Proof. It suffices to show that for any i, j, the measure ρ_{ij} given by

$$\rho_{ij}(A) = \int_{P_{ij}^{-1}(A)} P_{ij}^2(\boldsymbol{X})\, d\mu(\boldsymbol{X})$$

is absolutely continuous with respect to $|\cdot|$, the one dimensional Lebesgue measure. Let $A \subset \mathbb{R}^+$ satisfying $|A| = 0$. We denote $A^2 = \{x^2 : x \in A\}$, then $|A^2| = 0$ too. Notice that $P_{ij}^2 : \mathcal{M}^N \to \mathbb{R}^+$ is smooth, and the set of critical points (i.e. with zero gradient) of P_{ij}^2 is

$$\{\mathbf{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N) \in \mathcal{M}^N : \boldsymbol{x}_i = \boldsymbol{x}_j\},$$

which is of zero measure, therefore $V_{\mathcal{M}^N}((P_{ij}^2)^{-1}(A^2)) = 0$ whenever $|A^2| = 0$. It follows that $V_{\mathcal{M}^N}(P_{ij}^{-1}(A)) = V_{\mathcal{M}^N}((P_{ij}^2)^{-1}(A^2)) = 0$. Since μ is absolutely continuous with respect to $V_{\mathcal{M}^N}$, we obtain that $\mu(P_{ij}^{-1}(A)) = 0$, thus $\rho_{ij}(A) = 0$. This shows that ρ is absolute continuous with respect to the Lebesgue measure $|\cdot|$ on $\mathbb{R}^+$. □

In summary, the measure ρ is derived from μ and shares numerous properties with it, rendering it a valuable instrument in analyzing the operator A and characterizing the function space $\mathcal{H}$. With this groundwork laid, we are prepared to present the key outcome of this section.

Proposition 3.2. *Let $\mathcal{H}$ be a subspace of $L^2([0, R]; \rho \llcorner [0, R]; \mathbb{R})$ or itself. Then $A \in \mathcal{B}(\mathcal{H}, \mathcal{F})$ and $\|A\| \leq \frac{N-1}{N}$.*

Proof. It suffices to prove this result when $\mathcal{H} := L^2([0, R]; \rho \llcorner [0, R]; \mathbb{R})$. Observe that for any $\varphi \in \mathcal{H}$, one has that

$$\|\mathbf{f}_\varphi\|_{L^2(\mu)}^2 = \frac{1}{N} \sum_{i=1}^{N} \int_{\mathcal{M}^N} \|\frac{1}{N} \sum_{j \neq i} \varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_j))\omega(\boldsymbol{x}_i, \boldsymbol{x}_j)\|^2 \, d\mu(\boldsymbol{X}).$$

We now estimate the integrand using Cauchy-Schwartz and Young's inequality: for any $i = 1, \ldots, N$

$$\begin{aligned}
\|\frac{1}{N} \sum_{j \neq i} \varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_j))\omega(\boldsymbol{x}_i, \boldsymbol{x}_j)\|^2 &= g\Big(\frac{1}{N} \sum_{j \neq i} \varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_j))\omega(\boldsymbol{x}_i, \boldsymbol{x}_j), \frac{1}{N} \sum_{k \neq i} \varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_k))\omega(\boldsymbol{x}_i, \boldsymbol{x}_k)\Big) \\
&= \frac{1}{N^2} \sum_{j \neq i} \sum_{k \neq i} \varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_j))\varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_k)) g\left(\omega(\boldsymbol{x}_i, \boldsymbol{x}_j), \omega(\boldsymbol{x}_i, \boldsymbol{x}_k)\right) \\
&\leq \frac{1}{N^2} \sum_{j \neq i} \sum_{k \neq i} \big|\varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_j))\varphi(d(\boldsymbol{x}_i, \boldsymbol{x}_k))\big| d(\boldsymbol{x}_i, \boldsymbol{x}_j) d(\boldsymbol{x}_i, \boldsymbol{x}_k) \\
&\leq \frac{1}{2N^2} \sum_{j \neq i} \sum_{k \neq i} \varphi^2(d(\boldsymbol{x}_i, \boldsymbol{x}_j)) d^2(\boldsymbol{x}_i, \boldsymbol{x}_j) + \varphi^2(d(\boldsymbol{x}_i, \boldsymbol{x}_k)) d^2(\boldsymbol{x}_i, \boldsymbol{x}_k) \\
&= \frac{N-1}{N^2} \sum_{j \neq i} \varphi^2(d(\boldsymbol{x}_i, \boldsymbol{x}_j)) d^2(\boldsymbol{x}_i, \boldsymbol{x}_j).
\end{aligned}$$

From this estimate we get that

$$\|\mathbf{f}_\varphi\|_{L^2(\mu)}^2 \leq \frac{N-1}{N^3} \sum_{i \neq j} \int_{\mathcal{M}^N} \varphi^2(d(\boldsymbol{x}_i, \boldsymbol{x}_j)) d^2(\boldsymbol{x}_i, \boldsymbol{x}_j) \, d\mu(\boldsymbol{X})$$

Moreover, one has that

$$\|\varphi\|_{L^2(\rho)}^2 = \frac{1}{N(N-1)} \sum_{i \neq j} \int_{\mathcal{M}^N} \varphi^2(d(\boldsymbol{x}_i, \boldsymbol{x}_j)) d^2(\boldsymbol{x}_i, \boldsymbol{x}_j) \, d\mu(\boldsymbol{X}),$$

from which we get that

$$\|\mathbf{f}_\varphi\|^2_{L^2(\mu)} = \|A\varphi\|^2_{L^2(\mu)} \leq \frac{(N-1)^2}{N^2}\|\varphi\|^2_{L^2(\rho)},$$

as desired. $\square$

The above theorem establishes that the L^2 norm induced by ρ is a potent metric for data-driven estimators. Estimators that yield small $L^2(\rho)$ errors provide accurate approximations of the velocity field, thereby guaranteeing reliable trajectory prediction outcomes. Notably, the $L^2(\rho)$ norm has been utilized as an evaluation metric for the generalization error of least square estimators in [MMQZ21].

