Learning interacting kernels in mean-field equations of particle systems

Fei Lu

Department of Mathematics, Johns Hopkins University

Joint with **Quanjun Lang**

August 23, 2021;
Dynamics Days Europe 2021, Nice, France

Outline

An inverse problem

Mean-field equation of interacting particles

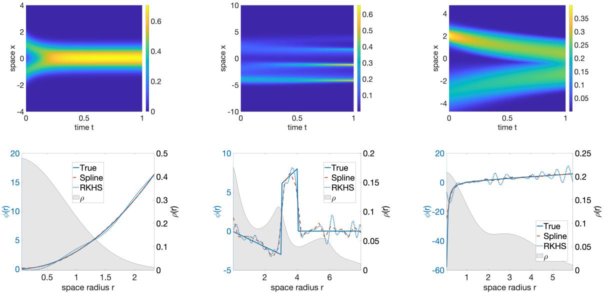
$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)], \quad x \in \mathbb{R}^d, t > 0,$$

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

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

Data

Kernel ϕ ?

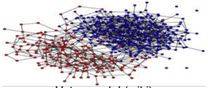


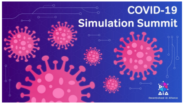
Systems of interacting particles/agents

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

- X_t^i: the i-th particle's position; B_t^i: Brownian motion
- From Newton's law (2nd-order) + 0-mass $\rightarrow$ 1st-order
- Applications in many disciplines:
  - Statistical physics, quantum mechanics
  - Biology [Keller-Segal1970, Cucker-Smale2000]
  - Social science [Motsch-Tadmor2014]
  - Monte Carlo sampling [Del Moral13]
  - Epidemiology (Agent-based models for COVID19 at Imperial)


Popkin. Nature(2016)


Voter model (wiki)



Previous work: finite N

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

- Data: many trajectories $\{X_{[0,T]}^{(m)}\}_{m=1}^{M}$.
- Nonparametric learning
  - Identifiability: $L^2(\rho_T)$ with $\rho_T \leftarrow |X_t^j - X_t^i|$
  - minimax convergence rate
  - various systems: Opinion Dynamics, Lennard-Jones, Prey-Predator

Large system challenge: $N \to \infty$

Data at macroscopic scale:

- lack of trajectories $\{X^{(m)}_{[0,T]}\}_{m=1}^{M}$ (recall $X_t \in \mathbb{R}^{Nd}$)
- only population density $u(x,t) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta(X_t^i - x)$

Discrete in space-time: $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$

$\Rightarrow$ Infer kernel ϕ in Mean-field equation:

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

Outline

1. Motivation and problem statement
2. Nonparametric learning
    - A probabilistic loss functional
    - Identifiability: function spaces of learning
    - Rate of convergence
3. Numerical examples

Variational approach: minimize a loss functional $\mathcal{E}(\psi)$

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

Loss functions that do not work

- discrepancy: $\mathcal{E}(\psi) = \|\partial_t u - \nu \Delta u - \nabla.(u(K_\psi * u))\|^2$
  derivatives approx. from discrete data
- Free energy: $\mathcal{E}(\psi) = C + |\int_{\mathbb{R}^d} u[(\Psi - \Phi) * u]dx|^2$
  limitted information from the 1st moment
- Wasserstein-2: $\mathcal{E}(\psi) = W_2(u^\psi, u)$
  costly: require many PDE simulations in optimization

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

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

A probabilistic loss functional

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

- $= -\mathbb{E}[$ 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-in-space free!
- Suitable for high-dimension: $K_\psi * u(\overline{X}_t) = \mathbb{E}[K_\psi(\overline{X}_t - \overline{X}_t') | \overline{X}_t]$

$$\mathcal{E}(\psi) = \frac{1}{T}\int_0^T \Big(\mathbb{E}|K_\psi * u(\overline{X}_t)|^2 + \partial_t \mathbb{E}\Psi(\overline{X}_t - \overline{X}_t') + 2\nu \mathbb{E}[\Delta \Psi(\overline{X}_t - \overline{X}_t')] \Big) \, dt$$

Least squares estimator

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

- 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 $\mathcal{H}_n = \mathrm{span}\{\phi_i\}_{i=1}^n$: $\psi = \sum_{i=1}^n c_i \phi_i$

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

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

- From data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$:

$$\widehat{\phi}_{n,M,L} = \sum_{i=1}^n \widehat{c}^i_{n,M,L} \phi_i, \qquad \text{with } \widehat{c}_{n,M,L} = A^{-1}_{n,M,L} b_{n,M,L}.$$

Three fundamental issues

- Identifiability: uniqueness of minimizer, A^{-1}
- Choice of $\mathcal{H}_n$: B-splines $\{\phi_i\}_{i=1}^n$ and n ?
- Convergence rate when $\Delta x = M^{-1/d} \to 0$?

