Learning interacting kernels of mean-field equations of particle systems

Fei Lu

Department of Mathematics, Johns Hopkins University

Joint work with Quanjun Lang

Related work with: Mauro Maggioni, Sui Tang, Ming Zhong, Zhongyang Li and Cheng Zhang

October 28, 2020
Junior Colloquium, JHU

FL acknowledges supports from JHU, NSF

Outline

1. Motivation and problem statement
2. Nonparametric Learning
3. Numerical examples
4. Ongoing work and open problems

An inverse problem

Consider the mean-field equation

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

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

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

Goal:

- An algorithm $\to \widehat{\phi}$
- identifiability: function space of learning
- convergence rate when $\Delta x = M^{-1/d} \to 0$

Motivation

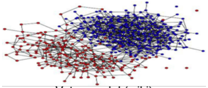
$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)]$$

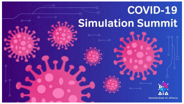
Interacting particles/agents:

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

- X_t^i: the i-th particle's position; B_t^i: Brownian motion
- $u(x,t) = \lim_{N\to\infty} \sum_{i=1}^N \delta(X_t^i - x)$ Propagation of chaos
- 1st- and 2nd-order models
- Application in many disciplines:
  - Statistical physics, quantum mechanics
  - Biology [Keller-Segal1970, Cucker-Smale2000]
  - Social science [Motsch-Tadmor2014]
  - Monte Carlo sampling [Del Moral13]
  - Epidemiology (Agent-based model for COVID19 at Imperial)


Popkin. Nature(2016)


Voter model (wiki)



Previous work: finite N

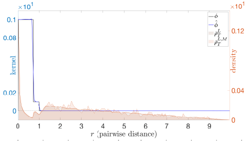
$$\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)]$$

Interacting particles/agents:

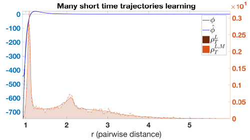
$$\frac{d}{dt} X_t^i = \frac{1}{N} \sum_{i'=1}^{N} \phi(|X_t^j - X_t^i|) \frac{X_t^j - X_t^i}{|X_t^j - X_t^i|} + \sqrt{2\nu} dB_t^i, \quad i = 1, \ldots, N$$

Maggioni JHU team: [M., L., Tang, Zhong, Miller, Li, Zhang: PNAS19, SPA20, etc]

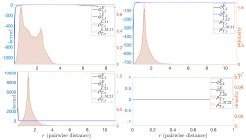
- Data: many trajectories $\{X_{[0,T]}^{(m)}\}_{m=1}^{M}$, $\nu = 0; \nu > 0$, **finite N**
- Function space of learning: $\phi \in L^2(\rho_T)$ with $\rho_T \leftarrow |X_t^j - X_t^i|$
- Nonparametric estimation ($Ac = b$)



Opinion Dynamics



Lennard-Jones



Prey-Predator

Previous work: finite N

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

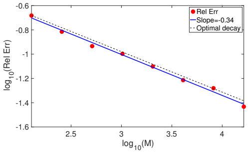
Interacting particles/agents:

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

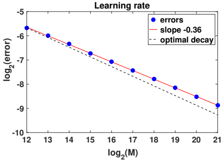
Maggioni JHU team: [M., L., Tang, Zhong, Miller, Li, Zhang: PNAS19, SPA20, etc]

- Identifiability: a coercivity condition for $L^2(\rho_T)$
- Optimal convergence rate:

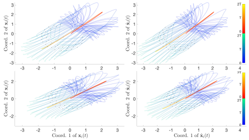
$$\mathbb{E}_{\mu_0}[\|\widehat{\phi}_{T,M,\mathcal{H}_{n_*}} - \phi_{true}\|_{L^2(\rho_T)}] \le C\left((\log M)/M\right)^{\frac{s}{2s+1}}.$$



Opinion Dynamics



Lennard-Jones



Prey-Predator

What if $N \to \infty$?