4 Well-posedness

In this section, we present stability results for a specific family of measures, namely the joint distribution of N independent and identically distributed random variables on $\mathcal{M}$. In the forward problem, it is often assumed that the initial conditions of particles are i.i.d. because these particles are indistinguishable. Under certain conditions, such as the absence of significant interactions between the particles and N is sufficiently large, the system may remain approximately i.i.d. over time. In our analysis, we leverage the i.i.d assumption to facilitate computation and prove that the associated integral operator is positive and even strictly positive. We later demonstrate that the stability results hold more generally for the joint distribution of N conditionally i.i.d. random variables by leveraging De Finetti's representation theorem. Finally, we show that the stability results hold for their equivalent measures in Corollary 4.5.

Proposition 4.1. *Suppose that μ is the joint distribution of N i.i.d random variables. Then we have that $\forall \varphi \in \mathcal{H} = L^2([0,R]; \rho \llcorner [0,R]; \mathbb{R})$*

$$\|A\varphi\|_{L^2(\mu)} \geq \frac{\sqrt{N-1}}{N}\|\varphi\|_{L^2(\rho)}$$

Proof. We calculate that

$$\begin{aligned}
\|A\phi\|^2_{L^2(\mu)} &= \mathbb{E}_{\boldsymbol{X}\sim\mu}\left[\|\mathbf{f}_\phi(\boldsymbol{X})\|^2_{T_{\boldsymbol{X}}\mathcal{M}^N}\right] \\
&= \mathbb{E}_{\boldsymbol{X}\sim\mu}\sum_{i=1}^N \frac{1}{N} g\left(\frac{1}{N}\sum_{j\neq i}\phi(P_{ij}(\boldsymbol{X}))\omega(\boldsymbol{x}_i,\boldsymbol{x}_j), \frac{1}{N}\sum_{k\neq i}\phi(P_{ik}(\boldsymbol{X}))\omega(\boldsymbol{x}_i,\boldsymbol{x}_k)\right) \\
&= \frac{1}{N^3}\sum_{j\neq i}\mathbb{E}_{\boldsymbol{X}\sim\mu}\Big[\phi^2(P_{ij}(\boldsymbol{X}))P^2_{ij}(\boldsymbol{X})\Big] \\
&+ \frac{1}{N^3}\sum_{j\neq i, k\neq j, k\neq i}\mathbb{E}_{\boldsymbol{X}\sim\mu}\Big[\phi(P_{ij}(\boldsymbol{X}))\phi(P_{ik}(\boldsymbol{X}))g\Big(\omega(\boldsymbol{x}_i,\boldsymbol{x}_j),\omega(\boldsymbol{x}_i,\boldsymbol{x}_k)\Big)\Big],
\end{aligned}$$

where we used the fact that in view of (9), one has that $\|\omega(\boldsymbol{x}_i,\boldsymbol{x}_j)\|^2 = P^2_{ij}(\boldsymbol{X})$. Note that

$$\begin{aligned}
&\frac{1}{N^3}\sum_{j\neq i}\mathbb{E}_{\boldsymbol{X}\sim\mu}\Big[\phi^2(P_{ij}(\boldsymbol{X}))P^2_{ij}(\boldsymbol{X})\Big] \\
&= \frac{N(N-1)}{N^3}\int_{\mathcal{M}^N}\frac{1}{N(N-1)}\sum_{i\neq j}\phi^2(P_{ij}(\boldsymbol{X}))P^2_{ij}(\boldsymbol{X})\,d\mu(\boldsymbol{X}) \\
&= \frac{N-1}{N^2}\|\phi\|^2_{L^2(\rho)}.
\end{aligned}$$

This gives that

$$
\begin{aligned}
&\|A\phi\|^2_{L^2(\mu)} \\
&= \frac{N-1}{N^2}\|\phi\|^2_{L^2(\rho)} + \frac{1}{N^3}\sum_{i\neq j\neq k}\mathbb{E}_{\boldsymbol{X}\sim\mu}\Big[\phi(P_{ij}(\boldsymbol{X}))\phi(P_{ik}(\boldsymbol{X}))g\Big(\omega(\boldsymbol{x}_i,\boldsymbol{x}_j),\omega(\boldsymbol{x}_i,\boldsymbol{x}_k)\Big)\Big].
\end{aligned}
$$

It therefore suffices to show that for $i \neq j \neq k$ (any three non-equal indices)

$$
C_{ijk} = \mathbb{E}_{\boldsymbol{X}\sim\mu}\Big[\phi(P_{ij}(\boldsymbol{X}))\phi(P_{ik}(\boldsymbol{X}))g\Big(\omega(\boldsymbol{x}_i,\boldsymbol{x}_j),\omega(\boldsymbol{x}_i,\boldsymbol{x}_k)\Big)\Big] \geq 0.
$$

Due to the special structure of μ, i.e. $\mu = \mu_0 \times \cdots \times \mu_0$ with μ_0 the same probability measure for each $\boldsymbol{x}_i$, $i = 1, \ldots, N$, the values of C_{ijk}'s are equal and therefore we just need to show it is positive for one such term. Without loss of generality, let $i = 1$, $j = 2$ and $k = 3$, and we further simplify the notations by denoting $\boldsymbol{x} := \boldsymbol{x}_1$, $\boldsymbol{y} := \boldsymbol{x}_2$ and $\boldsymbol{z} := \boldsymbol{x}_3$. We then get that

$$
\begin{aligned}
C_{123} &= \mathbb{E}_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}}\Big[g\Big(\phi(d(\boldsymbol{x},\boldsymbol{y}))w(\boldsymbol{x},\boldsymbol{y}),\phi(d(\boldsymbol{x},\boldsymbol{z}))w(\boldsymbol{x},\boldsymbol{z})\Big)\Big] \\
&= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{y},\boldsymbol{z}}\Big[\phi(d(\boldsymbol{x},\boldsymbol{y}))\phi(d(\boldsymbol{x},\boldsymbol{z}))g(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z}))\Big|\boldsymbol{x}\Big],
\end{aligned}
\tag{11}
$$

where the second equality is true due to the independence of $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$.