Identifiability and function space of learning

Recall that $\mathcal{H} = \mathrm{span}\{\phi_i\}_{i=1}^n$,

$$A_{ij} = \langle\!\langle \phi_i, \phi_j \rangle\!\rangle\,, \quad \langle\!\langle \phi, \psi \rangle\!\rangle = \int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \phi(r)\psi(s)\overline{R}_T(r,s)\rho_T(dr)\rho_T(ds)$$

- identifiability: on RKHS $H_{\overline{R}} \subset L^2(\rho_T)$ [LangLu21]
- Positive compact operator $L_{\overline{R}_T}$
  - normal matrix $A = L_{\overline{R}_T}|_{\mathcal{H}}$ in $L^2(\rho_T)$

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

  - $H_{\overline{R}}$ based Regularization needed as $Dim(\mathcal{H}) \uparrow \infty$.

Integration Error bounds

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

Theorem (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 conditions. Then

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

- Δx^α comes from numerical integrator (Riemann sum)
- Dominating order: $n\Delta x^\alpha$

Optimal dimension and rate of convergence

Total error: trade-off

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

Theorem (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 minimax rate:*

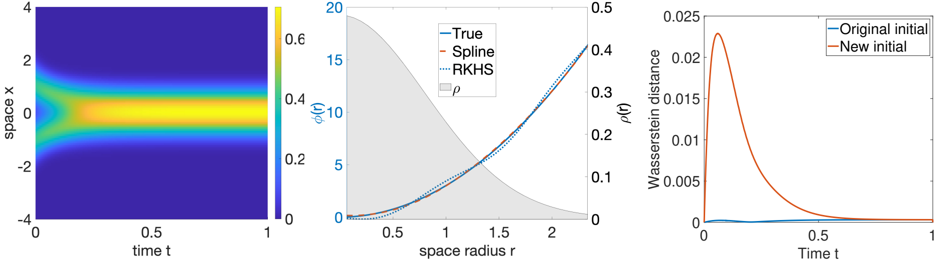
$$\|\widehat{\phi}_{n,M,\infty} - \phi\|_{\mathbb{H}} \lessapprox (\Delta x)^{\alpha s/(s+1)}$$

Outline

Example 1: granular media

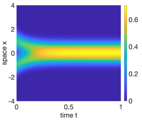
$\phi(r) = 3r^2$

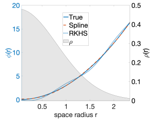
$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)], \quad x \in \mathbb{R}^d, t > 0, \quad K_\phi(x) = \phi(|x|)\frac{x}{|x|}$$

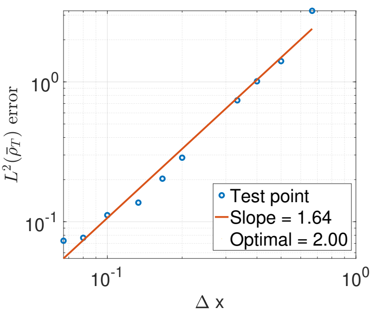


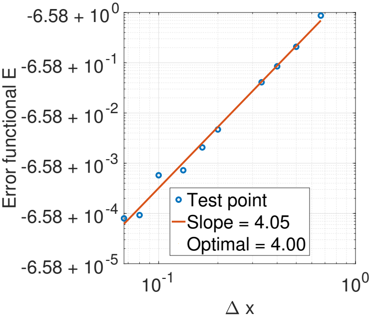
The solution $u(x,t)$ Estimators of ϕ Wasserstein $W_2(u,\widehat{u})$

Example 1: granular media







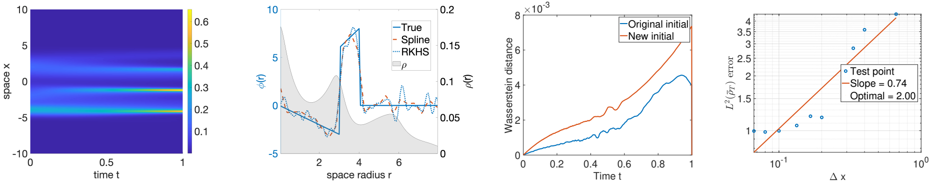


Convergence rate of $L^2(\rho_T)$ error   Convergence rate of $\mathcal{E}_{M,L}$

close to optimal

Example 2: Opinion dynamics

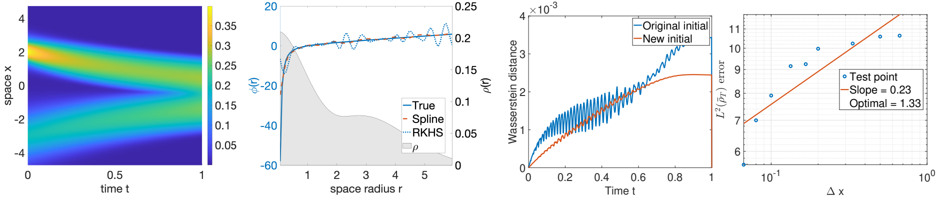
$\phi(r)$ piecewise linear



- 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

Summary and open problems

Problem: Estimate ϕ 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: Least-squares-based algorithm

- A probabilistic loss functional

Theory:

- Identifiability: intrinsic RKHSs
- Minimax convergence rate

Future directions

- General systems/settings:
  - 2nd-order systems / inviscid equations
  - High-dimensional cases (Monte Carlo)
- System-Intrinsic Data-Adaptive RKHSs:
  - Spare-aware regularization
  - feature selection

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. arXiv.2010.15694

- Q. Lang and F. Lu. Identifiability of interaction kernels in mean-field equations of interacting particles. arXiv2106.