Learning kernels in mean-field equations of interacting particle systems

Fei Lu

Department of Mathematics, Johns Hopkins University

Joint with **Quanjun Lang, Qingci An**

March 22, 2022; UMass. Amhurst

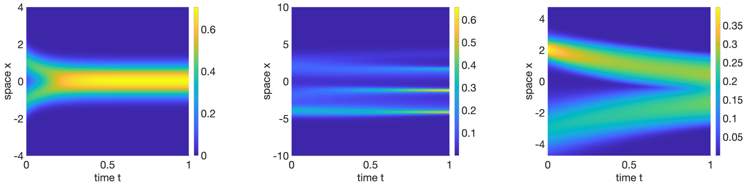
An inverse problem

Mean-field PDE

$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)], \quad x \in \mathbb{R}^d, t > 0,$$

where $K_\phi(x) = \nabla(\Phi(|x|)) = \phi(|x|)\frac{x}{|x|}$.

Question: identify ϕ from discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$?



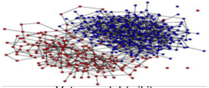
Dataset 1 Dataset 2 Dataset 3

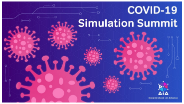
Systems of interacting particles/agents

$$\frac{d}{dt}X_t^i = \frac{1}{N}\sum_{i'=1}^{N} \phi(|X_t^j - X_t^i|)\frac{X_t^j - X_t^i}{|X_t^j - X_t^i|} + \sqrt{2\nu}\,dB_t^i, \quad i = 1, \ldots, N$$

- X_t^i: the i-th particle's position; B_t^i: Brownian motion
- Applications in many disciplines:
  - Statistical physics, quantum mechanics
  - Social science [Motsch-Tadmor2014]
  - Agent-based models (for COVID19)
  - Biology [Keller-Segal1970, Cucker-Smale2000]
  - Monte Carlo sampling [Del Moral13]


Popkin. Nature(2016)


Voter model (wiki)



- Problem: discover interaction law (ϕ) from data?

$$\frac{d}{dt}X_t^i = \frac{1}{N}\sum_{i'=1}^{N} \phi(|X_t^j - X_t^i|)\frac{X_t^j - X_t^i}{|X_t^j - X_t^i|} + \sqrt{2\nu}dB_t^i$$

Finite N: [Maggioni, L., Tang, Zhong, Miller, Li, Zhang: PNAS19, SPA20, et al]

- Data: many trajectories of particles
- Nonparametric learning:
  - Identifiability: a coercivity condition
  - Convergence rate = minimax rate 1D
  - ODE/SDEs: Opinion Dynamics, Lennard-Jones, Prey-Predator; 1st/2nd order

Learning interaction kernel

$$\frac{d}{dt}X_t^i = \frac{1}{N}\sum_{i'=1}^{N} \phi(|X_t^j - X_t^i|)\frac{X_t^j - X_t^i}{|X_t^j - X_t^i|} + \sqrt{2\nu}dB_t^i$$

Finite N: [Maggioni, L., Tang, Zhong, Miller, Li, Zhang: PNAS19, SPA20, et al]

- Data: many trajectories of particles
- Nonparametric learning:
  - Identifiability: a coercivity condition
  - Convergence rate = minimax rate 1D
  - ODE/SDEs: Opinion Dynamics, Lennard-Jones, Prey-Predator; 1st/2nd order

Large N ($>>$ 1) [Lang-Lu 20,21]

- Data: population density u

$$\partial_t u = \nu\Delta u + \nabla \cdot [u(K_\phi * u)]$$

- Inverse problem for mean-field PDE

Outline

1. Motivation and problem statement
2. Nonparametric learning
   - A probabilistic loss functional
   - Identifiability: function spaces of learning
   - Rate of convergence
3. Numerical examples
4. DARTR: regularization for linear inverse problems

Loss functional

$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)]$$

Candidates:

- Discrepancy: $\mathcal{E}(\psi) = \|\partial_t u - \nu \Delta u - \nabla.(u(K_\psi * u))\|^2$
- Free energy: $\mathcal{E}(\psi) = C + |\int_{\mathbb{R}^d} u[(\Psi - \Phi) * u]dx|^2$
- Wasserstein-2: $\mathcal{E}(\psi) = W_2(u^\psi, u)$
  costly: requires many PDE simulations in optimization
- A probabilistic loss functional