It therefore suffices to show that for any fixed $\boldsymbol{x}$, we have that

$$
\mathbb{E}_{\boldsymbol{y},\boldsymbol{z}}\Big[\phi(d(\boldsymbol{x},\boldsymbol{y}))\phi(d(\boldsymbol{x},\boldsymbol{z}))g(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z}))\Big|\boldsymbol{x}\Big] \geq 0.
$$

Now for each fixed $\boldsymbol{x}$, $\phi(d(\boldsymbol{x},\boldsymbol{y}))\phi(d(\boldsymbol{x},\boldsymbol{z}))$ and $g(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z}))$ are positive definite kernels with respect to $(\boldsymbol{y},\boldsymbol{z}) \in \mathcal{M}\times\mathcal{M}$ by Theorem 7.1 (2)(3) (notice that g is positive definite on $T_{\boldsymbol{x}}\mathcal{M}\times T_{\boldsymbol{x}}\mathcal{M}$), therefore their product is a positive definite kernel by Theorem 7.1 (1). Then by (4) of Theorem 7.1, the conclusion follows. □

Remark 2. *In fact, the same analysis can be extended to cover the case of N conditionally i.i.d. random variables, by invoking de Finetti's representation theorem. Specifically, in* (11)*,given three conditionally i.i.d random variables $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, one can do*

$$
C_{123} = E_\theta E_{\boldsymbol{x}|\theta,\boldsymbol{y}|\theta,\boldsymbol{z}|\theta}\Big[g\Big(\phi(d(\boldsymbol{x}|\theta,\boldsymbol{y}|\theta))w(\boldsymbol{x}|\theta,\boldsymbol{y}|\theta),\phi(d(\boldsymbol{x}|\theta,\boldsymbol{z}|\theta))w(\boldsymbol{x}|\theta,\boldsymbol{z}|\theta)\Big)\Big|\theta\Big]
\tag{12}
$$

where θ is the latent random variable conditioning on which that $\boldsymbol{x}|\theta$, $\boldsymbol{y}|\theta$, $\boldsymbol{z}|\theta$ are independent. Then with a slight modification on the proof, we can show that $C_{123} \geq 0$.

4.1 Discussion on the sharpness of the bound $\frac{\sqrt{N-1}}{N}$ on specific manifolds

The bound $\frac{\sqrt{N-1}}{N}$ has a drawback in that it decreases as N increases, which may appear counterintuitive at first. This is because, as N increases, the dimensionality of the data also increases while keeping ϕ as a radial kernel. Intuitively, this implies that we can acquire more information about ϕ. In this section, we explore whether this bound is sharp or can be enhanced by either confining $\mathcal{H}$ to a smaller subspace, such as a compact subspace of $L^2([0,R];\rho\llcorner[0,R];\mathbb{R})$, or by considering specific manifolds that enable us to refine the outcomes. We first discuss a criterion to decide whether, given a compact subset $\mathcal{H} \subset L^2([0,R];\rho\llcorner[0,R];\mathbb{R})$, there is going to be a larger bound in general or not, when restricting the map A to it.

Theorem 4.2. *Suppose that μ is the joint distribution of N i.i.d random variables on $\mathcal{M}$. Let $\mathcal{H} \subset L^2([0,R];\rho;\mathbb{R})$ be a compact subspace (i.e., the unit ball of $\mathcal{H}$ is compact with respect to the*

$L^2(\rho)$ *norm) consisting of continuous functions. If for any nonzero* $\varphi \in \mathcal{H}$, *there exists some* $\boldsymbol{x} \in \operatorname{supp}(\mu) \subset \mathcal{M}$ *such that*

$$\mathbb{E}_{\boldsymbol{y},\boldsymbol{z}}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(\omega(\boldsymbol{x},\boldsymbol{y}),\omega(\boldsymbol{x},\boldsymbol{z})\Big)\Big|\boldsymbol{x}\Big] > 0, \tag{13}$$

Then one has that

$$\inf_{\varphi\in\mathcal{H},\varphi\neq 0}\frac{\|A\varphi\|^2_{L^2(\mu)}}{\|\varphi\|^2_{L^2(\rho)}} \geq \frac{N-1}{N^2} + c_{\mathcal{H},\mathcal{M}}\frac{(N-1)(N-2)}{N^2},$$

where

$$c_{\mathcal{H},\mathcal{M}} := \inf_{\varphi\in\mathcal{H},\|\varphi\|_{L^2(\rho)}=1}\mathbb{E}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\Big)\Big] > 0. \tag{14}$$

Proof. We divide the proof into several steps.

Step 1. We claim that the function

$$\boldsymbol{x} \mapsto \mathbb{E}_{\boldsymbol{y},\boldsymbol{z}}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(\omega(\boldsymbol{x},\boldsymbol{y}),\omega(\boldsymbol{x},\boldsymbol{z})\Big)\Big|\boldsymbol{x}\Big]$$

is continuous. Take a sequence $\{\boldsymbol{x}_n\} \subset \mathcal{M}$ such that $\lim_{n\to\infty}\boldsymbol{x}_n = \boldsymbol{x}$. Due to the continuity of φ, the (Lipschitz) continuity of $d(\cdot,\boldsymbol{y})$, and $d(\cdot,\boldsymbol{z})$, respectively, we have $\varphi(d(\boldsymbol{x}_n,\boldsymbol{y}))\varphi(d(\boldsymbol{x}_n,\boldsymbol{z})) \to \varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))$ as $n\to\infty$ for each $\boldsymbol{y},\boldsymbol{z}\in\mathcal{M}$.

Recall that $w(\boldsymbol{x},\boldsymbol{y}) = \exp^{-1}_{\boldsymbol{x}}(\boldsymbol{y})$, $w(\boldsymbol{x},\boldsymbol{z}) = \exp^{-1}_{\boldsymbol{x}}(\boldsymbol{z})$ for $\boldsymbol{y},\boldsymbol{z} \notin \mathcal{CL}(\boldsymbol{x})$. Moreover, if $\boldsymbol{y},\boldsymbol{z} \notin \mathcal{CL}(\boldsymbol{x})$, then $\boldsymbol{y},\boldsymbol{z} \notin \mathcal{CL}(\boldsymbol{x}_n)$ for n sufficiently large. Therefore, $\exp^{-1}_{\boldsymbol{x}_n}(\boldsymbol{y}) \to \exp^{-1}_{\boldsymbol{x}}(\boldsymbol{y})$, $\exp^{-1}_{\boldsymbol{x}_n}(\boldsymbol{z}) \to \exp^{-1}_{\boldsymbol{x}}(\boldsymbol{z})$ as $n\to\infty$, for $\boldsymbol{y},\boldsymbol{z} \notin \mathcal{CL}(\boldsymbol{x})$. Note that $\mathcal{CL}(\boldsymbol{x})$ has zero measure with respect to μ_0, it follows that $g(w(\boldsymbol{x}_n,\boldsymbol{y}),w(\boldsymbol{x}_n,\boldsymbol{z})) \to g(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z}))$ a.e. on $\mathcal{M}\times\mathcal{M}$ with respect to the measure $\mu_0\times\mu_0$. The claim then follows from dominated convergence theorem as the integrand is continuous in $\boldsymbol{x}$ for almost every $(\boldsymbol{y},\boldsymbol{z}) \in \mathcal{M}\times\mathcal{M}$.

