

1 **IDENTIFIABILITY OF INTERACTION KERNELS IN**  
2 **MEAN-FIELD EQUATIONS OF INTERACTING PARTICLES**

Quanjun Lang and Fei Lu \*

Department of Mathematics, Johns Hopkins University  
3400 N. Charles Street, Baltimore, MD 21218, USA

ABSTRACT. This study examines the identifiability of interaction kernels in mean-field equations of interacting particles or agents, an area of growing interest across various scientific and engineering fields. The main focus is to identify data-dependent function spaces where a quadratic loss functional possesses a unique minimizer. We consider two data-adaptive  $L^2$  spaces: one weighted by a data-adaptive measure and the other using the Lebesgue measure. In each  $L^2$  space, we show that the function space of identifiability is the closure of the RKHS associated with the integral operator of inversion. Alongside prior research, our study completes a full characterization of identifiability in interacting particle systems with either finite or infinite particles, highlighting critical differences between these two settings. Moreover, the identifiability analysis has important implications for computational practice. It shows that the inverse problem is ill-posed, necessitating regularization. Our numerical demonstrations show that the weighted  $L^2$  space is preferable over the unweighted  $L^2$  space, as it yields more accurate regularized estimators.

3 **Keywords:** mean-field equations, identifiability, RKHS, regularization, inverse problem.

4 1. **Introduction.** Systems of interacting particles or agents have become increasingly used in many areas of  
5 science and engineering (see [2, 31, 26, 1] and the references therein). Driven by these applications, there is a  
6 growing interest in inferring the interaction kernel (or the interaction potential) from data, either parametrically  
7 [16, 7, 28, 9] or in a nonparametrically for broader applicability [4, 23, 21, 22, 17, 10, 32].

8 The inference problem can be classified into two categories: (i) a statistical learning problem when the  
9 system consists of finitely many particles and the data includes multiple trajectories of all particles, and (ii)  
10 a deterministic inverse problem for the mean-field equation (MFE) from data consisting of a solution to the  
11 MFE, which arises when the number of particles is so large that only the macroscopic density of particles can  
12 be observed. For the statistical learning of kernels in systems with finitely many particles, previous studies  
13 [23, 21, 22] minimize loss functionals based on the mean-square error or the likelihood of the data, estab-  
14 lishing computationally efficient algorithms that yield nonparametric estimators achieving the minimax rate  
15 of convergence. In particular, the studies [19, 18] show that any square-integrable kernel is identifiable under  
16 a coercivity condition, which imposes constraints on the distribution of the data trajectories. For the inverse  
17 problem of the MFE, the study [17] has introduced a *derivative-free probabilistic loss functional* and based on  
18 it, a scalable nonparametric regression algorithm that produces a convergent estimator robust to discrete noisy  
19 data. However, it remains open to study the identifiability of the kernel in the MFE from data.

20 This study provides a complete characterization of the identifiability of kernels in MFE by the probabilistic  
21 loss functional. The key is to determine the *data-dependent* function space of identifiability (FSOI), in which the  
22 quadratic loss functional has a unique minimizer. We consider two *data-adaptive*  $L^2$  spaces: one is unweighted  
23 with the Lebesgue measure, and the other is weighted with a data-dependent exploration measure. In each  $L^2$   
24 space, the second-order derivative of the loss functional defines a semi-positive integral operator, which acts  
25 as the operator of inversion. The FSOI is then the closure of this integral operator's eigenspace of nonzero  
26 eigenvalues. Furthermore, identifiability holds in the  $L^2$  space if and only if the integral operator is strictly  
27 positive. However, the inverse problem is ill-posed due to the inversion of a compact operator. Our results  
28 apply to both radial and non-radial interaction kernels.

2020 *Mathematics Subject Classification.* Primary: 35R30; Secondary: 68Q32.

*Key words and phrases.* mean-field equations, interacting particles, RKHS, identifiability, inverse problem.

FL is supported by NSF DMS-1913243, FA9550-20-1-0288 and DE-SC0021361.

\*Corresponding author: Fei Lu.

1 Together with [19, 18], this study completes a full characterization of the identifiability of kernels in inter-  
 2 acting particle systems with either finitely or infinitely many particles. Notably, there are significant differences  
 3 between these two settings. For systems with  $N$  particles, the identifiability holds in the weighted  $L^2$  space  
 4 because of a coercivity condition with a constant  $\frac{1}{N-2}$  (see [19, Proposition 2.1]), and the inverse problem is  
 5 well-posed. In contrast, for the inverse problem of the MFE, no coercivity holds in  $L^2$  (in agreement with the  
 6 above coercivity constant vanishes as  $N \rightarrow \infty$ ), the identifiability barely holds in  $L^2$ , and the inverse problem  
 7 is ill-posed.

8 The identifiability has important implications for computational practice. The ill-posedness implies that the  
 9 normal matrix in regression becomes ill-conditioned as the dimension of the hypothesis space increases. Thus,  
 10 regularization is necessary. The two ambient  $L^2$  spaces provide natural norms for the Tikhonov regularization.  
 11 We demonstrate numerically that the weighted  $L^2$  norm is preferable over the unweighted  $L^2$  space because it  
 12 leads to more accurate regularized estimators in the context of singular value decomposition (SVD) analysis.

13 Furthermore, the identifiability theory introduces adaptive RKHSs for regularization. They are different from  
 14 the widely-used kernel regression [11, 27, 9] or RKHS regularization [8], where the reproducing kernels are  
 15 pre-selected. They invite further study on data-adaptive regularization strategies for ill-posed statistical learning  
 16 and inverse problems [20].

17 The exposition in our manuscript proceeds as follows. In Section 2, we define identifiability and introduce the  
 18 main results. Section 3 studies identifiability for radial kernels and Section 4 extends the results to general non-  
 19 radial kernels. We discuss in Section 5 the implications of identifiability to computational practice. Appendix  
 20 A.1 provides a brief review of positive definite functions and reproducing kernel Hilbert spaces.

We shall use the notations in Table 1.

TABLE 1. Notations.

	Radial kernel	Non-radial kernel
Interaction potential	$\Psi( x )$ and $\psi = \Psi'$ ;	$\Psi(x)$
Interaction kernel	$K_\psi(x) = \psi( x )\frac{x}{ x }$	$K_\psi(x) = \nabla\Psi(x)$
Loss functional	$\mathcal{E}(\psi)$	$\mathcal{E}(K_\psi)$
Density of exploration measure	$\bar{\rho}_T, \mathcal{X} = \text{support}(\bar{\rho}_T)$ , in (2.9)	
Function space of learning	$L^2(\mathcal{X})$ and $L^2_{\bar{\rho}_T}$	
Mercer kernel & RKHS in $L^2(\mathcal{X})$	$\bar{G}_T$ in (2.11) and $\mathcal{H}_G$	$\bar{F}_T$ in (4.3) and $\mathcal{H}_F$
Mercer kernel & RKHS in $L^2_{\bar{\rho}_T}$	$\bar{R}_T$ in (3.8) and $\mathcal{H}_R$	$\bar{Q}_T$ in (4.3) and $\mathcal{H}_Q$

21

22 **2. Main results.** Consider the McKean-Vlasov mean-field equation (MFE) of interacting particles:

$$\begin{aligned} \partial_t u &= \nu \Delta u + \text{div}[u(K * u)], \quad x \in \mathbb{R}^d, t \in [0, T] \\ u(x, t) &\geq 0, \quad \int_{\mathbb{R}^d} u(x, t) dx = 1, \end{aligned} \tag{2.1}$$

23 where  $K * u$  denotes the convolution

$$(K * u)(x, t) = \int_{\mathbb{R}^d} K(y)u(x - y, t)dy.$$

24 Here  $K = \nabla\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is called an *interaction kernel* and  $\Phi$  is called the interaction potential. In particular,  
 25 if  $\Phi$  is radial, denoting  $\Phi(x) = \Phi(|x|)$  with an abuse of notation, we have

$$K(x) = \nabla(\Phi(|x|)) = \phi(|x|)\frac{x}{|x|}, \quad \text{with } \phi(r) = \Phi'(r). \tag{2.2}$$

26 where  $\phi$  is also called *interaction kernel* for simplicity.

1 The mean-field equation (also called aggregation-diffusion equation [5]) describes the macroscopic density  
 2 for the systems of interacting particles when  $N \rightarrow \infty$ :

$$\frac{d}{dt} \mathbf{X}_t^i = -\frac{1}{N} \sum_{j=1}^N K(\mathbf{X}_t^j - \mathbf{X}_t^i) + \sqrt{2\nu} d\mathbf{B}_t^i, \quad \text{for } i = 1, \dots, N, \quad (2.3)$$

3 where  $\mathbf{X}_t^i \in \mathbb{R}^d$  represents the position of agent  $i$  at time  $t$ . Denote by  $\mu^{(N)}(dx, t) = \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{X}_t^i - x)$   
 4 the empirical measure of the particles. Under suitable conditions on  $\Phi$ , it is well-known that  $\mu^{(N)}(dx, t) \rightarrow$   
 5  $u(x, t)dx$  in relative entropy of the invariant measure as the number of particles  $N \rightarrow \infty$  (see e.g., [25, 24, 6,  
 6 14]).

7 Our goal is to study the identifiability of the *interaction kernel*  $K$  or  $\phi$  from data consisting of a solution of  
 8 the MFE. Throughout the paper, we assume that the data  $u$  is a bounded weak solution to the MFE:

9 **Assumption 2.1** (Smoothness of data). *The data  $u$  is a bounded continuous weak solution to the MFE with*  
 10 *bounded support, that is,  $\sup_{(x,t) \in \mathbb{R}^d \times [0,T]} |u(x, t)| \leq C_u$  for a positive constant  $C_u$ , and  $\bigcup_{t \in [0,T]} \text{supp}(u(\cdot, t))$*   
 11 *is bounded, where  $\text{supp}(u(\cdot, t)) = \{x \in \mathbb{R}^d : u(x, t) > 0\}$ .*

12 Such a solution exists when the interaction kernel is local Lipschitz with polynomial growth:

$$|K(x) - K(y)| \leq C(|x - y| \wedge 1)(1 + |x|^m + |y|^m), \quad \forall x, y \in \mathbb{R}^d$$

13 for a constant  $C > 0$  and an integer  $m \geq 1$ . For further study on the forward problem of the MFE, we refer  
 14 to [30] for Lipschitz kernels, [25, 24] for uniform convex kernels and the existence of an equilibrium, and  
 15 the references in [6, 14, 15] for general (including singular) kernels. The assumptions on the solution being  
 16 continuous with bounded support are technical, and we discuss possible extensions to measure-valued solutions  
 17 with unbounded support in Remark 3.11.

18 In the rest of this section, we present the main results only for radial kernels, and similar results hold for  
 19 non-radial kernels (see Section 4).

20 **2.1. A loss functional in nonparametric regression.** We consider nonparametric approaches in which one  
 21 finds an estimator by minimizing a loss functional in a hypothesis space [4, 17, 10, 32]. Importantly, noticing  
 22 that the MFE in (2.1) depends linearly on the kernel, we can estimate the kernel by nonparametric regression,  
 23 in which we minimize a *quadratic loss functionals* efficiently by solving a least squares problem.

24 We consider the probabilistic loss functional introduced in [17],

$$\mathcal{E}(\psi) = \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} [ |K_\psi * u|^2 u + 2\partial_t u (\Psi(|\cdot|) * u) - 2\nu u (\Delta \Psi * u) ] dx dt. \quad (2.4)$$

25 where  $K_\psi(x) = \nabla \Psi(|x|) = \psi(|x|) \frac{x}{|x|}$ . It is the expectation of the log-likelihood of the McKean-Vlasov  
 26 stochastic differential equation (to be introduced in (2.7)). It has two appealing features: (i) it is derivative-free  
 27 (i.e., not using derivatives of the data  $u$ ), which plays a key role in obtaining a robust estimator in [17]; and  
 28 (ii) it applies to high-dimensional systems because the integrals in  $u$  can be written as expectations, which is  
 29 important when  $u$  is approximated by the empirical measure of the particles.

30 To focus on the identification of the kernel, we present an oracle version of the loss functional (see Lemma  
 31 3.2 for its derivation).

32 **Definition 2.2** (Oracle loss functional). *Let  $u$  be a solution of the mean-field equation (2.1) on  $[0, T]$ , and*  
 33 *denote  $K_\psi(x) = \nabla \Psi(|x|) = \psi(|x|) \frac{x}{|x|}$ , where  $\Psi : \mathbb{R}^+ \rightarrow \mathbb{R}$  is a radial interaction potential with derivative*  
 34  *$\psi(r) = \Psi'(r)$ . We consider the loss functional*

$$\mathcal{E}(\psi) = \langle\langle \psi, \psi \rangle\rangle - 2\langle\langle \psi, \phi_{true} \rangle\rangle, \quad (2.5)$$

35 where the bilinear form  $\langle\langle \cdot, \cdot \rangle\rangle$  is defined by

$$\langle\langle \varphi, \psi \rangle\rangle := \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} u(K_\varphi * u) \cdot (K_\psi * u) dx dt, \quad (2.6)$$

36 assuming that the integrals are well-defined.

1 Since the loss functional is quadratic, its minimizer in a finite-dimensional hypothesis space  $\mathcal{H}$  is computed  
2 by least squares in practice (with the term involving the true kernel approximated using data).