~~Data: many trajectories $\{X^{(m)}_{[0,T]}\}_{m=1}^{M}$;~~

Data: density $u(x,t) = \lim_{N\to\infty} \sum_{i=1}^{N} \delta(X_t^i - x)$

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

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

What if $N \to \infty$?

~~Data: many trajectories $\{X^{(m)}_{[0,T]}\}_{m=1}^{M}$;~~

Data: density $u(x,t) = \lim_{N\to\infty} \sum_{i=1}^{N} \delta(X_t^i - x)$

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

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

How to estimate ϕ from data?

$$\text{Minimize } \mathcal{E}_0(\psi) = \int_0^T \int_{\mathbb{R}^d} \left|\nabla.(u(K_\psi * u)) - g\right|^2 dx\, dt?$$

(with $g = \partial_t u - \nu \Delta u$)

Derivatives not available from data.

Outline

1. Motivation and problem statement
2. Nonparametric learning
   - A probabilistic error functional
   - Identifiability: function spaces of learning
   - Rate of convergence
3. Numerical examples
4. Ongoing work and open problems

A probabilistic error functional

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

- Expectation of the negative log-likelihood of the process

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

- Derivative-in-space free!
- $\overline{G}_T$ is a reproducing kernel for a RKHS

$$\langle \phi, \psi \rangle_{\overline{G}_T} := \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \langle (K_\phi * u), (K_\psi * u) \rangle u(x,t) dx\, dt = \int_{\mathbb{R}+} \int_{\mathbb{R}+} \phi(r)\psi(s)\, \overline{G}_T(r,s)\, dr\, ds$$

- $\psi = \sum_{i=1}^n c_i \phi_i \Rightarrow \mathcal{E}(\psi) = c^\top A c - 2b^\top c$ with $A_{ij} = \langle\!\langle \phi_i, \phi_j \rangle\!\rangle_{\overline{G}_T}$

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

Discrete data

From data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$: $\mathcal{H}_n = \mathrm{span}\{\phi_i\}_{i=1}^n$,

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

- Inverse problem: well-posed/ identifiable, A^{-1}?
- Choice of $\mathcal{H}_n$: $\{\phi_i\}$ and n ?
- Convergence rate when $\Delta x = M^{-1/d} \to 0$?
  $\to$ hypothesis testing and model selection

Invertibility of A and function space

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

$$A_{ij} = \langle\langle \phi_i, \phi_j \rangle\rangle_{\overline{G}_T},$$

with integral kernel $\overline{G}_T \to$ RKHS $H_{\overline{G}_T}$.

- if $\{\phi_i\}$ orthonormal in $H_{\overline{G}_T}$: $A = I_n$
- if $\{\phi_i\}$ orthonormal in $L^2(\rho_T)$: minimal eigenvalue of A =

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

  - measure $\rho_T \leftarrow |\overline{X}_t - \overline{X}'_t|$ (“pairwise distance”)

Error bounds

$\mathbb{H} = L^2(\rho_T)$ or RKHS $H_{\overline{G}_T}$.

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

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

- If if $\mathbb{H} = L^2(\rho_T)$: assume coercivity condition on $\mathcal{H}$ with $c_{\mathcal{H},T} > 0$, if $\mathbb{H}$ = RKHS, set $c_{\mathcal{H},T} = 1$
- $\Delta x + \Delta t$ comes from numerical integrator (Riemann sum)
- Dominating order: $n\Delta x$ (if $\Delta t = 0$)

Optimal dimension and rate of convergence

Total error: trade-off

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

Theorem (Lang-Lu20)

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

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

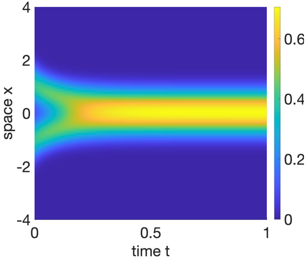
Outline

1. Motivation and problem statement
2. Nonparametric learning
3. Numerical examples
   - Granular media: smooth kernel $\phi(r) = 3r^2$
   - Opinion dynamics: piecewise linear ϕ
   - Repulsion-attraction: singular ϕ
4. Ongoing work and open problems