A probabilistic loss functional

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

- $= -\mathbb{E}[$ log-likelihood $]$ of the process

$$\begin{cases} d\overline{X}_t = - K_{\phi_{true}} * u(\overline{X}_t, t)dt + \sqrt{2\nu} dB_t, \\ \mathcal{L}(\overline{X}_t) = u(\cdot, t), \end{cases}$$

- Derivative free
- Suitable for high dimension

$$K_\psi * u(\overline{X}_t) = \mathbb{E}[K_\psi(\overline{X}_t - \overline{X}'_t)|\overline{X}_t]$$

Least squares estimator

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

$$= \langle\langle \psi, \psi \rangle\rangle - 2 \langle\langle \psi, \phi \rangle\rangle, \qquad K_\psi(x) = \psi(|x|)\frac{x}{|x|}$$

- bilinear form $\langle\langle \phi, \psi \rangle\rangle = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \langle (K_\phi * u), (K_\psi * u) \rangle u(x,t) dxdt$
- Hypothesis space $\mathcal{H}_n = \mathrm{span}\{\phi_i\}_{i=1}^n$: $\psi = \sum_{i=1}^n c_i \phi_i$

$$\Rightarrow \quad \mathcal{E}(\psi) = c^\top A c - 2b^\top c \text{ with } A_{ij} = \langle\langle \phi_i, \phi_j \rangle\rangle$$

$\Rightarrow$ **Estimator:** $\widehat{\phi}_n = \sum_{i=1}^n \widehat{c}_i \phi_i, \qquad \widehat{c} = A^{-1} b$

Three fundamental issues

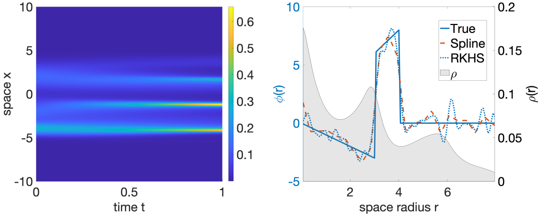
1. Identifiability, $A^{-1}b$, well-posedness: function space of learning
2. Choice of $\mathcal{H}_n$: $\{\phi_i\}_{i=1}^n$ and n ?
3. Convergence rate when $\Delta x = M^{-1/d} \to 0$?

Identifiability and function space of learning

$$A_{ij} = \langle\langle \phi_i, \phi_j \rangle\rangle = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \langle (K_\phi * u), (K_\psi * u)\rangle u(x,t)dxdt$$
$$= \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \phi_i(r)\phi_j(s)\overline{G}_T(r,s)\rho_T(dr)\rho_T(ds)$$

$$K_\psi * u(\overline{X}_t) = \mathbb{E}[K_\psi(\overline{X}_t - \overline{X}'_t)|\overline{X}_t]$$

- Exploration measure $\rho_T \leftarrow |\overline{X}_t - \overline{X}'_t|$



Identifiability and function space of learning

$$A_{ij} = \langle\langle \phi_i, \phi_j \rangle\rangle = \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \phi_i(r)\psi_j(s)\overline{G}_T(r,s)\rho_T(dr)\rho_T(ds)$$
$$= \langle L_{\overline{G}_T}\phi_i, \phi_j \rangle_{L^2(\rho_T)}$$

- Exploration measure $\rho_T \leftarrow |\overline{X}_t - \overline{X}'_t|$
- Positive compact operator $L_{\overline{G}_T}$
  - normal matrix $A \sim L_{\overline{G}_T}|_{\mathcal{H}}$ in $L^2(\rho_T)$

$$c_{\mathcal{H},T} = \inf_{\psi \in \mathcal{H}, \|\psi\|_{L^2(\rho_T)}=1} \langle\langle \psi, \psi \rangle\rangle > 0 \quad \text{(Coercivity condition)}$$

Identifiability and function space of learning

$$A_{ij} = \langle\!\langle \phi_i, \phi_j \rangle\!\rangle = \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \phi_i(r)\psi_j(s)\overline{G}_T(r,s)\rho_T(dr)\rho_T(ds)$$
$$= \langle L_{\overline{G}_T}\phi_i, \phi_j\rangle_{L^2(\rho_T)}$$