3 **Remark 2.3** (Least squares estimator). *For any hypothesis space  $\mathcal{H} = \text{span}\{\phi_1, \dots, \phi_n\}$  such that*

$$A_n(i, j) = \langle\langle \phi_i, \phi_j \rangle\rangle, \quad b_n(i) = \langle\langle \phi_i, \phi_{\text{true}} \rangle\rangle$$

4 *are well-defined, a minimizer of the loss functional  $\mathcal{E}$  in  $\mathcal{H}$  is given by least squares*

$$\hat{\phi}_{\mathcal{H}} = \sum_{i=1}^n \hat{c}_i \phi_i, \quad \text{where } \hat{c} = \arg \min_{c \in \mathbb{R}^n} \mathcal{E}(c), \quad \mathcal{E}(c) = c^\top A_n c - 2c^\top b_n.$$

5 *In particular, when  $A_n$  is invertible, we have  $\hat{c} = A_n^{-1} b_n$ , and hence  $\hat{\phi}_{\mathcal{H}}$  is the unique minimizer of  $\mathcal{E}$  in  $\mathcal{H}$ .*  
6 *When  $A_n$  is singular,  $A_n^{-1}$  denotes the Moore–Penrose pseudo inverse. Furthermore, when  $A_n$  is ill-conditioned*  
7 *and there are errors in the approximation of  $b_n$  due to measurement noise or numerical error, regularization*  
8 *helps to avoid amplifying the errors in  $b_n$  (see Section 5 for more discussions).*

9 In a parametric inference approach, the hypothesis space is determined by the parametric form of the kernel.  
10 In a nonparametric regression approach, one selects the optimal hypothesis space with proper smoothness and  
11 dimension.

12 Two fundamental elements are crucial for both approaches: the function space of learning and identifiability.  
13 The function space of learning provides a proper metric on the accuracy of the estimator, and the identifiability  
14 reveals if the inverse problem is ill-posed and sheds insights on regularization. In the next two subsections, we  
15 first introduce data-adaptive function spaces of learning, and then define the identifiability.

16 **2.2. The data-adaptive function spaces of learning.** We consider two data-adaptive function spaces of learn-  
17 ing: a weighted  $L^2$  space with a data-based measure and an unweighted  $L^2$  space on the support of the measure.

18 We introduce first a *exploration measure* to quantify the exploration of the kernel by data, because we can  
19 only learn the kernel in the region where the data explores. This measure originates from a probabilistic repre-  
20 sentation of the mean-field equation. Recall that Equation (2.1) is the Fokker-Planck equation (also called the  
21 Kolmogorov forward equation) of the following McKean-Vlasov stochastic differential equation

$$\begin{cases} d\bar{X}_t = -[K_{\text{true}} * u](\bar{X}_t, t)dt + \sqrt{2\nu}dB_t, \\ \mathcal{L}(\bar{X}_t) = u(x, t)dx, \end{cases} \quad (2.7)$$

22 for all  $t \geq 0$ . Here  $\mathcal{L}(\bar{X}_t)$  denotes the law of  $\bar{X}_t$ , whose probability density is  $u(\cdot, t)$ . Let  $(\bar{X}'_t, t \geq 0)$  be an  
23 independent copy of  $(\bar{X}_t, t \geq 0)$  and denote  $r_t = |\bar{X}_t - \bar{X}'_t|$ . We write the convolution  $K_\psi * u$  as

$$[K_\psi * u](\bar{X}_t, t) = \mathbb{E}[K_\psi(\bar{X}_t - \bar{X}'_t) \mid \bar{X}_t] = \mathbb{E}[\psi(r_t) \frac{\bar{X}_t - \bar{X}'_t}{r_t} \mid \bar{X}_t]. \quad (2.8)$$

24 This probabilistic representation of  $K_\psi * u$  indicates that the independent variable of  $\psi$  is explored by the  
25 process  $\{|\bar{X}_t - \bar{X}'_t|, t \in [0, T]\}$  (or  $\{\bar{X}_t - \bar{X}'_t, t \in [0, T]\}$  for non-radial kernels). Let  $\bar{\rho}_T$  denote the average of  
26 probability densities of the processes:

$$\bar{\rho}_T(r) = \frac{1}{T} \int_0^T \rho_t(r) dt = \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} \int_{\mathbb{S}^d} r^{d-1} u(y - r\xi, t) u(y, t) d\xi dy dt, \quad (2.9)$$

27 where  $\rho_t$  denotes the density of  $|\bar{X}'_t - \bar{X}_t|$  (or  $\bar{X}'_t - \bar{X}_t$  for the general case) for each  $t$ . Under Assumption 2.1,  
28 the probability density function  $\bar{\rho}_T$  is  $C^2$ . We denote the support of  $\bar{\rho}_T$  by  $\mathcal{X}$ :

$$\mathcal{X} := \text{supp}(\bar{\rho}_T).$$

29 Note that  $\mathcal{X}$  is bounded since the set  $\bigcup_{t \in [0, T]} \text{supp}(u(\cdot, t))$  is bounded.

30 Two data-adaptive function spaces emerge:  $L^2(\mathcal{X}, \bar{\rho}_T(r) dr)$  (denoted by  $L^2_{\bar{\rho}_T}$  hereafter) and the unweighted  
31 space  $L^2(\mathcal{X})$  with the Lebesgue measure on  $\mathcal{X}$ . Both spaces are viable choices because the loss functional in  
32 (2.5) is well-defined in either of them (see Lemma 3.1). However, we will show by numerical examples that the  
33 latter extracts more information from data and leads to more accurate regularized estimators (see Section 5).

1 **2.3. Definition of identifiability.** We define identifiability as the uniqueness of the minimizer of a quadratic  
 2 loss functional in a linear hypothesis space. This definition applies to general quadratic loss functionals. This  
 3 study focuses on the loss functional in (2.5).

4 **Definition 2.4** (Identifiability). *Given data consisting of a solution  $(u(x, t), x \in \mathbb{R}^d, t \in [0, T])$  to the mean-*  
 5 *field equation (2.1) and a quadratic loss functional  $\mathcal{E}$ , we say that the interaction kernel is identifiable by  $\mathcal{E}$  in*  
 6 *a linear subspace  $\mathcal{H}$  of  $L^2(\mathcal{X})$  or  $L^2_{\bar{\rho}_T}$  if the true kernel is the unique minimizer of the loss functional in  $\mathcal{H}$ . We*  
 7 *call the largest such linear subspace the function space of identifiability (FSOI).*

8 When  $\mathcal{H}$  is a finite-dimensional (e.g., in parametric inference), Remark 2.3 suggests that identifiability holds  
 9 in  $\mathcal{H}$  if the normal matrix  $A_n$  is invertible, in other words,  $c^\top A_n c > 0$  for all nonzero  $c \in \mathbb{R}^n$ . Similarly, when  
 10  $\mathcal{H}$  is infinite-dimensional, the identifiability is equivalent to the non-degeneracy of the bilinear form  $\langle\langle \cdot, \cdot \rangle\rangle$ , as  
 11 the following lemma shows.

12 **Lemma 2.5.** *Identifiability holds in a linear space  $\mathcal{H}$  for the loss functional  $\mathcal{E}$  in (2.4) if the bilinear form  $\langle\langle \cdot, \cdot \rangle\rangle$*   
 13 *in (2.6) is non-degenerate in  $\mathcal{H}$ . That is, the true kernel is the unique minimizer of the loss functional  $\mathcal{E}$  in (2.5)*  
 14 *in  $\mathcal{H}$  iff  $\langle\langle \phi, \phi \rangle\rangle > 0$  for all  $\phi \in \mathcal{H}$  except  $\phi = 0$ .*

15 *Proof.* Denote the true kernel by  $\phi_{true}$ . Note that

$$\mathcal{E}(\psi) = \langle\langle \psi - \phi_{true}, \psi - \phi_{true} \rangle\rangle - \langle\langle \phi_{true}, \phi_{true} \rangle\rangle. \quad (2.10)$$

16 Thus,  $\phi_{true} \in \mathcal{H}$  is the unique minimizer of  $\mathcal{E}$  iff  $\langle\langle \phi, \phi \rangle\rangle > 0$  for all nonzero  $\phi \in \mathcal{H}$ .  $\square$

17 The bilinear form plays a key role in our study of identifiability. We can write it as (see (3.4) for its derivation)

$$\langle\langle \varphi, \psi \rangle\rangle = \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \varphi(r)\psi(s) \bar{G}_T(r, s) dr ds =: \langle \mathcal{L}_{\bar{G}_T} \varphi, \psi \rangle_{L^2(\mathcal{X})},$$

18 where the integral kernel  $\bar{G}_T$  is a Mercer kernel (see Lemma 3.3) given by

$$\bar{G}_T(r, s) = \frac{1}{T} \int_0^T \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \int_{\mathbb{R}^d} \xi \cdot \eta (rs)^{d-1} u(z - r\xi, t) u(z - s\eta, t) u(z, t) dz d\xi d\eta dt, \quad (2.11)$$

19 where  $\mathbb{S}^d$  denotes the unit sphere in  $\mathbb{R}^d$ . Here  $\mathcal{L}_{\bar{G}_T}$  denotes the integral operator with kernel  $\bar{G}_T$  (see its  
 20 definition in (3.5)). It acts as the operator of inversion, and plays a key role in the connection between the  
 21 function space of identifiability and the RKHS of  $\bar{G}_T$ .

22 **2.4. Main results.** We characterize the data-dependent function spaces of identifiability (FSOI) of the loss  
 23 functional (2.4) in both  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$ , and compare them in computational practice.

- 24 • In  $L^2(\mathcal{X})$ , the FSOI is the closure of the RKHS  $\mathcal{H}_G$  with reproducing kernel  $\bar{G}_T$  in (2.11). The identifi-  
 25 ability holds in any linear subspace of the FSOI. Importantly, the identifiability holds in  $L^2(\mathcal{X})$  iff  $\mathcal{H}_G$  is  
 26 dense in it, or equivalently, the integral operator  $\mathcal{L}_{\bar{G}_T}$  in  $L^2(\mathcal{X})$  with integral kernel  $\bar{G}_T$  is strictly positive  
 27 (see Theorems 3.5–3.6).
- 28 • In  $L^2_{\bar{\rho}_T}$ , the same results hold with  $\bar{G}_T$  replaced by  $\bar{R}_T(r, s) = \frac{\bar{G}_T(r, s)}{\bar{\rho}_T(r)\bar{\rho}_T(s)}$  (see Theorem 3.7).
- 29 • Similar identifiability results hold for non-radial kernels (see Section 4).

30 We point out that identifiability is weaker than well-posedness. Identifiability in a linear space  $\mathcal{H}$  only  
 31 ensures that the loss functional has a unique minimizer in  $\mathcal{H}$ . It does not ensure the well-posedness of the  
 32 inverse problem unless the bilinear form satisfies a coercivity condition in  $\mathcal{H}$  (see Remark 3.8). When  $\mathcal{H}$   
 33 is finite-dimensional, identifiability is equivalent to the invertibility of the normal matrix in regression (see  
 34 Remark 2.3), and identifiability implies well-posedness. However, when  $\mathcal{H}$  is infinite-dimensional, the inverse  
 35 problem is ill-posed because the inverse of the integral operator  $\mathcal{L}_{\bar{G}_T}$  is unbounded.

36 The identifiability study has important implications for computational practice. The identifiability theory  
 37 implies that the regression matrix will become ill-conditioned as the dimension of the hypothesis space increases  
 38 (see Theorem 5.1). Thus, regularization becomes necessary. We compare two regularization norms, the norms  
 39 of  $L^2_{\bar{\rho}_T}$  and  $L^2(\mathcal{X})$ , in the context of singular value decomposition (SVD) analysis and the truncated SVD

1 regularization. Numerical tests suggest that the inversion in  $L^2_{\bar{\rho}_T}$  is less ill-conditioned and its regularization  
2 leads to more accurate estimators.

3 **3. Radial interaction kernels.** Radial interaction kernels are of particular interest because of their simplicity  
4 and efficiency in representing symmetric interactions. We will first show that the loss functional is well-defined  
5 and prove its oracle version. Then, we discuss the identifiability in the ambient function spaces  $L^2(\mathcal{X})$  and in  
6  $L^2_{\bar{\rho}_T}$  in Section 3.2-3.3, respectively.

7 Throughout this section, we let  $\psi \in L^2_{\bar{\rho}_T}$  with  $\bar{\rho}_T$  in (2.9). We let  $\Psi : \mathbb{R}^+ \rightarrow \mathbb{R}$  be  $\Psi(r) = \int_0^r \psi(s) ds$  and  
8 denote  $K_\psi(x) = \psi(|x|) \frac{x}{|x|}$ .

9 **3.1. The loss functional.** We show first that the oracle loss functional in (2.5) is well-defined and it is equiva-  
10 lent to the loss functional in practice.