Step 2. Consider the bilinear map

$$G : L^2([0,R];\rho \llcorner [0,R];\mathbb{R}) \times L^2([0,R];\rho \llcorner [0,R];\mathbb{R}) \to \mathbb{R}$$

defined by

$$G(\varphi,\psi) := \mathbb{E}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\psi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\Big)\Big].$$

We claim that the map

$$L^2([0,R];\rho \llcorner [0,R];\mathbb{R}) \to \mathbb{R}, \quad \varphi \mapsto G(\varphi,\varphi) := \mathbb{E}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\Big)\Big] \tag{15}$$

is continuous. Using Cauchy–Schwarz inequality

$$|G(\varphi,\psi)| \leq \|\varphi\|_{L^2(\rho)}\|\psi\|_{L^2(\rho)}$$

Then notice that

$$|G(\varphi_n,\varphi_n) - G(\varphi,\varphi)| \leq \|\varphi_n - \varphi\|_{L^2(\rho)}\|\varphi\|_{L^2(\rho)} + \|\varphi_n - \varphi\|_{L^2(\rho)}\|\varphi_n\|_{L^2(\rho)}.$$

Let $\varphi_n \to \varphi$ in $\mathcal{H}$ endowed with $L^2(\rho)$ norm, and then we know if n is large enough, we have $\|\varphi_n\|_{L^2(\rho)} \leq 2\|\varphi\|_{L^2(\rho)}$. Consequently, we have

$$\lim_{n\to\infty} G(\varphi_n,\varphi_n) = G(\varphi,\varphi).$$

Step 3. Now recall that

$$c_{\mathcal{H},\mathcal{M}} := \inf_{\varphi\in\mathcal{H},\|\varphi\|_{L^2(\rho)}=1} \mathbb{E}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\Big)\Big] \tag{16}$$

$$= \inf_{\varphi\in\mathcal{H},\|\varphi\|_{L^2(\rho)}=1} \mathbb{E}_{\boldsymbol{x}}\mathbb{E}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\Big)\Big|\boldsymbol{x}\Big] \tag{17}$$

and

$$\mathbb{E}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g\Big(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\Big)\Big|\boldsymbol{x}\Big] \geq 0$$

by the proof of Proposition 4.1.

By the property (13) and compactness of $\mathcal{H}$, the continuity of the conditional expectation w.r.t $\boldsymbol{x}$ and the continuity of the map (15) ensures that $c_{\mathcal{H},\mathcal{M}}$ is positive. Then using (11) in the proof of Proposition 4.1, we see that this implies

$$\|A\varphi\|^2_{L^2(\rho)} \geq \frac{N-1}{N^2}\|\varphi\|^2_{L^2(\rho)} + c_{\mathcal{H},\mathcal{M}}\frac{N(N-1)(N-2)}{N^3}\|\varphi\|^2_{L^2(\rho)}$$

which gives the result. □

Remark 3. *Theorem 4.2 provides a way to explore whether we can improve the bound by restricting on a smaller compact subspaces of $L^2([0,R];\rho\llcorner[0,R];\mathbb{R})$. Using exponential map, one can easily check the condition (13).*

Specifically, let $\varphi\in L^2(\rho)$ and $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mu_0$ i.i.d with densitiy function $p(\cdot)$. We denote $u:=\omega(\boldsymbol{x},\boldsymbol{y})$ and $v:=\omega(\boldsymbol{x},\boldsymbol{z})$. Notice that for $\boldsymbol{y}\,,\boldsymbol{z}\notin\mathcal{CL}(\boldsymbol{x})$,

$$u=\exp^{-1}_{\boldsymbol{x}}(\boldsymbol{y}) \quad \text{and} \quad v=\exp^{-1}_{\boldsymbol{x}}(\boldsymbol{z}),$$

so $\|u\|=d(\boldsymbol{x},\boldsymbol{y})$ and $\|v\|=d(\boldsymbol{x},\boldsymbol{z})$. Moreover, since the volume measure of $\mathcal{CL}(\boldsymbol{x})$ is zero, $\mu_0(\mathcal{CL}(\boldsymbol{x}))=0$ too. In view of (4), we obtain that

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{y},\boldsymbol{z}}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z}))\Big|\boldsymbol{x}\Big]\\
&=\int_{\mathcal{M}}\int_{\mathcal{M}}\varphi(d(\boldsymbol{x},\mathbf{y}))\varphi(d(\boldsymbol{x},\mathbf{z}))g_{\boldsymbol{x}}\left(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z})\right)p(\boldsymbol{y})p(\boldsymbol{z})\,dV_{\mathcal{M}}(\boldsymbol{y})dV_{\mathcal{M}}(\boldsymbol{z})\\
&=\int_{\mathcal{D}(\boldsymbol{x})}\int_{\mathcal{D}(\boldsymbol{x})}\varphi(\|u\|)\varphi(\|v\|)g(u,v)p(\exp_{\boldsymbol{x}}(u))p(\exp_{\boldsymbol{x}}(v))\sqrt{\det(G_{\boldsymbol{x}}(u))}\sqrt{\det(G_{\boldsymbol{x}}(v))}\,d\lambda^n(u)d\lambda^n(v).
\end{aligned}$$

We choose an orthonormal basis $(\mathbf{e}_i)_{i=1}^n$ at $\boldsymbol{x}$ with respect to the metric g on $T_{\boldsymbol{x}}\mathcal{M}$, so $u=\sum_{i=1}^n u^i\mathbf{e}_i$, $v=\sum_{i=1}^n v^i\mathbf{e}_i$ and

$$g_{\boldsymbol{x}}(u,v)=\sum_{i=1}^n u^i v^i.$$

Then

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{y},\boldsymbol{z}}\Big[\varphi(d(\boldsymbol{x},\boldsymbol{y}))\varphi(d(\boldsymbol{x},\boldsymbol{z}))g(w(\boldsymbol{x},\boldsymbol{y}),w(\boldsymbol{x},\boldsymbol{z}))\Big|\boldsymbol{x}\Big]\\
&=\sum_{i=1}^n\Big[\int_{\mathcal{D}(\boldsymbol{x})}u^i\varphi(\|u\|)p(\exp_{\boldsymbol{x}}(u))\sqrt{\det(G_{\boldsymbol{x}}(u))}\,d\lambda^n(u)\Big]^2\geq 0
\end{aligned}$$

It thus suffices to find $\boldsymbol{x}\in\operatorname{supp}(\mu_0)$ such that

$$\int_{\mathcal{D}(\boldsymbol{x})}u^i\varphi(\|u\|)p(\exp_{\boldsymbol{x}}(u))\sqrt{\det(G_{\boldsymbol{x}}(u))}\,d\lambda^n(u)\neq 0 \tag{18}$$

for some $i = 1, \dots, n$.