- Exploration measure $\rho_T \leftarrow |\overline{X}_t - \overline{X}_t'|$
- Positive compact operator $L_{\overline{G}_T}$
  - normal matrix $A \sim L_{\overline{G}_T}|_{\mathcal{H}}$ in $L^2(\rho_T)$

$$c_{\mathcal{H},T} = \inf_{\psi\in\mathcal{H}, \|\psi\|_{L^2(\rho_T)}=1} \langle\!\langle \psi, \psi \rangle\!\rangle > 0 \quad \text{(Coercivity condition)}$$

- Identifiability: $A^{-1}b \leftrightarrow L_{\overline{G}_T}^{-1}\phi^D$
  - RKHS $H_{\overline{G}} \subset L^2(\rho_T)$ [LangLu21]
  - DARTR: Data Adaptive RKHS Tikhonov Regularization

Error bound

$$\mathbb{H} = L^2(\rho_T)$$

Theorem (Numerical error bound [Lang-Lu20])

Let $\mathcal{H} = \mathrm{span}\{\phi_i\}_{i=1}^n$ *and* $\widehat{\phi}_n$ *the projection of* ϕ *on* $\mathcal{H} \subset \mathbb{H}$. *Assume regularity + coercivity conditions . Then*

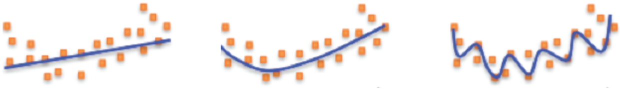
$$\|\widehat{\phi}_{n,M,L} - \widehat{\phi}_n\|_{\mathbb{H}} \leq 2c_{\mathcal{H},T}{}^{-1}\left(C^b\sqrt{n} + C^A n\,\|\phi\|_{\mathbb{H}}\right)(\Delta x^\alpha + \Delta t).$$

- from discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$
- Δx^α comes from numerical integrator (e.g.,Riemann sum)
- Dominating order: $n\Delta x^\alpha$

Optimal dimension and rate of convergence

Total error: trade-off

$$\|\widehat{\phi}_{n,M,\infty} - \phi\|_{\mathbb{H}} \leq \underbrace{\|\widehat{\phi}_{n,M,\infty} - \widehat{\phi}_n\|_{\mathbb{H}}}_{\text{numerical error}} + \underbrace{\|\widehat{\phi}_n - \phi\|_{\mathbb{H}}}_{\text{approximation error}}$$



Underfitting **Balanced** **Overfitting**

Theorem (Rate of convergence [Lang-Lu20])

Assume $\|\widehat{\phi}_{n,M,\infty} - \widehat{\phi}_n\|_{\mathbb{H}} \lessapprox n(\Delta x)^{\alpha}$ *and* $\|\widehat{\phi}_n - \phi\|_{\mathbb{H}} \lessapprox n^{-s}$. *Then, with dimension* $n \approx (\Delta x)^{-\alpha/(s+1)}$, *we have a rate:*

$$\|\widehat{\phi}_{n,M,\infty} - \phi\|_{\mathbb{H}} \lessapprox (\Delta x)^{\alpha s/(s+1)}$$

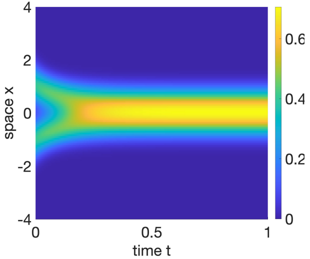
Outline

1. Motivation and problem statement
2. Nonparametric learning
   - A probabilistic loss functional
   - Identifiability
   - Rate of convergence
3. Numerical examples
   - Granular media: smooth kernel $\phi(r) = 3r^2$
   - Opinion dynamics: piecewise linear ϕ
   - Repulsion-attraction: singular $\phi = r - r^{-1.5}$
4. DARTR: regularization for linear inverse problems