11 **Lemma 3.1.** *For any  $\varphi, \psi \in L^2_{\bar{\rho}_T}$ , the bilinear form  $\langle\langle \varphi, \psi \rangle\rangle$  in (2.6) satisfies*

$$\langle\langle \varphi, \psi \rangle\rangle \leq \|\psi\|_{L^2_{\bar{\rho}_T}} \|\varphi\|_{L^2_{\bar{\rho}_T}}. \quad (3.1)$$

12 *Also, for any  $\psi \in L^2_{\bar{\rho}_T}$ , the radial loss functional in (2.5) is bounded above by*

$$\mathcal{E}(\psi) \leq \|\psi - \phi_{true}\|_{L^2_{\bar{\rho}_T}}^2 - \langle\langle \phi_{true}, \phi_{true} \rangle\rangle. \quad (3.2)$$

13 *Proof.* Recall that  $u(\cdot, t)$  is the law of  $\bar{X}_t$  defined in (2.7). By definition in (2.6), we have

$$\begin{aligned} \langle\langle \varphi, \psi \rangle\rangle &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} (K_\varphi * u) \cdot (K_\psi * u) u(x, t) dx dt \\ &= \frac{1}{T} \int_0^T \mathbb{E}[K_\varphi * u(\bar{X}_t, t) \cdot K_\psi * u(\bar{X}_t, t)] dt. \end{aligned} \quad (3.3)$$

14 Let  $\bar{X}'_t$  be an independent copy of  $\bar{X}_t$ . The above integrand in time is controlled by

$$\begin{aligned} &\mathbb{E}[K_\varphi * u(\bar{X}_t, t) \cdot K_\psi * u(\bar{X}_t, t)] \\ &\leq \mathbb{E}[|K_\varphi * u(\bar{X}_t, t)|^2]^{1/2} \mathbb{E}[|K_\psi * u(\bar{X}_t, t)|^2]^{1/2} \quad (\text{Cauchy-Schwartz}) \\ &\leq \mathbb{E}[\mathbb{E}[|K_\varphi(\bar{X}_t - \bar{X}'_t)|^2 | \bar{X}_t]]^{1/2} \mathbb{E}[\mathbb{E}[|K_\psi(\bar{X}_t - \bar{X}'_t)|^2 | \bar{X}_t]]^{1/2} \quad (\text{by (2.8)}) \\ &\leq \mathbb{E}[\mathbb{E}[|K_\varphi(\bar{X}_t - \bar{X}'_t)|^2 | \bar{X}_t]]^{1/2} \mathbb{E}[\mathbb{E}[|K_\psi(\bar{X}_t - \bar{X}'_t)|^2 | \bar{X}_t]]^{1/2} \quad (\text{by Jensen's inequality}) \\ &= \mathbb{E}[|K_\varphi(\bar{X}_t - \bar{X}'_t)|^2]^{1/2} \mathbb{E}[|K_\psi(\bar{X}_t - \bar{X}'_t)|^2]^{1/2} = \|\varphi\|_{L^2(\rho_t)} \|\psi\|_{L^2(\rho_t)}. \end{aligned}$$

15 Then, we obtain (3.1).

16 The upper bound in (3.2) follows from (3.1) and (2.10).  $\square$

17 **Lemma 3.2.** *The loss functional in (2.4) can be written in the oracle version in (2.5) if  $u$  is a solution of the*  
18 *MFE (2.1) with kernel  $\phi_{true}$ .*

19 *Proof.* The proof follows from the MFE and integration by parts. More specifically, note that  $u$  vanishes at the  
20 boundary of its support, we have  $\int_{\mathbb{R}^d} u(\Delta \Psi * u) dx = \int_{\mathbb{R}^d} \Delta u(\Psi * u)$  by integration by parts. Then, the MFE  
21  $\partial_t u - \nu \Delta u = \text{div}[u(K_{\phi_{true}} * u)]$  implies that

$$\begin{aligned} &\frac{1}{T} \int_0^T \int_{\mathbb{R}^d} [2\partial_t u(\Psi(|\cdot|) * u) - 2\nu u(\Delta \Psi * u)] dx dt \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} 2(\partial_t u - \Delta u)(\Psi(|\cdot|) * u) dx dt \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} 2\text{div}[u(K_{\phi_{true}} * u)](\Psi(|\cdot|) * u) dx dt \\ &= -\frac{1}{T} \int_0^T \int_{\mathbb{R}^d} 2u(K_{\phi_{true}} * u) \cdot (K_\psi * u) dx dt = -2\langle\langle \psi, \phi_{true} \rangle\rangle, \end{aligned}$$

1 where in the third equality, we used integration by parts along with the fact that  $K_\psi(x) = \nabla\Psi(|\cdot|)$ .  $\square$

2 **3.2. Identifiability in the unweighted L2 space.** Now we show that the interaction kernel is identifiable by  
 3 the loss functional (2.4) in the  $L^2(\mathcal{X})$ -closure of the RKHS  $\mathcal{H}_G$  with reproducing kernel  $\overline{G}_T$  defined in (2.11).  
 4 The key element is the integral operator of this reproducing kernel: it connects the RKHS with the space  $L^2(\mathcal{X})$   
 5 and allows for a spectral characterization of identifiability.

6 The reproducing kernel  $\overline{G}_T$  emerges from the bilinear form in the loss functional. Specifically, by a change  
 7 of variable to polar coordinates with  $K_\varphi(y) = \varphi(|y|)\frac{y}{|y|} = \varphi(r)\xi$  by setting  $r = |y|$  and  $\xi = \frac{y}{|y|} \in \mathbb{S}^d$ , we can  
 8 write the bilinear form as

$$\begin{aligned} \langle\langle \varphi, \psi \rangle\rangle &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} (K_\varphi * u) \cdot (K_\psi * u) u(x, t) dx dt \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_\varphi(y) \cdot K_\psi(z) \int_{\mathbb{R}^d} u(x-y, t) u(x-z, t) u(x, t) dx dy dz dt \\ &= \int_0^\infty \int_0^\infty \varphi(r) \psi(s) \overline{G}_T(r, s) dr ds, \end{aligned} \quad (3.4)$$

9 where the last equality follows from the definition of  $\overline{G}_T$  in (2.11),

$$\overline{G}_T(r, s) = \frac{1}{T} \int_0^T \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \int_{\mathbb{R}^d} \xi \cdot \eta (rs)^{d-1} u(x-r\xi, t) u(x-s\eta, t) u(x, t) dx d\xi d\eta dt.$$

10 **Lemma 3.3.** Under Assumption 2.1, the kernel  $\overline{G}_T$  is a Mercer kernel, i.e., it is symmetric, continuous and  
 11 positive definite.

12 *Proof.* The symmetry is clear from its definition and the continuity follows from the continuity of  $u$ . To show  
 13 that it is positive definite (see Definition A.1), for any  $\{r_1, \dots, r_k\} \subset \mathbb{R}^+$  and  $(c_1, \dots, c_k) \in \mathbb{R}^k$ , we have

$$\begin{aligned} &\sum_{i,j} c_i c_j \overline{G}_T(r_i, r_j) \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} \sum_{ij} c_i c_j (r_i r_j)^{d-1} \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \xi \cdot \eta u(x-r_i \xi, t) u(x-r_j \eta, t) d\xi d\eta u(x, t) dx dt \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} \left| \sum_i c_i r_i^{d-1} \int_{\mathbb{S}^d} \xi u(x-r_i \xi, t) d\xi \right|^2 u(x, t) dx dt \geq 0. \end{aligned}$$

14 Thus, it is positive definite.  $\square$

15 Since  $\overline{G}_T$  is a Mercer kernel, it determines an RKHS  $\mathcal{H}_G$  with  $\overline{G}_T$  as reproducing kernel (see Appendix  
 16 A.1). We show next that the identifiability holds on the closure of  $\mathcal{H}_G$ , by studying the integral operator with  
 17 kernel  $\overline{G}_T$ :

$$\mathcal{L}_{\overline{G}_T} f(r) = \int_{\mathbb{R}^+} \overline{G}_T(r, s) f(s) ds. \quad (3.5)$$

18 Note that by definition,

$$\langle\langle \varphi, \psi \rangle\rangle = \langle \varphi, \mathcal{L}_{\overline{G}_T} \psi \rangle_{L^2(\mathcal{X})}. \quad (3.6)$$

19 We start with a lemma on the boundedness and integrability of  $\overline{G}_T$ .

20 **Lemma 3.4.** Under Assumption 2.1, the kernel function  $\overline{G}_T$  in (2.11) satisfies:

- 21 (a) For all  $r, s \in \mathbb{R}^+$ ,  $\overline{G}_T(r, s) \leq C_u C_d |\mathcal{X}|^{d-1} \min\{\overline{\rho}_T(r), \overline{\rho}_T(s)\}$ , where  $|\mathcal{X}|$  denotes the Lebesgue measure  
 22 of  $\mathcal{X}$ , and  $C_d = |\mathbb{S}^d| = 2\pi^{d/2} \Gamma(\frac{d}{2})$ .  
 23 (b)  $\overline{G}_T$  is in  $L^2(\mathcal{X} \times \mathcal{X})$  and  $\frac{\overline{G}_T}{\overline{\rho}_T \otimes \overline{\rho}_T} \in L^2(\overline{\rho}_T \otimes \overline{\rho}_T)$ .

- 1 *Proof.* Recall that  $\mathcal{X}$  is the support of  $\bar{\rho}_T(r) = \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} \int_{\mathbb{S}^d} r^{d-1} u(y - r\xi, t) u(y, t) d\xi dy dt$  defined in (2.9).  
 2 Since  $u$  has bounded support, so the set  $\mathcal{X}$  is bounded and  $s \leq |\mathcal{X}|$  for each  $s \in \mathcal{X}$ . Then, by the uniform  
 3 boundedness of  $u$ , we have  $s^{d-1} \int_{\mathbb{S}^d} u(x - s\eta, t) d\eta \leq |\mathcal{X}|^{d-1} C_u$  for any  $s \in \mathbb{R}^+$  and  $x \in \mathbb{R}^d$ . Hence, we have

$$\begin{aligned} \bar{G}_T(r, s) &= \frac{1}{T} \int_0^T \int_{\mathbb{S}^d} \int_{\mathbb{S}^d} \int_{\mathbb{R}^d} \xi \cdot \eta (rs)^{d-1} u(x - r\xi, t) u(x - s\eta, t) u(x, t) dx d\xi d\eta dt \\ &\leq |\mathcal{X}|^{d-1} C_u C_d \frac{1}{T} \int_0^T \int_{\mathbb{S}^d} \int_{\mathbb{R}^d} r^{d-1} u(x - r\xi, t) u(x, t) dx d\xi dt \\ &= |\mathcal{X}|^{d-1} C_u C_d \bar{\rho}_T(r), \end{aligned}$$

- 4 for any  $s \in \mathcal{X}$ . Similarly,  $\bar{G}_T(r, s) \leq |\mathcal{X}|^{d-1} C_u C_d \bar{\rho}_T(s)$  for any  $r$ . Then, (a) follows.  
 5 For (b), we obtain  $\bar{G}_T$  is in  $L^2(\mathcal{X} \times \mathcal{X})$  by applying (a):

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^+} |\bar{G}_T(r, s)|^2 dr ds \leq |\mathcal{X}|^{2d-2} C_u^2 C_d^2 \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \bar{\rho}_T(r) \bar{\rho}_T(s) dr ds = |\mathcal{X}|^{2d-2} C_u^2 C_d^2.$$

- 6 Similarly, we obtain  $\bar{R}_T(r, s) = \frac{\bar{G}_T}{\bar{\rho}_T \otimes \bar{\rho}_T} \in L^2(\bar{\rho}_T \otimes \bar{\rho}_T)$  by applying (a) to get that that  $\bar{R}_T(r, s)^2 \bar{\rho}_T(r) \bar{\rho}_T(s) =$   
 7  $\frac{|\bar{G}_T(r, s)|^2}{\bar{\rho}_T(r) \bar{\rho}_T(s)} \leq |\mathcal{X}|^{2d-2} C_u^2 C_d^2$ , and hence,

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \bar{R}_T(r, s)^2 \bar{\rho}_T(r) \bar{\rho}_T(s) dr ds \leq \int_{\mathcal{X}} \int_{\mathcal{X}} |\mathcal{X}|^{2d-2} C_u^2 C_d^2 dr ds = |\mathcal{X}|^{2d} C_u^2 C_d^2.$$

8 □

- 9 By Lemma 3.4 and Theorem A.3,  $\mathcal{L}_{\bar{G}_T}$  is a positive compact self-adjoint operator in  $L^2(\mathcal{X})$ , and it has count-  
 10 ably many positive eigenvalues  $\{\lambda_i\}_{i=1}^\infty$  with orthonormal eigenfunctions  $\{\varphi_i\}_{i=1}^\infty$  (note that the eigenfunctions  
 11 of the eigenvalue  $\lambda = 0$  is excluded). In particular,  $\{\sqrt{\lambda_i} \varphi_i\}_{i=1}^\infty$  is an orthonormal basis of  $\mathcal{H}_G$ . The following  
 12 theorem follows directly.

- 13 **Theorem 3.5.** *The function space of identifiability by  $\mathcal{E}$  in (2.4) in  $L^2(\mathcal{X})$  is the  $L^2(\mathcal{X})$ -closure of  $\mathcal{H}_G$ , the*  
 14 *RKHS with reproducing kernel  $\bar{G}_T$  in (2.11).*

- 15 *Proof.* By Lemma 2.5, it suffices to show that  $\langle\langle f, f \rangle\rangle^2 > 0$  for any nonzero  $f$  in the  $L^2(\mathcal{X})$ -closure of  $\mathcal{H}_G$ .  
 16 Since  $\{\sqrt{\lambda_i} \varphi_i\}$  is an orthonormal basis of  $\mathcal{H}_G$ , its  $L^2(\mathcal{X})$ -closure is the closure of the eigenspace  $\text{span}\{\varphi_i\}$   
 17 corresponding to nonzero eigenvalues. Thus, if  $f = \sum_{i=1}^\infty c_i \varphi_i$  is nonzero, we have  $\langle f, f \rangle_{L^2(\mathcal{X})} = \sum_{i=1}^\infty c_i^2 > 0$ ,  
 18 which ensures that  $\langle\langle f, f \rangle\rangle^2 = \sum_{i=1}^\infty c_i^2 \lambda_i > 0$ . □

- 19 The RKHS  $\mathcal{H}_G$  has the nice feature of being data-informed: its reproducing kernel  $\bar{G}_T$  depends solely on  
 20 the data  $(u(x, t), x \in \mathbb{R}^d, t \in [0, T])$ . It provides a tool to investigate when the kernel is identifiable in  $L^2(\mathcal{X})$ .