The reference [MMQZ21] provides numerical examples of first-order opinion dynamics and Lennard-Jones dynamics on both the sphere and the Poincaré disk. These examples motivated our exploration of refining stability results on these spaces, as well as on hyperbolic space more generally.

4.1.1 Sphere

We first consider $\mathcal{M} = \mathbb{S}^n$, and $\mu = \mathrm{Unif}((\mathbb{S}^n)^N)$ is the uniform distribution on sphere.

Proposition 4.3. *If* $\mathcal{M} = \mathbb{S}^n$ *and we choose* $\mu = \mathrm{Unif}((\mathbb{S}^n)^N)$*. Then one has that*

$$\inf_{\varphi \in \mathcal{H}, \varphi \neq 0} \frac{\|A\varphi\|_{L^2(\mu)}}{\|\varphi\|_{L^2(\rho)}} = \frac{\sqrt{N-1}}{N}$$

for any nontrivial subspace $\mathcal{H} \subset L^2([0,R]; \rho \llcorner [0,R]; \mathbb{R})$.

Proof. Since we consider the uniform distribution, the density function of each particle is $p(\mathbf{u}) = \frac{1}{\mathrm{Vol}(\mathbb{S}^n)}$. Recall that in the sphere case, we have that

$$\begin{aligned} &\int_{D(\boldsymbol{x})} \mathbf{u}_i \varphi(\|\mathbf{u}\|) p(\exp_{\mathbf{x}}(\mathbf{u})) \sqrt{G_{\mathbf{x}}(\mathbf{u})}\, d\mathbf{u} \\ &= \frac{1}{\mathrm{Vol}(\mathbb{S}^n)} \int_0^\pi \sin^{n-1}(t) t \varphi(t)\, dt \int_{\mathbb{S}^{n-1}} \mathbf{u}_i \, d\sigma(\mathbf{u})\, dt = 0 \end{aligned}$$

for any $i = 1, \dots, n$, as the functions $\mathbf{u}_i$ are odd and domain of integration is symmetric. The claim then follows from the computation done in Remark 3. □

We see that in the sphere case, (13) is always zero. As a consequence, on any compact subsets $\mathcal{H} \subset L^2([0,R]; \rho \llcorner [0,R]; \mathbb{R})$, we find that the $c_{\mathcal{H},\mathcal{M}}$ is equal to zero and consequently, the bound $\frac{\sqrt{N-1}}{N}$ is sharp.

4.1.2 Hyperbolic space

Next we show a positive result on the hyperbolic space $\mathbb{H}^n$, where μ is a product of uniform distribution on a ball of radius $R_0 > 0$. In this case, we can establish a better bound, depending on the compact set $\mathcal{H} \subset L^2(\rho) \cap C_c^1[0,\infty)$.

Proposition 4.4. *Let* $R_0 > 0$, $\mathbf{p} \in \mathcal{M} := \mathbb{H}^n$ *(see definition in section 2) and* μ *have the density function* $\frac{1}{Vol_{\mathbb{H}^n}(B_{R_0}(\mathbf{p}))^N} 1_{B_{R_0}(\mathbf{p}) \times \cdots \times B_{R_0}(\mathbf{p})}(\boldsymbol{X})$*. Then for any* $\mathcal{H} \subset L^2([0,R]; \rho \llcorner [0,R]; \mathbb{R}) \cap C_c^1[0,\infty)$ *compact (with respect to the topology of* $L^2(\rho)$*), we get that*

$$\inf_{\varphi \in \mathcal{H}, \varphi \neq 0} \frac{\|A\varphi\|^2_{L^2(\mu)}}{\|\varphi\|^2_{L^2(\rho)}} \geq \frac{N-1}{N^2} + c_{\mathcal{H},\mathcal{M}} \frac{(N-1)(N-2)}{N^2}.$$

where $c_{\mathcal{H},\mathcal{M}} > 0$ *is defined in* (14).

Proof. The density function $p(\boldsymbol{x})$ for marginal distribution of each particle is given by

$$p(\boldsymbol{x}) = \frac{1}{\mathrm{Vol}_{\mathbb{H}^n}(B_{R_0}(\mathbf{p}))} 1_{B_{R_0}(\mathbf{p})}(\boldsymbol{x})$$

It is well known that

$$\mathcal{CL}(\boldsymbol{x}) = \emptyset \quad \text{for any } \boldsymbol{x} \in \mathbb{H}^n.$$

In view of Equality (18), we only need to consider the following integrals.

$$\int_{\mathbb{R}^n} 1_{\exp_{\boldsymbol{x}}^{-1}(B_{R_0}(\mathbf{p}))}(\mathbf{u})\varphi(\|\mathbf{u}\|)\sqrt{G_{\mathbf{x}}(\mathbf{u})}\,d\mathbf{u} \quad \text{for } i=1,\dots n. \tag{19}$$