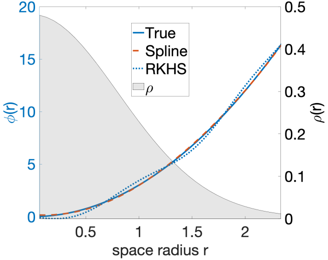
Example 1: granular media

$$\phi(r) = 3r^2$$

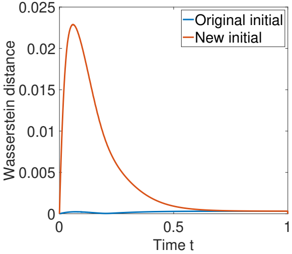
$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)], \quad x \in \mathbb{R}^d, t > 0, \quad K_\phi(x) = \phi(|x|)\frac{x}{|x|}$$



The solution $u(x,t)$

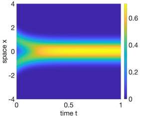


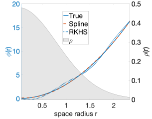
Estimators of ϕ

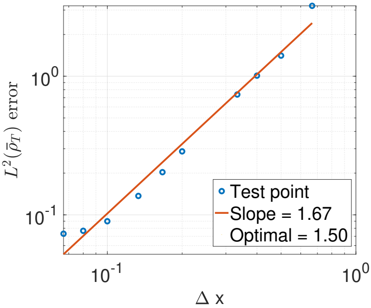


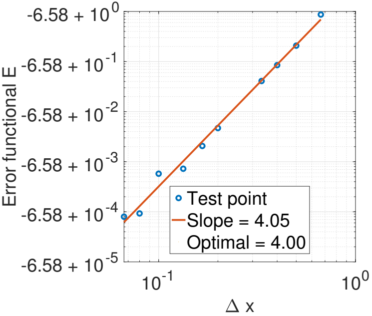
Wasserstein $W_2(u, \widehat{u})$

Example 1: granular media







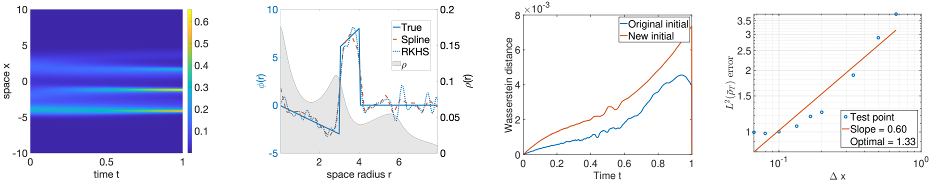


Convergence rate of $L^2(\rho_T)$ error | Convergence rate of $\mathcal{E}_{M,L}$

close to optimal

Example 2: Opinion dynamics

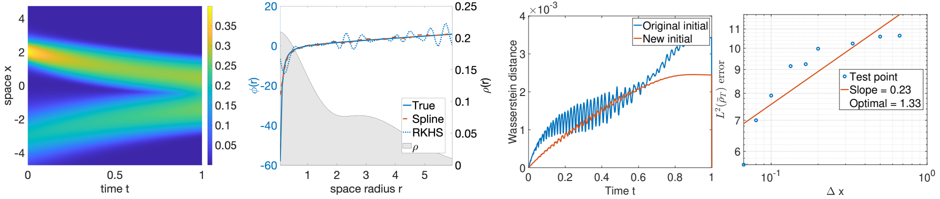
$\phi(r)$ piecewise linear



- Acceptable estimator
- Accurate prediction: small Wasserstein-2
- sub-optimal rate ($\phi \notin W^{1,\infty}$)

Example 3: repulsion-attraction

$\phi(r) = r - r^{-1.5}$ (singular)



- Acceptable estimator
- Accurate prediction: small Wasserstein-2
- low rate: theory does not apply

Outline

1. Motivation and problem statement
2. Nonparametric learning
   - A probabilistic loss functional
   - Identifiability: RKHS
   - Rate of convergence
3. Numerical examples
4. DARTR: regularization for linear inverse problems

Learning kernels in operators

Learn the kernel ϕ: $$R_\phi[u] = f$$

from data: $$\mathcal{D} = \{(u_k, f_k)\}_{k=1}^N, \quad (u_k, f_k) \in \mathbb{X} \times \mathbb{Y}$$