- 21 **Theorem 3.6** (Identifiability in  $L^2(\mathcal{X})$ ). *For the loss functional  $\mathcal{E}$  in (2.4), the following statements are equiv-*  
 22 *alent.*

- 23 (a) *Identifiability holds in  $L^2(\mathcal{X})$ , i.e.,  $\langle\langle h, h \rangle\rangle > 0$  for any nonzero  $h \in L^2(\mathcal{X})$ .*  
 24 (b)  *$\mathcal{L}_{\bar{G}_T}$  is strictly positive.*  
 25 (c)  *$\mathcal{H}_G$  is dense in  $L^2(\mathcal{X})$ .*

- 26 *Moreover, for any  $\phi = \sum_{i=1}^n c_i \varphi_i \in \mathcal{H} = \text{span}\{\varphi_1, \dots, \varphi_n\}$  with  $\{\varphi_i\}$  being orthonormal eigenfunctions of*  
 27  *$\mathcal{L}_{\bar{G}_T}$  corresponding to positive eigenvalues  $\{\lambda_i\}$ , we have*

$$\langle\langle \phi, \phi \rangle\rangle = \sum_{i=1}^n c_i^2 \lambda_i, \quad \|\phi\|_{\mathcal{H}_G}^2 = \sum_{i=1}^n c_i^2 \lambda_i^{-1}, \quad \|\phi\|_{L^2(\mathcal{X})}^2 = \sum_{i=1}^n c_i^2. \quad (3.7)$$

- 28 *In particular, the bilinear form  $\langle\langle \cdot, \cdot \rangle\rangle$  satisfies the coercivity condition in  $\mathcal{H}$ :*

$$\langle\langle \phi, \phi \rangle\rangle \geq \min\{\sqrt{\lambda_i}\}_{i=1}^n \|\phi\|_{L^2(\mathcal{X})}^2, \quad \forall \phi \in \mathcal{H}.$$



1 *Proof.* (a)  $\Rightarrow$  (b). Suppose  $\mathcal{L}_{\overline{G}_T}$  is not strictly positive, then there exists an eigenfunction  $\phi \in L^2(\mathcal{X})$  corre-  
 2 sponding to eigenvalue 0. But we would also have  $\langle\langle \phi, \phi \rangle\rangle = \langle \mathcal{L}_{\overline{G}_T} \phi, \phi \rangle_{L^2(\mathcal{X})} = 0$ .

3 (b)  $\Rightarrow$  (a). If  $\overline{G}_T$  is strictly positive, then  $\{\varphi_i\}_{i=1}^\infty$  would be an orthonormal basis for  $L^2(\mathcal{X})$  and all eigenva-  
 4 lues of  $\mathcal{L}_{\overline{G}_T}$  are positive. Take  $\phi = \sum_{j=1}^\infty c_j \varphi_j$ . Then,  $\langle\langle \phi, \phi \rangle\rangle = 0$  implies

$$\langle \mathcal{L}_{\overline{G}_T} \phi, \phi \rangle_{L^2(\mathcal{X})} = \langle \sum_{j=1}^\infty c_j \lambda_j \varphi_j, \sum_{j=1}^\infty c_j \varphi_j \rangle_{L^2(\mathcal{X})} = \sum_{j=1}^\infty c_j^2 \lambda_j = 0.$$

5 Hence  $\phi = 0$  in  $L^2(\mathcal{X})$ .

6 (b)  $\Leftrightarrow$  (c). Note that  $\{\sqrt{\lambda_i} \varphi_i\}_{i=1}^\infty$  is a basis of  $\mathcal{H}_G$ . Thus,  $\mathcal{H}_G$  is dense in  $L^2(\mathcal{X})$  iff  $\{\varphi_i\}_{i=1}^\infty$  is a basis of  
 7  $L^2(\mathcal{X})$ , i.e.  $\mathcal{L}_{\overline{G}_T}$  is strictly positive.

8 At last, for any  $\phi = \sum_{j=1}^n c_j \varphi_j \in \mathcal{H} = \text{span}\{\varphi_1, \dots, \varphi_n\}$ , we have

$$\langle\langle \phi, \phi \rangle\rangle^2 = \langle \mathcal{L}_{\overline{G}_T} \phi, \phi \rangle_{L^2(\mathcal{X})} = \sum_{j=1}^n c_j^2 \lambda_j \geq \min\{\lambda_i\}_{i=1}^n \|\phi\|_{L^2(\mathcal{X})}^2.$$

9 Also, we have  $\phi = \sum_{j=1}^n c_j \lambda_j^{-1/2} \sqrt{\lambda_j} \varphi_j \in \mathcal{H}_G$  with  $\|\phi\|_{\mathcal{H}_G}^2 = \sum_{i=1}^n c_i^2 \lambda_i^{-1}$ ; □

10 **3.3. Identifiability in the weighted L2 space.** In this section, we study the identifiability in  $L^2_{\overline{\rho}_T}$  through the  
 11 RKHS whose reproducing kernel is a weighted integral kernel. Since the results are mostly the same as those  
 12 in  $L^2(\mathcal{X})$ , so we only briefly state the main results, then focus on discussing their relations.

13 We define the following kernel  $\overline{R}_T$  on the set  $\mathcal{X} \times \mathcal{X}$ :

$$\overline{R}_T(r, s) = \frac{\overline{G}_T(r, s)}{\overline{\rho}_T(r) \overline{\rho}_T(s)}, \quad (r, s) \in \mathcal{X} \times \mathcal{X}. \quad (3.8)$$

14 The function  $\overline{R}_T$  is a positive definite kernel, since  $\overline{G}_T$  is by Lemma 3.3. Additionally, by Lemma 3.4, the  
 15 kernel  $\overline{R}_T \in L^2(\overline{\rho}_T \otimes \overline{\rho}_T)$ , so that  $\mathcal{L}_{\overline{R}_T}$ . Thus, it defines a compact integral operator

$$\mathcal{L}_{\overline{R}_T} \varphi(s) = \int_{\mathbb{R}^+} \overline{R}_T(r, s) \varphi(r) \overline{\rho}_T(r) dr, \quad (3.9)$$

16 and it satisfies  $\langle\langle \varphi, \psi \rangle\rangle = \langle \varphi, \mathcal{L}_{\overline{R}_T} \psi \rangle_{L^2_{\overline{\rho}_T}}$ .

17 All the results for  $L^2(\mathcal{X})$  extends to  $L^2_{\overline{\rho}_T}$ . Importantly,  $\mathcal{L}_{\overline{R}_T}$  is a positive compact self-adjoint operator in  
 18  $L^2_{\overline{\rho}_T}$ , and it has countably many positive eigenvalues  $\{\gamma_i\}_{i=1}^\infty$  with orthonormal eigenfunctions  $\{\psi_i\}_{i=1}^\infty$ . In  
 19 particular,  $\{\sqrt{\gamma_i} \psi_i\}_{i=1}^\infty$  is an orthonormal basis of  $\mathcal{H}_R$ . Similar to Theorem 3.5–Theorem 3.6, the identifiability  
 20 holds in  $L^2_{\overline{\rho}_T}$  if the RKHS  $\mathcal{H}_R$  is dense in it. The following theorem summarizes these results.

21 **Theorem 3.7** (Identifiability in  $L^2_{\overline{\rho}_T}$ ). *The function space of identifiability in  $L^2_{\overline{\rho}_T}$  is the  $L^2_{\overline{\rho}_T}$ -closure of  $\mathcal{H}_R$ , the*  
 22 *RKHS with reproducing kernel  $\overline{R}_T$  in (3.8). The following are equivalent.*

23 (a) *Identifiability holds in  $L^2_{\overline{\rho}_T}$ .*

24 (b)  *$\mathcal{L}_{\overline{R}_T}$  in (3.9) is strictly positive.*

25 (c)  *$\mathcal{H}_R$  is dense in  $L^2_{\overline{\rho}_T}$ .*

26 *Moreover, for any  $\phi = \sum_{i=1}^n c_i \psi_i \in \mathcal{H} = \text{span}\{\psi_1, \dots, \psi_n\}$  with  $\{\psi_i\}$  being orthonormal eigenfunctions of*  
 27  *$\mathcal{L}_{\overline{R}_T}$  corresponding to eigenvalues  $\{\gamma_i > 0\}$ , we have*

$$\langle\langle \phi, \phi \rangle\rangle^2 = \sum_{i=1}^n c_i^2 \gamma_i, \quad \|\phi\|_{\mathcal{H}_R}^2 = \sum_{i=1}^n c_i^2 \gamma_i^{-1}, \quad \|\phi\|_{L^2_{\overline{\rho}_T}}^2 = \sum_{i=1}^n c_i^2. \quad (3.10)$$

28 *In particular, the bilinear form  $\langle\langle \cdot, \cdot \rangle\rangle$  satisfies the coercivity condition in  $\mathcal{H}$ :*

$$\langle\langle \phi, \phi \rangle\rangle \geq \min\{\sqrt{\gamma_i}\}_{i=1}^n \|\phi\|_{L^2_{\overline{\rho}_T}}^2, \quad \forall \phi \in \mathcal{H}.$$

1 **Remark 3.8** (Relation to the coercivity condition of the bilinear form). *Recall that a bilinear form  $\langle\langle \cdot, \cdot \rangle\rangle$  is*  
2 *said to be coercive on a subspace  $\mathcal{H} \subset L^2_{\bar{\rho}_T}$  if there exists a constant  $c_{\mathcal{H}} > 0$  such that  $\langle\langle \psi, \psi \rangle\rangle \geq c_{\mathcal{H}} \|\psi\|_{L^2_{\bar{\rho}_T}}^2$*   
3 *for all  $\psi \in \mathcal{H}$ . Such a coercivity condition has been introduced on subspaces of  $L^2_{\bar{\rho}_T}$  in [4, 23, 21, 22, 19]*  
4 *for systems of finitely many particles. Theorem 3.7 shows that, for any finite-dimensional hypothesis space*  
5  *$\mathcal{H} = \text{span}\{\psi_1, \dots, \psi_n\}$ , the coercivity condition holds with  $c_{\mathcal{H}} = \min\{\gamma_i\}_{i=1}^n$ , but the coercivity constant*  
6 *vanishes as the dimension of  $\mathcal{H}$  increases to infinity.*

7 Now we have two RHKSs,  $\mathcal{H}_G$  and  $\mathcal{H}_R$ , whose closures in  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$  are the function spaces of  
8 identifiability. They are the images  $\mathcal{L}_{\bar{G}_T}^{1/2}(L^2(\mathcal{X}))$  and  $\mathcal{L}_{\bar{R}_T}^{1/2}(L^2_{\bar{\rho}_T})$  (see Appendix A.1). The following remarks  
9 discuss their relations.

10 **Remark 3.9** (The two integral operators). *The integral operators  $\mathcal{L}_{\bar{G}_T}$  and  $\mathcal{L}_{\bar{R}_T}$  are derived from the same*  
11 *bilinear form: for any  $\varphi, \psi \in L^2(\mathcal{X}) \subset L^2_{\bar{\rho}_T}$ , we have*

$$\langle\langle \varphi, \psi \rangle\rangle = \langle \mathcal{L}_{\bar{G}_T} \varphi, \psi \rangle_{L^2(\mathcal{X})} = \langle \mathcal{L}_{\bar{R}_T} \varphi, \psi \rangle_{L^2(\bar{\rho}_T)}.$$

12 *Since  $L^2(\mathcal{X})$  is dense in  $L^2_{\bar{\rho}_T}$ , the second equality implies that the null-space of and  $\mathcal{L}_{\bar{G}_T}$  is a subset of  $\mathcal{L}_{\bar{R}_T}$ .*  
13 *However, there is no correspondence between their eigenfunctions of nonzero eigenvalues. To see this, let*  
14  *$\varphi_i \in L^2(\mathcal{X})$  be an eigenfunction of  $\mathcal{L}_{\bar{G}_T}$  with eigenvalue  $\lambda_i > 0$ . Then, it follows from the second equality that*

$$\langle \lambda_i \varphi_i, \psi \rangle_{L^2(\mathcal{X})} = \langle \mathcal{L}_{\bar{G}_T} \varphi_i, \psi \rangle_{L^2(\mathcal{X})} = \langle \mathcal{L}_{\bar{R}_T} \varphi_i, \psi \rangle_{L^2(\bar{\rho}_T)} = \langle \bar{\rho}_T \mathcal{L}_{\bar{R}_T} \varphi_i, \psi \rangle_{L^2(\mathcal{X})}$$

15 *for any  $\psi \in L^2(\mathcal{X})$ . Thus,  $\lambda_i \varphi_i = \bar{\rho}_T \mathcal{L}_{\bar{R}_T} \varphi_i$  in  $L^2(\mathcal{X})$ . Then, neither  $\bar{\rho}_T \varphi_i$  nor  $\frac{\varphi_i}{\bar{\rho}_T}$  is an eigenfunction of  $\mathcal{L}_{\bar{R}_T}$ .*

16 **Remark 3.10** (Metrics on  $\mathcal{H}_G$  and  $\mathcal{H}_R$ ). *We have three metrics on  $\mathcal{H}_G \subset L^2(\mathcal{X})$ : the RKHS norm, the  $L^2(\mathcal{X})$*   
17 *norm and the norm induced by the bilinear form. By (3.7), these three metrics satisfy*

$$\sqrt{\langle\langle \phi, \phi \rangle\rangle} \leq \max_i \{\sqrt{\lambda_i}\} \|\phi\|_{L^2(\mathcal{X})} \leq \max_i \{\lambda_i\} \|\phi\|_{\mathcal{H}_G}$$

18 *for any  $\phi \in \mathcal{H}_G$ . Similarly, there are three metrics on  $\mathcal{H}_R \subset L^2_{\bar{\rho}_T}$  satisfying, for any  $\phi \in \mathcal{H}_R$ ,*

$$\sqrt{\langle\langle \phi, \phi \rangle\rangle} \leq \max_i \{\sqrt{\gamma_i}\} \|\phi\|_{L^2_{\bar{\rho}_T}} \leq \max_i \{\gamma_i\} \|\phi\|_{\mathcal{H}_R}.$$

19 **Remark 3.11** (Relaxing the assumption on data). *It is possible to relax the technical assumptions that the*  
20 *solution  $u$  is continuous with bounded support and consider measure-valued solutions or unbounded support.*  
21 *These assumptions are used to prove the integrability of the integral kernel  $\bar{G}_T$  in Lemmas 3.3-3.4 so that the*  
22 *integrator  $\mathcal{L}_{\bar{G}_T}$  is a bounded operator. But the  $\mathcal{L}_{\bar{G}_T}$  can remain to be bounded when the data  $u$  is measure-*  
23 *valued or has unbounded support. Furthermore, the weighted space  $L^2(\bar{\rho}_T)$  can be used to study the inference*  
24 *of singular kernels. We leave the more involving analysis as future work.*

25 **4. Non-radial interaction kernels.** Non-radial interaction kernels are important because they provide more  
26 flexibility for modeling than radial kernels. We extend the identifiability analysis to non-radial kernels, using  
27 the same arguments as for the radial case.