Since $\varphi \in C_c^1(\mathbb{R}_+)$, for any nonzero φ we get that there is $T, \varepsilon > 0$ such that $\varphi(t) = 0$ for any $t \geq T$ and φ does not change the sign on $(T-\varepsilon, T)$ and is not identically equal to zero. We now have to find $\boldsymbol{x} \in \mathbb{H}^n$ such that (19) is not zero for some $i = 1, \dots n$. To do so, let us choose a point $\boldsymbol{x} \in \mathbb{H}^n$ such that $d(\mathbf{p}, \boldsymbol{x}) = T + R_0 - \varepsilon > 0$. Then we choose a minimal geodesic γ such that $\gamma(0) = \boldsymbol{x}$ and $\gamma(d(\boldsymbol{x}, \mathbf{p})) = \mathbf{p}$. Let $\mathbf{e}_1 = \dot{\gamma}(0) \in T_{\boldsymbol{x}}\mathbb{H}^n$. Then, completing this into an orthonormal basis $(\mathbf{e}_i)$, we now consider the integral (19) with respect to this orthonormal basis. We thus get that $\mathbf{u}_1 > 0$ on $\{|\mathbf{u}| \leq T\} \cap \exp_{\mathbf{x}}^{-1}(B_{R_0}(\mathbf{p}))$ and that $\{|\mathbf{u}| \leq T\} \cap \exp_{\boldsymbol{x}}^{-1}(B_{R_0}(\mathbf{p})) \subset \{T - \varepsilon < |\mathbf{u}| \leq T\}$. This implies

$$\begin{aligned}&\left(\int_{\mathbb{R}^n} 1_{\exp_{\boldsymbol{x}}^{-1}(B_{R_0}(\mathbf{p}))}(\mathbf{u})\varphi(\|\mathbf{u}\|)\mathbf{u}_1\sqrt{G_{\mathbf{x}}(\mathbf{u})}\,d\mathbf{u}\right)^2 \\ &= \left(\int_{\{|\mathbf{u}|\leq T\}\cap\exp_{\boldsymbol{x}}^{-1}(B_{R_0}(\mathbf{p}))} \varphi(\|\mathbf{u}\|)\mathbf{u}_1\sqrt{G_{\mathbf{x}}(\mathbf{u})}\,d\mathbf{u}\right)^2 > 0.\end{aligned}$$

This proves the claim. □

4.2 Generalization

Finally, we remark that the stability results also hold on equivalent families of measure μ, which are not necessarily the product of the conditionally i.i.d distribution on $\mathcal{M}^N$. Below we provide an empirical example of learning interaction kernels in opinion dynamics on sphere. Starting with initial conditions that follows a distribution with i.i.d components, the distribution μ for observation data coming from i.i.d trajectories over the time interval [0,5] does not have i.i.d components any more. But its associated measure ρ is equivalent to the one formed by initial condition. So our stability results can be applied to this case as well.

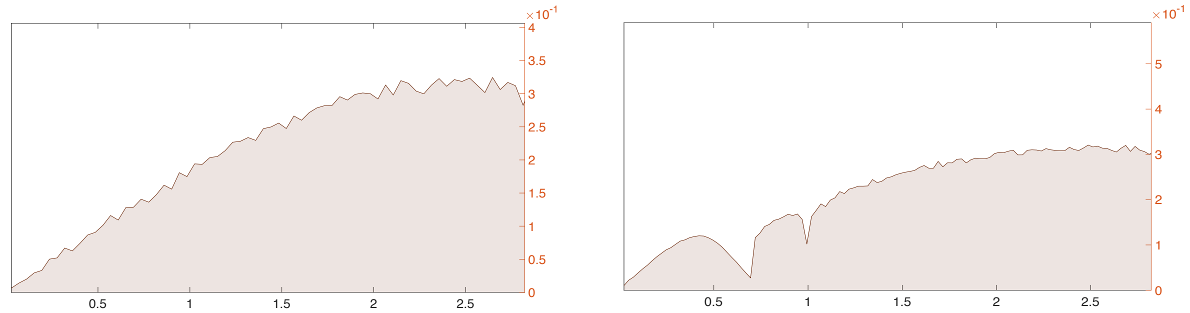

Figure 2: Two examples of ρ from different observation regimes in the same opinion dynamics on $\mathbb{S}^2$ with a piecewise linear kernel ϕ that is compactly supported on $[0, \frac{5}{\sqrt{\pi}}]$, where there are 20 agents/opinions evolving as in [MMQZ21] and the injective radius is $\frac{5}{\sqrt{\pi}}$. To obtain empirical approximations of ρ_1 and ρ_2, we use 3000 trajectories. In the first example, denoted by ρ_1, we choose μ_1 to be the product of uniform distribution on $\mathbb{S}^2$ and observe the position and velocity at $t = 0$ using i.i.d samples from μ_1. In the second example, denoted by ρ_2, we choose μ_2 to be the distribution of positions by observing infinite i.i.d trajectories at time interval [0,5] using μ_1 as the initial distribution. Note that while μ_1 has i.i.d components, μ_2 does not due to time evolution. Despite this difference, we observe numerical evidence that ρ_2 is equivalent to ρ_1 on $[0, \frac{5}{\sqrt{\pi}}]$. Hence, Corollary 4.5 suggests that the stability result also holds for ρ_2.

Corollary 4.5. *Suppose that $\tilde{u}$ and $\tilde{\rho}$ are equivalent measures to μ and $\rho \llcorner [0,R]$ that are defined on $\mathcal{M}^N$ and $[0,R]$ respectively, then $A \in \mathcal{B}(\tilde{H}, \tilde{F})$ where $\tilde{H} = L^2([0,R]; \tilde{\rho} \llcorner [0,R]; \mathbb{R})$ and $\tilde{F} = L^2(\mathcal{M}^N; \tilde{\mu}; T_{\boldsymbol{X}}\mathcal{M}^N)$. Moreover, A admits a bounded inverse on $\tilde{H}$ provided it has a bounded inverse on $\mathcal{H} = L^2(\mathcal{M}^N; \mu; T_{\boldsymbol{X}}\mathcal{M}^N)$*

Proof. The conclusion follow by the fact that Radon-Nikodym derivative of two measures are bounded below and above by two positive constants on $[0,R]$. □

5 Examples where the stability result would fail

In the preceding sections, we demonstrated that if μ is a product of conditionally independent and identically distributed (i.i.d) distributions, then A has a bounded inverse and we derived an estimate on the lower bound of its operator norm. While restricting the space $\mathcal{H}$ to a compact subset of $L^2([0,R]; \rho \llcorner [0,R]; \mathbb{R})$ for some manifold can improve this lower bound, it remains sharp for $\mathbb{S}^n$, irrespective of the choice of $\mathcal{H}$. It is worth recalling that we assumed that μ is absolutely continuous with respect to the volume measure on $\mathcal{M}^N$, and this assumption is fundamental in establishing stability, which was utilized at several key points in the proofs. We shall now provide examples of singular measures where stability may not hold.