- R_ϕ linear in ϕ, but linear/nonlinear in u:

$$R_\phi[u] = \nabla \cdot [u(K_\phi * u)] = \partial_t u - \nu \Delta u$$

- Other operators: integral operators, nonlocal operators,...
- General linear inverse problems:
  - deconvolution
  - homogenization
  - inverse Laplace transform

DARTR: Data Adaptive RKHS Tikhonov Regularization

Quadratic loss functional:

$$\mathcal{E}(\psi) = \|R_\psi[u] - f\|_{\mathbb{Y}}^2 = \langle L_G\psi, \psi\rangle_{L^2(\rho)} - 2\langle \phi^f, \psi\rangle_{L^2(\rho)}$$

$$\nabla\mathcal{E}(\psi) = L_G\psi - \phi^f = 0 \quad \to \widehat{\phi} = L_G^{-1}\phi^f$$

Regularization:

$$\mathcal{E}_\lambda(\psi) = \mathcal{E}(\psi) + \lambda\|\psi\|_*^2 \to c^\top A c - 2b^\top c + \lambda\|c\|_*^2$$

Which norm $\|\cdot\|_*$ to use?

DARTR: Data Adaptive RKHS Tikhonov Regularization

Quadratic loss functional:

$$\mathcal{E}(\psi) = \|R_\psi[u] - f\|_{\mathbb{Y}}^2 = \langle L_G \psi, \psi \rangle_{L^2(\rho)} - 2\langle \phi^f, \psi \rangle_{L^2(\rho)}$$
$$\nabla \mathcal{E}(\psi) = L_G \psi - \phi^f = 0 \quad \rightarrow \widehat{\phi} = L_G^{-1} \phi^f$$

Regularization:

$$\mathcal{E}_\lambda(\psi) = \mathcal{E}(\psi) + \lambda \|\psi\|_*^2 \rightarrow c^\top A c - 2b^\top c + \lambda \|c\|_*^2$$

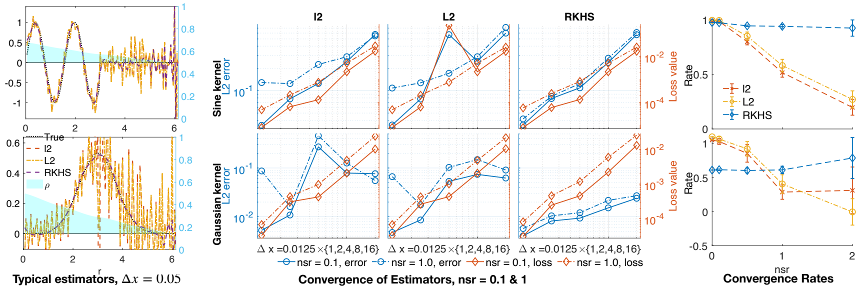
Which norm $\|\cdot\|_*$ to use?

norm of RKHS with G: $H_G \subset L^2(\rho)$ [Lu+Lang+An22]
DARTR: Data Adaptive RKHS Tikhonov Regularization

DARTR: Data Adaptive RKHS Tikhonov Regularization

$$R_\phi[u] = \nabla \cdot [u(K_\phi * u)] = f$$

- Recover kernel from discrete noisy data
- Consistent convergence as mesh refines



Summary and future directions

Problem: Estimate kernel of Mean-field equation

$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)]$$

from discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$.

Solution: Robust efficient learning algorithm by least-squares

- A probabilistic loss functional
- Identifiability, RKHS
- Convergent estimator

DARTR: regularization for general linear inverse problems

Future directions

- General systems/settings:
  - Aggression equations (inviscid MFE)
  - Learning from steady-states
  - High-D, non-radial kernels (Monte Carlo)
- DARTR:
  - DARTR for NN
  - Applications of DARTR: deconvolution, homogenization

References (@ `http://www.math.jhu.edu/~feilu`)

- Q. Lang and F. Lu. Learning interaction kernels in mean-field equations of 1st-order systems of interacting particles. SISC22
- Q. Lang and F. Lu. Identifiability of interaction kernels in mean-field equations of interacting particles. arXiv2106.
- F.Lu, Q .An and Y. Yu. Nonparametric learning of kernels in nonlocal operators. 2201

- F.Lu, Q .Lang and Q. An. Data adaptive RKHS Tikhonov regularization for learning kernels in operators. arXiv2203