28 Throughout this section, we consider the non-radial vector-valued kernels  $K_\phi = \nabla \Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  being  
29 a gradient of an interaction potential  $\Phi$ , where the subscript  $\phi$  emphasizes that the kernel is a gradient. We  
30 analyze the function space of identifiability by the loss functional in (2.4). The next theorem presents the main  
31 results.

32 **Theorem 4.1** (Identifiability of non-radial kernel). *Given data  $(u(x, t) : (x, t) \in \mathbb{R}^d \times [0, T])$  satisfying*  
33 *Assumption 2.1, consider the estimation of the kernel  $K_{\text{true}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  by minimizing the loss functional*

$$\mathcal{E}(K_\psi) := \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} \left[ |K_\psi * u|^2 + 2\partial_t u(\Psi * u) + 2\nu \nabla u \cdot (K_\psi * u) \right] dx dt, \quad (4.1)$$

34 where  $K_\psi(x) = \nabla\Psi(x)$  in either  $L^2(\mathcal{X})$  or  $L^2_{\bar{\rho}_T} := \{K : \mathbb{R}^d \rightarrow \mathbb{R}^d, \int_{\mathbb{R}^d} |K(x)|^2 \bar{\rho}_T(x) dx < \infty\}$ , where  
 1  $\mathcal{X} = \text{supp}(\bar{\rho}_T)$  with  $\bar{\rho}_T$  is defined by

$$\bar{\rho}_T(x) = \frac{1}{T} \int_0^T \rho_t(x) dt = \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} u(y-x, t) u(y, t) dy dt. \quad (4.2)$$

2 Let  $F_t(x, y) := \int_{\mathbb{R}^d} u(z-y, t) u(z-x, t) u(z, t) dz$  and define  $\bar{F}_T$  and  $\bar{Q}_T$  as

$$\bar{F}_T(x, y) = \frac{1}{T} \int_0^T F_t(x, y) dt, \quad \bar{Q}_T(x, y) = \frac{\bar{F}_T(x, y)}{\bar{\rho}_T(x) \bar{\rho}_T(y)}, \quad (x, y) \in \mathcal{X} \times \mathcal{X}. \quad (4.3)$$

3 Then, the function space of identifiability in  $L^2_{\bar{\rho}_T}$  by the loss functional  $\mathcal{E}$  is the  $L^2_{\bar{\rho}_T}$ -closure of  $\mathcal{H}_Q$ , the RKHS  
 4 with  $\bar{Q}_T$  as the reproducing kernel. Additionally, the following are equivalent:

- 5 (a) Identifiability holds in  $L^2_{\bar{\rho}_T}$ .  
 6 (b) The operator  $\mathcal{L}_{\bar{Q}_T} : L^2(\bar{\rho}_T) \rightarrow L^2(\bar{\rho}_T)$  in (4.4) is strictly positive:

$$\mathcal{L}_{\bar{Q}_T} f(x) = \int_{\mathbb{R}^d} \bar{Q}_T(y, x) f(y) \bar{\rho}_T(y) dy \quad (4.4)$$

7 (c)  $\mathcal{H}_Q$  is dense in  $L^2_{\bar{\rho}_T}$ .

8 Similarly, these claims hold in  $L^2(\mathcal{X})$  by considering the  $L^2(\mathcal{X})$ -closure of  $\mathcal{H}_F$ , the RKHS with  $\bar{F}_T$  as the  
 9 reproducing kernel, and the corresponding integral operator  $\mathcal{L}_{\bar{F}_T}$ .

10 *Proof.* The proof is mostly the same as the proofs of Theorems 3.5– 3.6. It consists of three steps.

- 11 • Show that  $\bar{Q}_T$  and  $\bar{F}_T$  are square-integrable reproducing kernels, so their RKHSs are well-defined. Con-  
 12 sequently, their integral operators  $\mathcal{L}_{\bar{Q}_T}$  and  $\mathcal{L}_{\bar{F}_T}$  are semi-positive.  
 13 • Extend Lemma 2.5 to vector-valued functions by showing that loss functional  $\mathcal{E}$  has a unique minimizer  
 14 in a linear space  $\mathcal{H}$  if and only if  $\langle\langle K, K \rangle\rangle > 0$  for any nonzero  $K \in \mathcal{H}$ . Here  $\langle\langle K_1, K_2 \rangle\rangle$  is a bilinear form  
 15 for vector-valued functions  $K_1, K_2 : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , defined by

$$\begin{aligned} \langle\langle K_1, K_2 \rangle\rangle &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} (K_1 * u) \cdot (K_2 * u) u(x, t) dx dt \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_1(y) \cdot K_2(z) \bar{F}_T(y, z) dy dz = \sum_{i=1}^d \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_1^i(y) K_2^i(z) \bar{F}_T(y, z) dy dz. \end{aligned} \quad (4.5)$$

16 The extension is straightforward, because  $\mathcal{E}(K_\psi) = \langle\langle K_\psi, K_\psi \rangle\rangle - 2\langle\langle K_\psi, K_{true} \rangle\rangle$  if  $K_{true}$  is the true  
 17 kernel generating the data  $u$ .

- 18 • Show the spectral characterization of the FSOI and the equivalence between (a)-(c), which follow from  
 19 the same proofs for Theorems 3.5– 3.6.

20 Thus, we only need to prove that  $\bar{F}_T$  and  $\bar{Q}_T$  are square-integrable reproducing kernels, which we do in Lemma  
 21 4.2 below.  $\square$

22 **Lemma 4.2.** *Under Assumption 2.1, the functions  $\bar{F}_T, \bar{Q}_T : \mathbb{R}^d \rightarrow \mathbb{R}$  in (4.3) are symmetric, positive definite,*  
 23 *and satisfy the following properties:*

- 24 (a) For all  $x, y \in \mathbb{R}^d$ ,  $\bar{F}_T(x, y) \leq C_u \bar{\rho}_T(x)$ , where  $C_u = \sup_{x \in \mathbb{R}^d, t \in [0, T]} u(x, t)$ .  
 25 (b) The function  $\bar{F}_T$  is in  $L^2(\mathbb{R}^d \times \mathbb{R}^d)$ , and  $\bar{Q}_T$  is in  $L^2_{\bar{\rho}_T \otimes \bar{\rho}_T}$ .

26 *Proof.* Both functions are symmetry by definition. They are positive definite similar to the proof of Lemma 3.3.

27 Part (a) follows from (4.3) and that  $C_u = \sup_{z \in \mathbb{R}^d, t \in [0, T]} u(z-x, t)$  for any  $x$ ,

$$\begin{aligned} \bar{F}_T(x, y) &= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} u(z-y, t) u(z-x, t) u(z, t) dz dt \\ &\leq C_u \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} u(z-y, t) u(z, t) dz dt = C_u \bar{\rho}_T(y), \end{aligned}$$

28 where the last equality follows from the definition of  $\bar{\rho}_T$ .

1 For (b), note that by symmetry, we have  $\bar{F}_T(x, y) \leq C_u \bar{\rho}_T(y)$  for any  $x, y \in \mathbb{R}^d$ . Then,

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\bar{F}_T(x, y))^2 dx dy \leq C_u^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \bar{\rho}_T(x) \bar{\rho}_T(y) dx dy = C_u^2.$$

2 To show that  $\bar{Q}_T$  is square-integrable, we make use of the assumption that the data  $u$  has bounded support,  
3 which implies that the support of  $\bar{\rho}_T$ , denoted by  $\mathcal{X} = \text{supp}(\bar{\rho}_T)$ , is bounded. Hence,

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \bar{Q}_T^2(x, y) \bar{\rho}_T(x) \bar{\rho}_T(y) dx dy = \int_{\mathcal{X}} \int_{\mathcal{X}} \frac{(\bar{F}_T(x, y))^2}{\bar{\rho}_T(x) \bar{\rho}_T(y)} dx dy \leq C_u^2 |\mathcal{X}|^2,$$

4 where the inequality follows from (a).  $\square$

5 We note that the assumption on  $u$  having a bounded support is sufficient but not necessary for  $\bar{Q}_T \in L^2_{\bar{\rho}_T \otimes \bar{\rho}_T}$ .  
6 When the support of  $u$  is unbounded, the function  $\bar{Q}_T$  may not be square-integrable. The following two ex-  
7 amples show that  $\bar{Q}_T$  is square-integrable when  $u$  is the probability density function of a stationary Gaussian  
8 process, but it is not when  $u$  is the density of a Cauchy distribution.

9 **Example 4.3** (Square-integrable  $\bar{Q}_T$ ). We show that  $\bar{Q}_T \in L^2_{\bar{\rho}_T \otimes \bar{\rho}_T}$  when  $d = 1$  and  $u(x, 0) = U(x) =$   
10  $\frac{1}{\sqrt{2\pi\nu}} e^{-\frac{x^2}{2\nu}}$ . First, we show that  $U(x)$  is a stationary solution to the mean-field equation (2.1) (equivalently,  
11  $\mathcal{N}(0, \nu)$  is an invariant density of the SDE (2.7)) with  $K(x) = x$ . In fact, noting that  $K * U(x) = \int_{\mathbb{R}} (x -$   
12  $y)U(y)dy = x$ , one can verify directly that  $\nu U''(x) + [xU(x)]' = 0$  (similarly, the SDE (2.7) becomes the  
13 Ornstein-Uhlenbeck process  $d\bar{X}_t = -\bar{X}_t dt + \sqrt{2\nu} dB_t$  and  $U$  is its invariant density). Second, we compute  $\bar{F}_T$   
14 and  $\bar{\rho}_T$  directly from their definitions. Since  $u(x, t) = U(x)$  for each  $t$ , by definition of  $\bar{F}_T$  in (4.5):

$$\begin{aligned} \bar{F}_T(x, y) &= \int_{\mathbb{R}^d} U(z-x)U(z-y)U(z)dz \\ &= \frac{1}{2\pi\nu} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{1}{2\nu}((z-x)^2 + (z-y)^2 + z^2)} dz = \frac{\sqrt{3}}{2\pi\nu} e^{-\frac{1}{3\nu}(x^2 + y^2 - xy)}. \end{aligned}$$

15 Since  $\bar{\rho}_T$  is the density of  $\bar{X}_t - \bar{X}'_t$  with  $\bar{X}'_t$  being an independent copy of  $\bar{X}_t$ , which has the stationary density  
16  $U$ , we have  $\bar{\rho}_T(x) = \frac{1}{2\sqrt{\pi\nu}} e^{-\frac{x^2}{4\nu}}$ . Hence, the kernel  $\bar{Q}_T$  is square-integrable due to the fast decay of  $\bar{F}_T$ :

$$\|\bar{Q}_T\|_{L^2(\bar{\rho}_T \otimes \bar{\rho}_T)}^2 = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\bar{F}_T(x, y)^2}{\bar{\rho}_T(x) \bar{\rho}_T(y)} dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{3}{\pi\nu} e^{-\frac{1}{12\nu^2}[4(x-y)^2 + x^2 + y^2]} dx dy < \infty.$$

17 **Example 4.4** (Non-square-integrable  $\bar{Q}_T$ ). We show that  $\bar{Q}_T \notin L^2_{\bar{\rho}_T \otimes \bar{\rho}_T}$  when  $d = 1$ ,  $\nu = 1$  and  $u(x, 0) =$   
18  $U(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ , which is the density of Cauchy distribution. Suppose that  $K$  has a Fourier transform satisfying  
19  $\hat{K}(\xi) \hat{U}(\xi) = \int_{\mathbb{R}} \frac{2x}{1+x^2} e^{i\xi x} dx$ , in other words,  $[K * U](x) = \frac{2x}{1+x^2}$ . First, note that  $U$  is a steady solution to  
20 (2.1) because

$$\nu U''(x) = -\frac{d}{dx} \left[ \frac{2x}{1+x^2} U(x) \right] = -\frac{d}{dx} [U(x)[K * U](x)].$$

21 Second, direction computation (see Appendix A.2 for the details) yields

$$\begin{aligned} \bar{F}_T(x, y) &= \int_{\mathbb{R}} U(z-x)U(z-y)U(z)dz = \frac{1}{\pi^3} \int_{\mathbb{R}} \frac{1}{1+(z-x)^2} \frac{1}{1+(z-y)^2} \frac{1}{1+z^2} dz \\ &= \frac{2}{\pi^2} \frac{(x^2 - xy + y^2 + 12)}{(x^2 + 4)(y^2 + 4)(x^2 - 2xy + y^2 + 4)}. \end{aligned}$$

22 Meanwhile, since  $\bar{\rho}_T$  is the density of  $\bar{X}_t - \bar{X}'_t$  with  $\bar{X}'_t$  being an independent copy of  $\bar{X}_t$ , which has the  
23 stationary density  $U$ , we have  $\bar{\rho}_T(x) = (U(\cdot) * U(\cdot))(x) = \frac{2}{\pi} \frac{1}{(x^2+4)}$  (see Appendix A.2 for the computation  
24 details). Lastly,  $\bar{Q}_T$  is not square integrable because

$$\|\bar{Q}_T\|_{L^2(\bar{\rho}_T \otimes \bar{\rho}_T)}^2 = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\bar{F}_T(x, y)^2}{\bar{\rho}_T(x) \bar{\rho}_T(y)} dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(x^2 - xy + y^2 + 12)}{2(x^2 - 2xy + y^2 + 4)} dx dy = \infty.$$

25 **5. Identifiability in computational practice.** In this section, we discuss the implications of the identifiability  
 1 theory for computational practice. For simplicity, we consider only radial interaction kernels and  $d = 1$ . We  
 2 show that the regression matrix becomes ill-conditioned as the dimension of the hypothesis space increases  
 3 (see Theorem 5.1). Thus, regularization becomes necessary to avoid amplification of the numerical errors.  
 4 We compare  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$  in the context of truncated singular value decomposition (SVD) regularization.  
 5 Numerical tests in Section 5.3 suggest that the  $L^2_{\bar{\rho}_T}$  norm leads to more accurate regularized estimators, and a  
 6 better-conditioned inversion (see Figure 1–3).