Specifically, we start by demonstrating an example of a measure μ such that $\mathbf{f}_\phi(\boldsymbol{X}) = 0$ for $\boldsymbol{X} \sim \mu$. We choose $\mathcal{M} = \mathbb{S}^n$ for $n \geq 1$ and we pick $N = n+2$ interacting particles. We begin by considering the case where $\mu = \delta_{\boldsymbol{X}_0}$ where the vector $\boldsymbol{X}_0 = [\cdots, \mathbf{v_i}, \cdots]^\top \in (\mathbb{S}^n)^N$ stacks the vertices of an $n+1$ dimensional regular unit simplex. In this case, we have

$$\sum_{i=1}^N \mathbf{v}_i = \mathbf{0},$$
$$|\langle \mathbf{v}_i, \mathbf{v}_j \rangle| \equiv c_0, i \neq j$$
$$\omega_{\mathbb{S}^n}(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{v}_j - \langle \mathbf{v}_j, \mathbf{v}_i \rangle \mathbf{v}_i,$$

where c_0 is a nonzero constant that is the cosine of the angle between two edges. For example, $c_0 = \frac{1}{2}$ when $n = 1$. From (5), we have $d_{\mathbb{S}^n}(\mathbf{v}_i, \mathbf{v}_j) \equiv \arccos(c_0)$ for all $i \neq j$. Using these expressions, we have for $i = 1, \cdots, N$

$$\begin{aligned}
[\mathbf{f}_\phi(\boldsymbol{X}_0)]_i &= \frac{1}{N} \sum_{j \neq i} \phi(\arccos(c_0)) \omega(\mathbf{v}_i, \mathbf{v}_j) \\
&= \frac{1}{N} \phi(\arccos(c_0)) (\sum_{j=1}^N \mathbf{v}_j - \langle \sum_{j=1}^N \mathbf{v}_j, \mathbf{v}_i \rangle \mathbf{v}_i) \\
&= 0
\end{aligned}$$

Therefore, $\mathbf{f}_\phi(\boldsymbol{X}) = 0$ whenever $X \sim \mu$. More generally, let U be any orthogonal matrix, and note that $\mathbf{f}_\phi(U\boldsymbol{X}_0) \equiv 0$. If we take μ to be the uniform distribution over the image of $\boldsymbol{X}_0$ under the orthogonal group, then we shall also have $\|A\phi\|_{L^2(\mu)}$=0. However, even in this case, μ is still not absolutely continuous with respect to the volume measure on $(\mathbb{S}^n)^N$.

In conclusion, our findings emphasize the importance of the assumptions made in establishing stability results for $\mathbf{f}_\phi(\boldsymbol{X})$, particularly with regards to the absolute continuity of the measure with respect to the volume measure on the underlying manifold. Understanding the impact of singular measures on stability is crucial for the development of more robust and accurate data-driven discovery methods.

6 Conclusion

We investigate the problem of identifiability of interaction kernels in a first-order system from data. We show that for a large family of distributions μ, the statistical inverse problem is well-posed in specific function spaces. In particular, when μ is also the distribution of independent trajectories with random initializations, our results provide a positive answer to the geometric coercivity conjecture in [MMQZ21], which implies that the least square estimators constructed in [MMQZ21] can achieve mini-max rate of convergence. In addition, our results indicate that in the mean-field case ($N \to \infty$), the inverse problem becomes ill-posed and effective regularization techniques are needed.

There are several avenues for future work, including investigating identifiability under stronger norms such as the RKHS norm, generalizing to second-order systems, possibly with multiple interaction kernels, and studying mean-field systems on Riemannian manifolds. These directions can potentially lead to a deeper understanding of the identifiability problem and can have important applications in various fields, such as physics, biology, and social sciences.

7 Appendix

Definition 2. *Let $\mathbb{X}$ be a nonempty set. A map $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is called a positive definite kernel if k is symmetric and*

$$\sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) \geq 0$$

for all $m \in \mathbb{N}$, $\{x_1, x_2, \cdots, x_m\} \subset \mathbb{X}$ and $\{c_1, c_2, \cdots, c_m\} \subset \mathbb{R}$.

Theorem 7.1 (Properties of positive-definite kernels)**.** *Suppose that $k, k_1, k_2 : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ are positive-definite kernels. Then*

1. *$k_1 k_2$ is positive-definite.*
2. *$f(x)f(y)$ is positive-definite for any function $f : \mathbb{X} \to \mathbb{R}$.*
3. *$k(g(u), g(v))$ is positive-definite for any map $g : \mathbb{Y} \to \mathbb{X}$ with another nonempty set $\mathbb{Y}$.*
4. *Let $(\mathbb{X}, \mathcal{U})$ be a measurable space with a σ-finite measure μ. If $k(x, y)$ is measurable and integrable with respect to $\mu \times \mu$, then $\iint k(x, y) d\mu(x) d\mu(y) \geq 0$.*

Proof. Property 1 and 2 can be found in [VDBCR12, p.69]. Property 3 can be easily checked by applying the definition 2. Property 4 can be found in [RKSF13, p. 524, Property 21.2.12]. $\square$

Acknowledgement

S. Tang was partially supported by Faculty Early Career Development Award sponsored by University of California Santa Barbara, Hellman Family Faculty Fellowship and the NSF DMS-2111303. H. Zhou was partly supported by NSF grant DMS-2109116. M. Tuerkoen wishes to thank Quyuan Lin for helpful comments and discussions.

References

[AHPS21] Hyunjin Ahn, Seung-Yeal Ha, Hansol Park, and Woojoo Shim. Emergent behaviors of cucker–smale flocks on the hyperboloid. *Journal of Mathematical Physics*, 62(8):082702, 2021.

[AHS21] Hyunjin Ahn, Seung-Yeal Ha, and Woojoo Shim. Emergent dynamics of a thermodynamic cucker-smale ensemble on complete riemannian manifolds. *Kinetic & Related Models*, 14(2):323, 2021.

[AMD17] Aylin Aydoğdu, Sean T. McQuade, and Nastassia Pouradier Duteil. Opinion dynamics on a general compact riemannian manifold. *Networks & Heterogeneous Media*, 12(3):489–523, 2017.

[B$^+$11] Jaya Prakash Narayan Bishwal et al. Estimation in interacting diffusions: Continuous and discrete sampling. *Applied Mathematics*, 2(9):1154–1158, 2011.

[BFHM17] Mattia Bongini, Massimo Fornasier, Markus Hansen, and Mauro Maggioni. Inferring interaction rules from observations of evolutive systems i: The variational approach. *Mathematical Models and Methods in Applied Sciences*, 27(05):909–951, 2017.