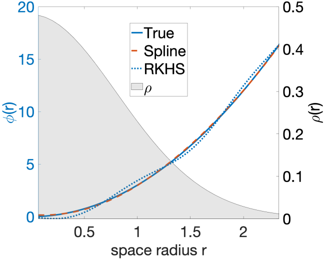
Numerical example 1: granular media

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

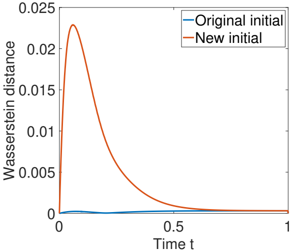
where $K_\phi(x) = \nabla(\Phi(|x|)) = \phi(|x|)\frac{x}{|x|}$. $\phi(r) = 3r^2$



The solution $u(x, t)$

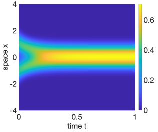


Estimators of ϕ

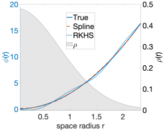


Wasserstein $W_2(u, \widehat{u})$

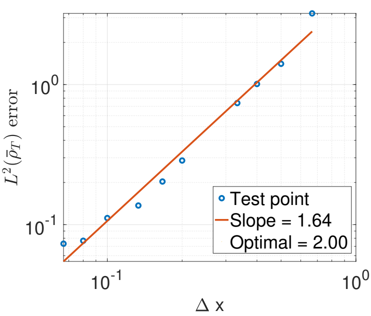
Numerical example 1: granular media



The solution $u(x,t)$

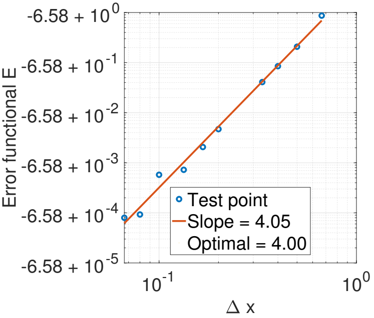


Estimators of ϕ



Convergence rate of $L^2(\rho_T)$ error

almost optimal



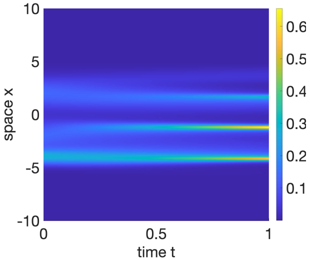
Convergence rate of $\mathcal{E}_{M,L}$

almost optimal

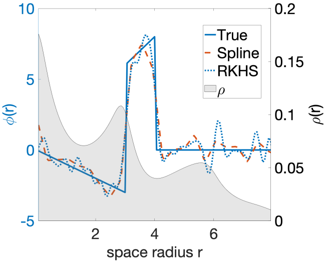
Numerical Example 2: opinion dynamics

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

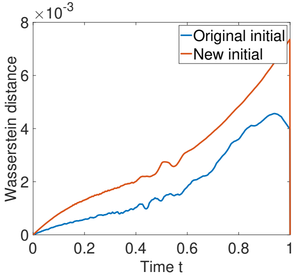
where $K_\phi(x) = \nabla(\Phi(|x|)) = \phi(|x|)\frac{x}{|x|}$. $\phi(r)$ piecewise linear