7 **5.1. Nonparametric regression in practice.** In computational practice, the data is on discrete space mesh  
 8 grids, and our goal is to find a minimizer of the loss functional by least squares as in Remark 2.3. We review  
 9 only those fundamental elements, and we refer to [17] for more details.

10 First, we select a set of data-adaptive basis functions in  $L^2_{\bar{\rho}_T}$  or  $L^2(\mathcal{X})$  to avoid a singular normal matrix.  
 11 The starting point is to approximate empirically the measure  $\bar{\rho}_T$  in (2.9) from data and to obtain its support  
 12  $\mathcal{X}$ . Let  $\{r_i\}_{i=0}^n$  be a uniform partition of  $\mathcal{X}$  and denote the width of each interval as  $\Delta r$ . They provide the  
 13 knots for the B-spline basis functions of  $\mathcal{H}$ . Here we use piecewise constant basis functions to facilitate the rest  
 14 discussions, that is,  $\mathcal{H} = \text{span}\{\phi_1, \dots, \phi_n\}$ , with  $\phi_i(r) = \mathbf{1}_{[r_{i-1}, r_i]}(r)$ . One may also use other partitions, for  
 15 example, a partition with uniform probability for all intervals, as well as other basis functions, such as higher  
 16 degree B-splines or weighted orthogonal polynomials.

17 Second, as outlined in Remark 2.3, we compute the normal matrix  $A_n$  and vector  $b_n$  from data. Since  $\phi_{true}$   
 18 is unknown, the vector  $b_n$  is computed from data, following the loss functional in (2.4):

$$\begin{aligned} A_n(ij) &= \langle\langle \phi_i, \phi_j \rangle\rangle = \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} (K_{\phi_i} * u) \cdot (K_{\phi_j} * u) u(x, t) dx dt, \\ b_n(i) &= -\frac{1}{T} \int_0^T \int_{\mathbb{R}^d} [\partial_t u (\Phi_i * u) + \nu \nabla u \cdot (K_{\phi_i} * u)] dx dt, \end{aligned} \quad (5.1)$$

19 where  $\Phi_i(r) = \int_0^r \phi_i(s) ds$  is an anti-derivative of  $\phi_i$ . The integrals in the entries of  $A_n$  and  $b_n$  are approximated  
 20 from data by the Riemann sum.

21 Then, the minimizer of the loss functional in  $\mathcal{H}$  is solved from the linear equation  $A_n \hat{c} = b_n$ . When the  
 22 normal matrix  $A_n$  is well-conditioned, we compute the minimizer by  $\hat{c} = A_n^{-1} b_n$ . When  $A_n$  is ill-conditioned  
 23 or singular, which happens often as  $n$  increases, the (pseudo-)inverse of  $A_n$  tends to amplify the numerical error  
 24 in  $b_n$ . Thus, we need regularization (see Section 5.3).

25 **5.2. Identifiability and ill-conditioned normal matrix.** We show first that the eigenvalues of the integral  
 26 operators are generalized eigenvalues of the normal matrix.

27 **Theorem 5.1.** *Let  $\mathcal{H} = \text{span}\{\phi_i\}_{i=1}^n \subset L^2(\mathcal{X})$ , where the basis functions are linearly independent in  $L^2(\mathcal{X})$   
 28 and  $L^2_{\bar{\rho}_T}$ . Recall the normal matrix  $A_n$  in (5.1), the operators  $\mathcal{L}_{\bar{G}_T} : L^2(\mathcal{X}) \rightarrow L^2(\mathcal{X})$  in (3.5) and  $\mathcal{L}_{\bar{R}_T} :$   
 29  $L^2_{\bar{\rho}_T} \rightarrow L^2_{\bar{\rho}_T}$  in (3.9). The following statements hold true.*

30 (a) *If  $\mathcal{L}_{\bar{G}_T} \varphi = \lambda \varphi$  for some  $\varphi = \sum_{i=1}^n c_i \phi_i \in \mathcal{H}$ , then  $\lambda$  is a generalized eigenvalue of  $A_n$ :*

$$A_n c = \lambda B_n^G c, \quad \text{with } B_n^G = (\langle \phi_i, \phi_j \rangle_{L^2(\mathcal{X})})_{1 \leq i, j \leq n}. \quad (5.2)$$

31 *In particular, if  $\{\phi_i\}$  are piecewise constants on intervals with length  $\Delta r$  in a uniform partition of  $\mathcal{X}$ ,  
 32 then,  $\lambda \Delta r$  is an eigenvalue of  $A_n$ .*

33 (b) *Similarly, if  $\mathcal{L}_{\bar{R}_T} \varphi = \lambda \varphi$  for some  $\varphi = \sum_{i=1}^n c_i \phi_i \in \mathcal{H}$ , then  $\lambda$  is generalize eigenvalue of  $A_n$ :*

$$A_n c = \lambda B_n^R c, \quad \text{with } B_n^R = (\langle \phi_i, \phi_j \rangle_{L^2_{\bar{\rho}_T}})_{1 \leq i, j \leq n}. \quad (5.3)$$

34 *Proof.* For Part (a), since  $\mathcal{L}_{\bar{G}_T} \varphi = \lambda \varphi$  with  $\varphi = \sum_{i=1}^n c_i \phi_i$ , we have

$$\lambda (B_n^G c)_k = \langle \lambda \varphi, \phi_k \rangle_{L^2(\mathcal{X})} = \langle \mathcal{L}_{\bar{G}_T} \varphi, \phi_k \rangle_{L^2(\mathcal{X})} = \sum_{i=1}^n c_i \langle \mathcal{L}_{\bar{G}_T} \phi_i, \phi_k \rangle_{L^2(\mathcal{X})} = (A_n c)_k,$$

where the last equality follows from (3.6). Thus,  $\lambda$  is a generalized eigenvalue of  $(A_n, B_n^G)$ . Note that  $B_n^G = \Delta r I_n$  if  $\{\phi_i\}$  are piecewise constants on intervals with length  $\Delta r$ . Thus, by (5.2),  $\lambda \Delta r$  is an eigenvalue of  $A_n$ . Part (b) follows similarly.  $\square$

We summarize the notations in these two generalized eigenvalue problems in Table 2.

TABLE 2. Notation of variables in the eigenvalue problems on  $\mathcal{H}$ .

	in $L^2(\mathcal{X})$	in $L^2_{\rho_T}$
integral kernel and operator	$\widehat{G}_T \approx \overline{G}_T, \mathcal{L}_{\widehat{G}_T} \approx \mathcal{L}_{\overline{G}_T}$	$\widehat{R}_T \approx \overline{R}_T, \mathcal{L}_{\widehat{R}_T} \approx \mathcal{L}_{\overline{R}_T}$
eigenfunction and eigenvalue	$\mathcal{L}_{\widehat{G}_T} \widehat{\varphi}_k = \widehat{\lambda}_k \widehat{\varphi}_k$	$\mathcal{L}_{\widehat{R}_T} \widehat{\psi}_k = \widehat{\gamma}_k \widehat{\psi}_k$
eigenvector and eigenvalue	$A_n \widehat{\varphi}_k = \widehat{\lambda}_k B_n^G \widehat{\varphi}_k$	$A_n \widehat{\psi}_k = \widehat{\gamma}_k B_n^R \widehat{\psi}_k$

**Remark 5.2** (Ill-conditioned normal matrix). *As the dimensions of  $\mathcal{H} = \text{span}\{\phi_i\}_{i=1}^n$  increases, the normal matrix  $A_n$  becomes ill-conditioned. This is because  $A_n$  approximates the compact operators  $\mathcal{L}_{\overline{G}_T}$  in  $L^2(\mathcal{X})$  and  $\mathcal{L}_{\overline{R}_T}$  in  $L^2_{\rho_T}$ , in the sense that*

$$A_n(i, j) = \langle\langle \phi_i, \phi_j \rangle\rangle = \langle \mathcal{L}_{\overline{G}_T} \phi_i, \phi_j \rangle_{L^2(\mathcal{X})} = \langle \mathcal{L}_{\overline{R}_T} \phi_i, \phi_j \rangle_{L^2_{\rho_T}}, \forall \phi_i, \phi_j \in \mathcal{H}.$$

Thus, Theorem 5.1 indicates that, as  $n$  increases, the generalized eigenvalues of  $A_n$ , with respect to  $B_n^G$  and  $B_n^R$ , converge to those of  $\mathcal{L}_{\overline{G}_T}$  and  $\mathcal{L}_{\overline{R}_T}$ , respectively. Then, the ratio  $\lambda_{max}^{G,n}/\lambda_{min}^{G,n}$  increases to infinity, where we let  $\lambda_{max}^{G,n}$  and  $\lambda_{min}^{G,n}$  be the maximal and minimal generalized eigenvalues of  $(A_n, B_n^G)$ . Let  $\lambda_{max}^{A_n}$  and  $\lambda_{min}^{A_n}$  be the maximal and minimal eigenvalues of  $A_n$ , and similarly,  $\lambda_{max}^{B_n^G}$  and  $\lambda_{min}^{B_n^G}$  for  $B_n^G$ . Note that

$$\lambda_{max}^{A_n} = \max_{c \in \mathbb{R}^n} \frac{c^\top A_n c}{c^\top B_n^G c} \frac{c^\top B_n^G c}{c^\top c} \geq \max_{c \in \mathbb{R}^n} \frac{c^\top A_n c}{c^\top B_n^G c} \min_{c \in \mathbb{R}^n} \frac{c^\top B_n^G c}{c^\top c} = \lambda_{max}^{G,n} \lambda_{min}^{B_n^G},$$

$$\lambda_{min}^{A_n} = \min_{c \in \mathbb{R}^n} \frac{c^\top A_n c}{c^\top B_n^G c} \frac{c^\top B_n^G c}{c^\top c} \leq \min_{c \in \mathbb{R}^n} \frac{c^\top A_n c}{c^\top B_n^G c} \max_{c \in \mathbb{R}^n} \frac{c^\top B_n^G c}{c^\top c} = \lambda_{min}^{G,n} \lambda_{max}^{B_n^G}.$$

Therefore, the conditional number of  $A_n$  is bounded below as

$$\frac{\lambda_{max}^{A_n}}{\lambda_{min}^{A_n}} \geq \frac{\lambda_{max}^{G,n}}{\lambda_{min}^{G,n}} \times \frac{\lambda_{min}^{B_n^G}}{\lambda_{max}^{B_n^G}}.$$

Consequently, the matrix  $A_n$  becomes increasing ill-conditioned as  $\mathcal{H}$  enlarges, since the ratio  $\lambda_{min}^{B_n^G}/\lambda_{max}^{B_n^G}$  remains bounded for suitable basis functions.

**Remark 5.3** (Ill-posed inverse problem). *The inverse problem is ill-posed in general: since  $A_n$  becomes ill-conditioned as  $n$  increases, a small perturbation in  $b_n$  may lead to large errors in the estimator. More specifically, we are solving the inverse problem  $\widehat{\phi} = \mathcal{L}_{\overline{R}_T}^{-1}(\mathcal{L}_{\overline{R}_T} \phi_{true})$  in  $L^2_{\rho_T}$ , where  $\mathcal{L}_{\overline{R}_T}^{-1}$  is an unbounded operator. The normal matrix  $A_n$  approximates the operator  $\mathcal{L}_{\overline{R}_T}$ , and  $b_n$  approximates  $\mathcal{L}_{\overline{R}_T} \phi_{true}$ . The error in  $b_n$  in the eigenspace of small eigenvalues will be amplified by the inversion, leading to an ill-posed inverse problem.*

**5.3. Truncated SVD regularization in the L2 spaces.** We compare the  $L^2(\mathcal{X})$  and  $L^2_{\rho_T}$  in the context of truncated Singular value decomposition (SVD) regularization. We show by numerical examples that the space  $L^2_{\rho_T}$  leads to more accurate regularized estimators.

**Truncated SVD regularization.** The truncated SVD regularization methods (see [12] and references therein) discard the smallest singular values of  $A_n$  and solve the normal equation in the remaining eigenspace. To take into account the function spaces of learning, we present here a generalized version using generalized eigenvalues of  $A_n$ . More precisely, let  $B_n(i, j) = \langle \phi_i, \phi_j \rangle$  be a basis matrix, e.g.,  $B_n = B_n^G$  or  $B_n^R$ , which are extensions of  $B_n = I_n$  in [12]. Write

$$A_n = \sum_{i=1}^n \sigma_i \mathbf{u}_i \mathbf{u}_i^\top,$$

27 where  $\{\sigma_i\}$  are the decreasingly-ordered generalized eigenvalues of  $(A_n, B_n)$  and  $\{\mathbf{u}_i\} \in \mathbb{R}^n$  are the corre-  
 1 sponding  $B_n$ -orthonormal eigenvectors (i.e.,  $\mathbf{u}_i^\top B_n \mathbf{u}_j = \delta_{ij}$ ). The truncated SVD regularizer keeps only the  
 2  $n_*$  largest eigenvalues above a proper threshold and leads to an estimator

$$\hat{\phi} = \sum_{j=1}^n \hat{c}_j \phi_j, \quad \text{where } \hat{c} = \sum_{i=1}^{n_*} \frac{\mathbf{u}_i^\top b}{\sigma_i} \mathbf{u}_i. \quad (5.4)$$

3 This regularized estimator removes the error-prone contributions from  $\frac{\mathbf{u}_i^\top b}{\sigma_i} \mathbf{u}_i$  when  $\sigma_i$  is small. Also, the  
 4 estimator is regularized by expressing it as a linear combination of eigenfunctions corresponding to large ei-  
 5 genvalues, which have a resemblance to low-frequency trigonometric functions.