[BPK16] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[Car92] Manfredo Perdigão do. Carmo. *Riemannian geometry / Manfredo do Carmo ; translated by Francis Flaherty.* Mathematics. Theory and applications. Birkhäuser, Boston, 1992.

[Che21] Xiaohui Chen. Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. *Electronic Communications in Probability*, 26:1–13, 2021.

[DGYZ22] Qiang Du, Yiqi Gu, Haizhao Yang, and Chao Zhou. The discovery of dynamics via linear multistep methods and deep learning: Error estimation. *SIAM Journal on Numerical Analysis*, 60(4):2014–2045, 2022.

[DMH22] Laetitia Della Maestra and Marc Hoffmann. The lan property for mckean-vlasov models in a mean-field regime. *arXiv preprint arXiv:2205.05932*, 2022.

[DRS20] X Duan, JE Rubin, and D Swigon. Identification of affine dynamical systems from a single trajectory. *Inverse Problems*, 36(8):085004, 2020.

[FPP21] Razvan C Fetecau, Hansol Park, and Francesco S Patacchini. Well-posedness and asymptotic behavior of an aggregation model with intrinsic interactions on sphere and other manifolds. *Analysis and Applications*, 19(06):965–1017, 2021.

[FRT21] Jinchao Feng, Yunxiang Ren, and Sui Tang. Data-driven discovery of interacting particle systems using gaussian processes. *arXiv preprint arXiv:2106.02735*, 2021.

[FZ18] Razvan C Fetecau and Beril Zhang. Self-organization on riemannian manifolds. *arXiv preprint arXiv:1802.06089*, 2018.

[GCL22] Valentine Genon-Catalot and Catherine Larédo. Inference for ergodic mckean-vlasov stochastic differential equations with polynomial interactions. 2022.

[GHL04] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian Geometry.* Springer, Berlin, Heidelberg, third edition, 2004.

[GSW19] Susana N Gomes, Andrew M Stuart, and Marie-Therese Wolfram. Parameter estimation for macroscopic pedestrian dynamics models from microscopic data. *SIAM Journal on Applied Mathematics*, 79(4):1475–1500, 2019.

[HKK22] Seung-Yeal Ha, Myeongju Kang, and Dohyun Kim. Emergent behaviors of high-dimensional kuramoto models on stiefel manifolds. *Automatica*, 136:110072, 2022.

[HYM+18] Markus Heinonen, Cagatay Yildiz, Henrik Mannerström, Jukka Intosalmi, and Harri Lähdesmäki. Learning unknown ode models with gaussian processes. In *International Conference on Machine Learning*, pages 1959–1968. PMLR, 2018.

[Kas90] Raphael A Kasonga. Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, 50(3):865–875, 1990.

[LLM+21] Zhongyang Li, Fei Lu, Mauro Maggioni, Sui Tang, and Cheng Zhang. On the identifiability of interaction functions in systems of interacting particles. *Stochastic Processes and their Applications*, 132:135–163, 2021.

[LMT21] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories. *Foundations of Computational Mathematics*, pages 1–55, 2021.

[LZTM19] Fei Lu, Ming Zhong, Sui Tang, and Mauro Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. *Proceedings of the National Academy of Sciences*, 116(29):14424–14433, 2019.

[MB21a] Daniel A Messenger and David M Bortz. Learning mean-field equations from particle data using wsindy. *arXiv preprint arXiv:2110.07756*, 2021.

[MB21b] Daniel A Messenger and David M Bortz. Weak sindy: Galerkin-based data-driven model selection. *Multiscale Modeling & Simulation*, 19(3):1474–1497, 2021.

[MMQZ21] Mauro Maggioni, Jason Miller, H. P. Truong Qui, and Ming Zhong. Learning interaction kernels for agent systems on riemannian manifolds. *International Conference on Machine Learning*, abs/2102.00327, 2021.

[MT14] Sebastien Motsch and Eitan Tadmor. Heterophilious dynamics enhances consensus. *SIAM review*, 56(4):577–621, 2014.

[MTZM20] Jason Miller, Sui Tang, Ming Zhong, and Mauro Maggioni. Learning theory for inferring interaction kernels in second-order interacting agent systems. *arXiv preprint arXiv:2010.03729*, 2020.

[MXPW11] Hongyu Miao, Xiaohua Xia, Alan S Perelson, and Hulin Wu. On identifiability of nonlinear ode models and applications in viral dynamics. *SIAM review*, 53(1):3–39, 2011.

[QWX19] Tong Qin, Kailiang Wu, and Dongbin Xiu. Data driven governing equations approximation using deep neural networks. *Journal of Computational Physics*, 395:620–635, 2019.

[RKSF13] Svetlozar T Rachev, Lev B Klebanov, Stoyan V Stoyanov, and Frank Fabozzi. *The methods of distances in the theory of probability and statistics*, volume 10. Springer, 2013.

[RPK17] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Machine learning of linear differential equations using gaussian processes. *Journal of Computational Physics*, 348:683–693, 2017.

[SKPP21] Louis Sharrock, Nikolas Kantas, Panos Parpas, and Grigorios A Pavliotis. Parameter estimation for the mckean-vlasov stochastic differential equation. *arXiv preprint arXiv:2106.13751*, 2021.

[SL09] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009.

[SRS14] Shelby Stanhope, Jonathan E Rubin, and David Swigon. Identifiability of linear and linear-in-parameters dynamical systems from a single trajectory. *SIAM Journal on Applied Dynamical Systems*, 13(4):1792–1815, 2014.

[Str00] Steven H Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1-4):1–20, 2000.

[STW18] Hayden Schaeffer, Giang Tran, and Rachel Ward. Extracting sparse high-dimensional dynamics from limited data. *SIAM Journal on Applied Mathematics*, 78(6):3279–3295, 2018.

[VDBCR12] C Van Den Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*, volume 100. Springer Science & Business Media, 2012.

[YCY22] Rentian Yao, Xiaohui Chen, and Yun Yang. Mean-field nonparametric estimation of interacting particle systems. *arXiv preprint arXiv:2205.07937*, 2022.

[YWK21] Shihao Yang, Samuel WK Wong, and SC Kou. Inference of dynamic systems from noisy and sparse data via manifold-constrained gaussian processes. *Proceedings of the National Academy of Sciences*, 118(15):e2020397118, 2021.