6 **Truncated SVD estimators in  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$ .** To apply the truncated SVD regularization, we first compute  
 7 the eigenvalues and eigenvectors corresponding to  $L^2_{\bar{\rho}_T}$  and  $L^2(\mathcal{X})$ . As suggested by Theorem 5.1, they are  
 8 from the generalized eigenvalue problems:

$$\begin{aligned} \text{Unweighted SVD } L^2(\mathcal{X}) : \quad & \Phi D_G \Phi^\top = A_n, \text{ with } \Phi^\top B_n^G \Phi = I, \\ \text{Weighted SVD } L^2_{\bar{\rho}_T} : \quad & \Psi D_R \Psi^\top = A_n, \text{ with } \Psi^\top B_n^R \Psi = I. \end{aligned} \quad (5.5)$$

9 Here  $D_G$  and  $D_R$  are diagonal matrices consisting of the generalized eigenvalues of  $(A_n, B_n^G)$  and  $(A_n, B_n^R)$ .  
 10 We compare the truncated SVD estimators in  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$  for three examples:

- 11 • cubic potential with  $\phi(r) = 3r^2$ ;
- 12 • opinion dynamics with  $\phi$  being piecewise linear;
- 13 • the attraction-repulsion potential with  $\phi(r) = r - r^{-1.5}$ .

14 These examples are studied in [17], and our numerical settings and simulations are the same as those in [17,  
 15 Section 4.1] (except for simplicity we consider  $v = 0.001$ ,  $\Delta t = 0.01$  with  $T = 0.1$  and only  $M = 400$ ).

16 We consider the basis functions being piecewise constants on a uniform partition of  $\mathcal{X}$ . One can obtain better  
 17 results by using spline basis functions with higher order regularity [17]. Here the piecewise constant basis  
 18 functions can highlight the regularity of the eigenfunctions in  $L^2_{\bar{\rho}_T}$  with  $L^2(\mathcal{X})$ , making it easier to compare  
 19  $L^2_{\bar{\rho}_T}$  with  $L^2(\mathcal{X})$  in the truncated SVD regularization. The basis matrices become

$$B_n^G = \Delta r I_n, \quad B_n^R = \text{diag}(\hat{\rho}_T(r_1), \dots, \hat{\rho}_T(r_n)) \Delta r, \quad (5.6)$$

20 where  $\{\hat{\rho}_T(r_i)\}$  is the average density on the interval  $[r_{i-1}, r_i]$ , i.e.  $\hat{\rho}_T(r_i) = \frac{1}{\Delta r} \int_{r_{i-1}}^{r_i} \bar{\rho}_T(r) dr$ . Note that we  
 21 can represent  $\bar{\rho}_T$  by  $\hat{\rho}_T = \sum_{i=1}^n \hat{\rho}_T(r_i) \phi_i$ .

22 Figure 1 shows the regularized estimators via truncated SVD in  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$  for these examples. We  
 23 pick a truncation level such that the sum of the largest  $n_*$  singular values takes 99% of the total summation of  
 24 the singular values. The corresponding eigenfunctions are presented in Figure 2. As can be seen, the weighted  
 25 SVD leads to significantly more accurate estimators than the unweighted SVD.

26 **SVD analysis in  $L^2(\mathcal{X})$  and  $L^2_{\bar{\rho}_T}$ .** SVD analysis helps to understand the truncated SVD regularization. The  
 27 truncated SVD regularization aims to remove the error-prone terms  $\frac{\mathbf{u}_i^\top b}{\sigma_i} \mathbf{u}_i$ , particularly when the eigenvalue  
 28  $\sigma_i$  is small. Thus, it is helpful to analyze the Picard ratio  $\frac{\mathbf{u}_i^\top b}{\sigma_i}$  [12]. Clearly, when the ratio converges to zero,  
 29 (called the discrete Picard condition), the term  $\frac{\mathbf{u}_i^\top b}{\sigma_i} \mathbf{u}_i$  is error-immune; when the ratio increases largely, the  
 30 inverse problem is ill-posed.

31 Figure 3 shows the singular values and the Picard ratios for the weighted and unweighted SVD (5.5). Here  
 32 the unweighted SVD is denoted by  $G$ , and ‘‘G: b-projection’’ refers to  $|\mathbf{u}_i^\top b|$  with  $\mathbf{u}_i$  being the columns of  $\Phi$ .  
 33 Similarly,  $R$  denote the weighted SVD, and ‘‘R: b-projection’’ refers to  $|\mathbf{u}_i^\top b|$  with  $\mathbf{u}_i$  being the columns of  
 34  $\Psi$ . In all these examples, the weighted SVD has larger eigenvalues than those of the unweighted SVD; and it

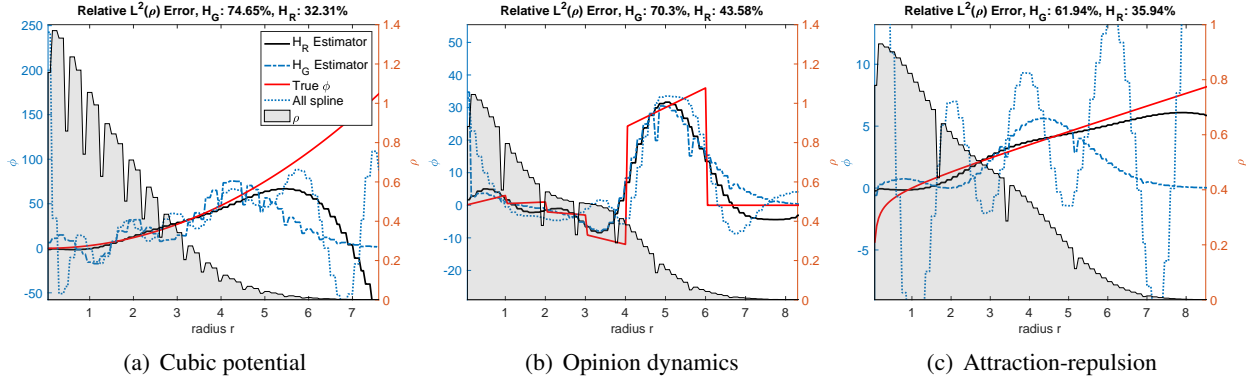


FIGURE 1. Regularized estimators via truncated SVD for the three examples, superimposed with the exploration measure  $\bar{\rho}_T$ . The weighted SVD (“ $H_R$  Estimator”) has smaller errors than the unweighted SVD (“ $H_G$  Estimator”), while both are significantly more accurate than the un-regularized estimator (“all spline”).

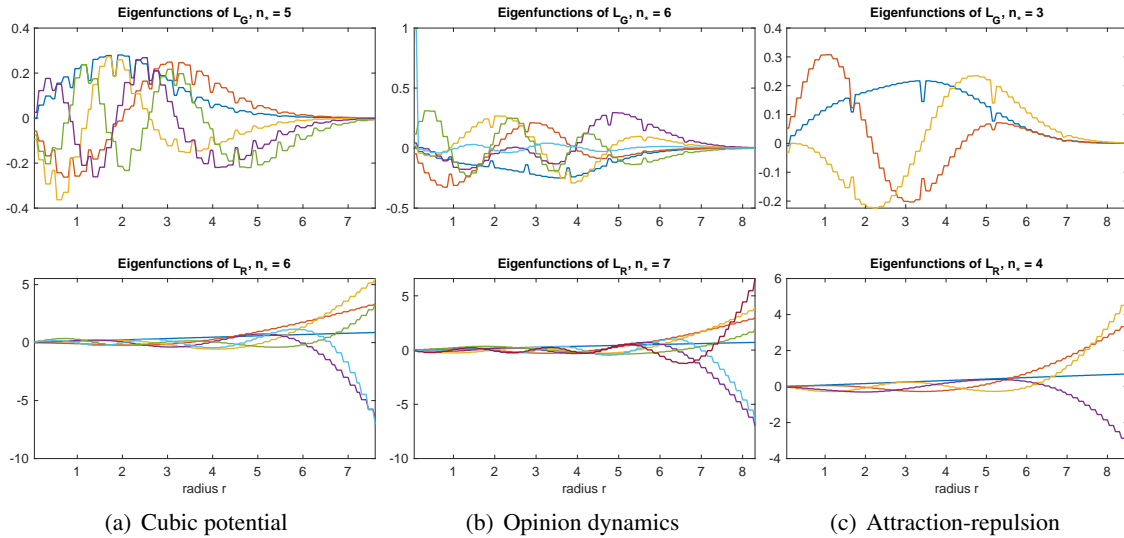


FIGURE 2. Eigenfunctions in the estimation via weighted and unweighted SVD in Figure 1. The weighted operator  $\mathcal{L}_R$  has smoother eigenfunctions than the unweighted operator  $\mathcal{L}_G$ .

35 has smaller Picard ratios. Thus, the weighted SVD leads to less ill-conditioned inversions and more accurate  
1 estimators.

2 **Remark 5.4** (Tikhonov regularization with L-curve). *The widely-used Tikhonov regularization [12] works well*  
3 *for this ill-posed inverse problem [17]. It minimizes  $\mathcal{E}_\lambda(\psi) = \mathcal{E}(\psi) + \lambda \|\psi\|^2$ , where  $\|\cdot\|$  is a regularization*  
4 *norm. When the norm  $\|\cdot\|$  defines an inner product, it leads to a basis matrix  $B_n$  for the basis functions  $\{\phi_i\}_{i=1}^n$ .*  
5 *The minimizer of  $\mathcal{E}_\lambda$  in the hypothesis space  $\mathcal{H}$  is  $\hat{c}_\lambda = (A_n + \lambda B_n)^{-1} b_n$ . There are two factors in the method,*  
6 *the regularization norm and the hyper-parameter. Given a regularization norm  $\|\cdot\|$ , an optimal parameter  $\lambda$*   
7 *aims to balance the decrement of the loss functional  $\mathcal{E}$  and the increment of the norm. Two groups of methods*  
8 *are successful. The L-curve method [13] selects  $\lambda$  at where the largest curvature occurs on a parametric curve*  
9 *of  $(\log(\mathcal{E}(\hat{c}_\lambda)), \log(\|\hat{c}_\lambda\|))$ . The truncated SVD methods with the SVD analysis can also be used to select*  
10  *$\lambda$ . However, the choice of regularization norm is problem-dependent and various norms have been explored,*  
11 *including the  $H^1$ -norm in [17] and the RKHS norms (see [20] and the references therein).*

## 12 Appendix A. Appendix.

13 **A.1. Review of RKHS and positive definite functions. Positive definite functions.** We review the definitions  
14 and properties of positive definite kernels. The following is a real-variable version of the definition in [3, p.67].



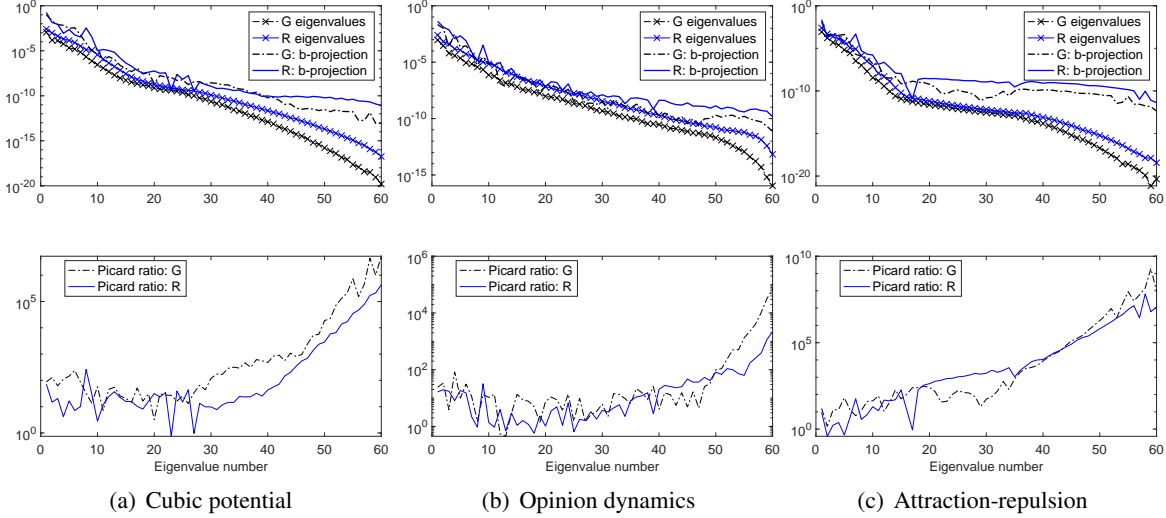


FIGURE 3. SVD analysis of the regression in three examples. Here  $R$  represents the weighted SVD and  $G$  represents the unweighted SVD. In all three examples, the weighted SVD has larger eigenvalues than those of the unweighted SVD; and it has slightly smaller ratios  $\frac{\mathbf{u}_i^\top \mathbf{b}}{\sigma_i}$ .

15 **Definition A.1** (Positive definite function). *Let  $X$  be a nonempty set. A function  $G : X \times X \rightarrow \mathbb{R}$  is positive*  
 1 *definite if and only if it is symmetric (i.e.  $G(x, y) = G(y, x)$ ) and  $\sum_{j,k=1}^n c_j c_k G(x_j, x_k) \geq 0$  for all  $n \in \mathbb{N}$ ,*  
 2  *$\{x_1, \dots, x_n\} \subset X$  and  $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ . The function  $\phi$  is strictly positive definite if the equality holds*  
 3 *only when  $\mathbf{c} = \mathbf{0} \in \mathbb{R}^n$ .*

4 **Theorem A.2** (Properties of positive definite kernels). *The following statements hold true.*

- 5 (a) *Suppose that  $k_1, k_2 : X \times X \subset \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  are positive definite kernels. Then, the product  $k_1 k_2$  is*  
 6 *positive definite ([3, p.69]).*  
 7 (b) *Inner product  $\langle u, v \rangle = \sum_{j=1}^d u_j v_j$  is positive definite ([3, p.73]).*  
 8 (c)  *$f(u)f(v)$  is positive definite for any function  $f : X \rightarrow \mathbb{R}$  ([3, p.69]).*

### 9 RKHS and positive integral operators.

10 We review the definitions and properties of Mercer kernel, RKHS, and related integral operators on compact  
 11 domains (see e.g., [8]) and non-compact domains (see e.g., [29]).

12 Let  $(X, d)$  be a metric space and  $G : X \times X \rightarrow \mathbb{R}$  be continuous and symmetric. We say that  $G$  is a  
 13 Mercer kernel if it is positive definite (as in Definition A.1). The reproducing kernel Hilbert space (RKHS)  $\mathcal{H}_G$   
 14 associated with  $G$  is defined to be closure of  $\text{span}\{G(x, \cdot) : x \in X\}$  with the inner product

$$\langle f, g \rangle_{\mathcal{H}_G} = \sum_{i=1, j=1}^{n, m} c_i d_j G(x_i, y_j)$$

15 for any  $f = \sum_{i=1}^n c_i G(x_i, \cdot)$  and  $g = \sum_{j=1}^m d_j G(y_j, \cdot)$ . It is the unique Hilbert space such that  $\text{span}\{G(\cdot, y), y \in$   
 16  $X\}$  is dense in  $\mathcal{H}_G$  and having reproducing kernel property in the sense that for all  $f \in \mathcal{H}_G$  and  $x \in X$ ,  
 17  $f(x) = \langle G(x, \cdot), f \rangle_{\mathcal{H}_G}$  (see [8, Theorem 2.9]).

18 By means of the Mercer Theorem, we can characterize the RKHS  $\mathcal{H}_G$  through the integral operator associ-  
 19 ated with the kernel. Let  $\mu$  be a non-degenerate Borel measure on  $(X, d)$  (that is,  $\mu(U) > 0$  for every open set  
 20  $U \subset X$ ). Define the integral operator  $\mathcal{L}_G$  on  $L^2(X, \mu)$  by

$$\mathcal{L}_G f(x) = \int_X G(x, y) f(y) d\mu(y).$$

21 The RKHS has the operator characterization (see e.g., [8, Section 4.4] and [29]).

22 **Theorem A.3** (Operator characterization of RKHS). *Assume that the  $G$  is a Mercer kernel and  $G \in L^2(X \times$   
 23  $X, \mu \otimes \mu)$ . Then*

- 24 1.  $\mathcal{L}_G$  is a compact positive self-adjoint operator. It has countably many positive eigenvalues  $\{\lambda_i\}_{i=1}^\infty$  and  
 1 corresponding orthonormal eigenfunctions  $\{\phi_i\}_{i=1}^\infty$ .  
 2 2.  $\{\sqrt{\lambda_i}\phi_i\}_{i=1}^\infty$  is an orthonormal basis of the RKHS  $\mathcal{H}_G$ .  
 3 3. The RKHS is the image of the square root of the integral operator, i.e.,  $\mathcal{H}_G = \mathcal{L}_G^{1/2}L^2(X, \mu)$ .

4 **A.2. Computation details for Example 4.4.** We provide here the computation details in evaluating convolu-  
 5 tions to obtain  $\bar{F}_T(x, y)$  and  $\bar{\rho}_T(x, y)$  in Example 4.4.

6 **1. Computation of  $\bar{F}_T(x, y)$ .** Recall that  $U(x) = \frac{1}{\pi} \frac{1}{1+x^2}$  and

$$\bar{F}_T(x, y) = \int_{\mathbb{R}} U(z-x)U(z-y)U(z)dz = \frac{1}{\pi^3} \int_{\mathbb{R}} \frac{1}{1+(z-x)^2} \frac{1}{1+(z-y)^2} \frac{1}{1+z^2} dz.$$

7 Using a separation of rational functions, we can write

$$\frac{1}{1+(z-x)^2} \frac{1}{1+(z-y)^2} \frac{1}{1+z^2} = \frac{Az+B}{1+(z-x)^2} + \frac{Cz+D}{1+(z-y)^2} + \frac{Ez+F}{1+z^2}, \quad (\text{A.1})$$

8 where each term can be integrated analytically. We first solve for the constants  $A, B, C, D, E$  and  $F$  from a  
 9 system of linear equations that match the coefficients of the powers of  $z^p$  with  $p \in \{0, 1, \dots, 5\}$ . For example,  
 10 we have  $A + C + E = 0$  from the coefficient of  $z^5$ . Using a symbolic numerical solver, we obtain

$$\begin{aligned} A &= \frac{-2(2x-y)}{x(x^2+4)((x-y)^3+4(x-y))}, & B &= \frac{5x^2-3xy-4}{x(x^2+4)((x-y)^3+4(x-y))}, \\ C &= \frac{2(x-2y)}{y(y^2+4)((x-y)^3+4(x-y))}, & D &= \frac{-5y^2+3xy+4}{y(y^2+4)((x-y)^3+4(x-y))}, \\ F &= \frac{xy-4}{xy(x^2+4)(y^2+4)}. \end{aligned}$$

11 Meanwhile, notice that each of the three terms in (A.1) can be computed similarly and

$$\int_{\mathbb{R}} \frac{Az+B}{1+(z-x)^2} dz = \int_{\mathbb{R}} \frac{A(z-x) + (Ax+B)}{1+(z-x)^2} dz = A \int_{\mathbb{R}} \frac{t}{1+t^2} dt + (Ax+B) \int_{\mathbb{R}} \frac{1}{1+t^2} dt. \quad (\text{A.2})$$

12 The terms with  $\int_{\mathbb{R}} \frac{t}{1+t^2} dt$  is canceled because  $A + C + E = 0$ . Also  $\int_{\mathbb{R}} \frac{1}{1+t^2} dt = \pi$ . Hence, we have

$$\bar{F}_T(x, y) = \frac{1}{\pi^2} [(Ax+B) + (Cy+D) + F] = \frac{2}{\pi^2} \frac{(x^2 - xy + y^2 + 12)}{(x^2+4)(y^2+4)(x^2 - 2xy + y^2 + 4)}.$$

13 **2. Computation of  $\bar{\rho}_T(x, y)$ .** The computation is similar to that of  $\bar{F}_T(x, y)$ . Notice that,

$$\bar{\rho}_T(x) = (U(\cdot) * U(-\cdot))(x) = \frac{1}{\pi^2} \int_{\mathbb{R}} \frac{1}{1+(z-x)^2} \frac{1}{1+z^2} dz.$$

14 Thus, using a separation of rational functions  $\frac{1}{1+(z-x)^2} \frac{1}{1+z^2} = \frac{Az+B}{1+(z-x)^2} \frac{Cz+D}{1+z^2}$ , Eq.(A.2), and the fact that  
 15  $A + C = 0$  from the coefficient of  $z^3$ , we have

$$\bar{\rho}_T(x) = \frac{1}{\pi} (Ax + B + D) = \frac{2}{\pi(x^2+4)}.$$

16 **Acknowledgments.** FL would like to thank Mauro Maggioni and P-E Jabin for helpful discussions on mean-  
 17 field equations and the inverse problem.

18

## REFERENCES

- 19 [1] A. S. Baumgarten and K. Kamrin. A general constitutive model for dense, fine-particle suspensions validated in many geometries.  
 20 *Proc Natl Acad Sci USA*, 116(42):20828–20836, 2019.  
 21 [2] N. Bell, Y. Yu, and P. J. Mucha. Particle-based simulation of granular materials. In *Proceedings of the 2005 ACM SIG-*  
 22 *GRAPH/Eurographics Symposium on Computer Animation - SCA '05*, page 77, Los Angeles, California, 2005. ACM Press.  
 23 [3] C. Berg, J. P. R. Christensen, and P. Ressel. *Harmonic analysis on semigroups: theory of positive definite and related functions*,  
 24 volume 100. New York: Springer, 1984.

- 25 [4] M. Bongini, M. Fornasier, M. Hansen, and M. Maggioni. Inferring interaction rules from observations of evolutive systems I: The  
 1 variational approach. *Mathematical Models and Methods in Applied Sciences*, 27(05):909–951, 2017.
- 2 [5] J. A. Carrillo, K. Craig, and Y. Yao. Aggregation-diffusion equations: dynamics, asymptotics, and singular limits. In *Active*  
 3 *Particles, Volume 2*, pages 65–108. Springer, 2019.
- 4 [6] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepčev. Global-in-time weak measure solutions and finite-time  
 5 aggregation for nonlocal interaction equations. *Duke Math. J.*, 156(2):229–271, 2011.
- 6 [7] X. Chen. Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data.  
 7 *ArXiv200711048 Math Stat*, 2021.
- 8 [8] F. Cucker and D. X. Zhou. *Learning theory: an approximation theory viewpoint*, volume 24. Cambridge University Press, 2007.
- 9 [9] L. Della Maestra and M. Hoffmann. The LAN property for McKean-Vlasov models in a mean-field regime, 2022.
- 10 [10] L. Della Maestra and M. Hoffmann. Nonparametric estimation for interacting particle systems: McKean–Vlasov models. *Proba-*  
 11 *bility Theory and Related Fields*, 182(1):551–613, 2022.
- 12 [11] J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York, NY, 2003.
- 13 [12] P. C. Hansen. REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems. *Numer*  
 14 *Algor*, 6(1):1–35, 1994.
- 15 [13] P. C. Hansen. The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in*  
 16 *Electrocardiology*, ed. P. Johnston, *Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.
- 17 [14] P.-E. Jabin and Z. Wang. Mean Field Limit for Stochastic Particle Systems. In N. Bellomo, P. Degond, and E. Tadmor, editors,  
 18 *Active Particles, Volume 1*, pages 379–402. Springer International Publishing, Cham, 2017.
- 19 [15] P.-E. Jabin and Z. Wang. Quantitative estimates of propagation of chaos for stochastic systems with  $w^{-1,\infty}$  kernels. *Invent. math.*,  
 20 214(1):523–591, 2018.
- 21 [16] R. A. Kasonga. Maximum Likelihood Theory for Large Interacting Systems. *SIAM J. Appl. Math.*, 50(3):865–875, 1990.
- 22 [17] Q. Lang and F. Lu. Learning interaction kernels in mean-field equations of first-order systems of interacting particles. *SIAM*  
 23 *Journal on Scientific Computing*, 44(1):A260–A285, 2022.
- 24 [18] Z. Li and F. Lu. On the coercivity condition in the learning of interacting particle systems. *arXiv preprint arXiv:2011.10480*, 2020.
- 25 [19] Z. Li, F. Lu, M. Maggioni, S. Tang, and C. Zhang. On the identifiability of interaction functions in systems of interacting particles.  
 26 *Stochastic Processes and their Applications*, 132:135–163, 2021.
- 27 [20] F. Lu, Q. Lang, and Q. An. Data adaptive RKHS Tikhonov regularization for learning kernels in operators. *Proceedings of Math-*  
 28 *ematical and Scientific Machine Learning, PMLR 190:158-172*, 2022.
- 29 [21] F. Lu, M. Maggioni, and S. Tang. Learning interaction kernels in heterogeneous systems of agents from multiple trajectories.  
 30 *Journal of Machine Learning Research*, 22(32):1–67, 2021.
- 31 [22] F. Lu, M. Maggioni, and S. Tang. Learning interaction kernels in stochastic systems of interacting particles from multiple trajec-  
 32 tories. *Foundations of Computational Mathematics*, pages 1–55, 2021.
- 33 [23] F. Lu, M. Zhong, S. Tang, and M. Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory  
 34 data. *Proceedings of the National Academy of Sciences of the United States of America*, 116(29):14424–14433, 2019.
- 35 [24] F. Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560,  
 36 2003.
- 37 [25] S. Méléard. *Asymptotic Behaviour of Some Interacting Particle Systems; McKean-Vlasov and Boltzmann Models*, volume 1627,  
 38 pages 42–95. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996.
- 39 [26] S. Mostch and E. Tadmor. Heterophilious Dynamics Enhances Consensus. *Siam Review*, 56(4):577 – 621, 2014.
- 40 [27] C. E. Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.
- 41 [28] L. Sharrock, N. Kantas, P. Pappas, and G. A. Pavliotis. Parameter estimation for the McKean-Vlasov stochastic differential equa-  
 42 tion. *arXiv preprint arXiv:2106.13751*, 2021.
- 43 [29] H. Sun. Mercer theorem for rkhs on noncompact sets. *Journal of Complexity*, 21(3):337 – 349, 2005.
- 44 [30] A.-S. Sznitman. *Topics in Propagation of Chaos*, volume 1464, pages 165–251. Springer Berlin Heidelberg, Berlin, Heidelberg,  
 45 1991.
- 46 [31] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517:71 – 140, 2012.
- 47 [32] R. Yao, X. Chen, and Y. Yang. Mean-field nonparametric estimation of interacting particle systems. *arXiv preprint*  
 48 *arXiv:2205.07937*, 2022.

49 E-mail address: [qlang1@math.jhu.edu](mailto:qlang1@math.jhu.edu)

50 E-mail address: [feilu@math.jhu.edu](mailto:feilu@math.jhu.edu)