The solution $u(x,t)$

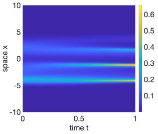


Estimators of ϕ

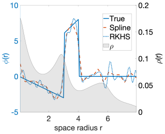


Wasserstein $W_2(u, \widehat{u})$

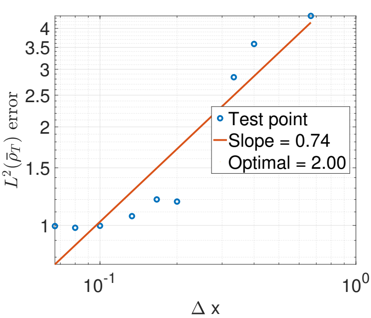
Numerical Example 2: opinion dynamics



The solution $u(x, t)$

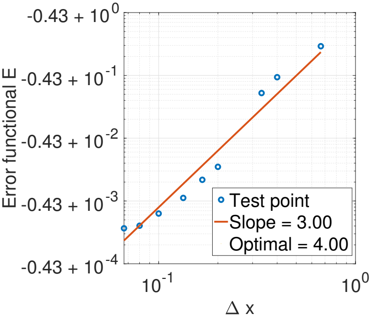


Estimators of ϕ



Convergence rate of $L^2(\rho_T)$ error

sub-optimal ($\phi \notin W^{1,\infty}$)



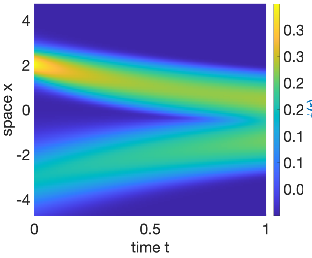
Convergence rate of $\mathcal{E}_{M,L}$

sub-optimal

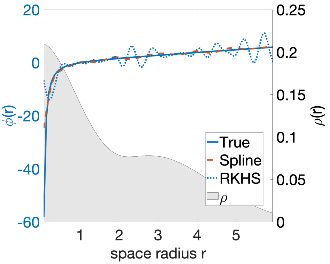
Numerical example 3: repulsion-attraction

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

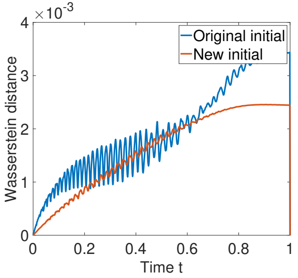
where $K_\phi(x) = \nabla(\Phi(|x|)) = \phi(|x|)\frac{x}{|x|}$. $\phi(r) = r - r^{-1.5}$ singular



The solution $u(x,t)$

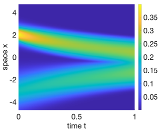


Estimators of ϕ

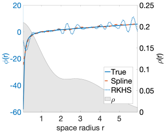


Wasserstein $W_2(u, \widehat{u})$

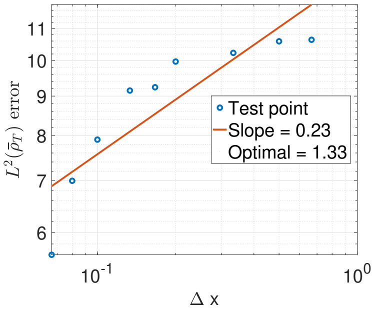
Numerical example 3: repulsion-attraction



The solution $u(x, t)$

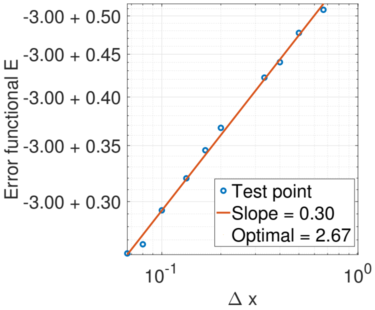


Estimators of ϕ



Convergence rate of $L^2(\rho_T)$ error

low rate: theory does not apply



Convergence rate of $\mathcal{E}_{M,L}$

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: Algorithm

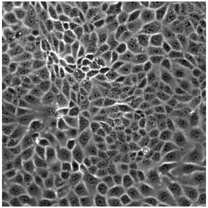
- A probabilistic error functional
- Estimator by least squares

Theory guidance

- Choice of hypothesis space basis functions & dimension
- Function space of learning: RKHS v.s. $L^2(\rho_T)$
- Optimal learning rate

Open problem and future directions

- Learning/computation:
  - 2nd-order systems
  - High-dimensional state space (Monte Carlo)
  - non-radial interaction kernel
  - partial observation of large systems
- Coercivity condition on $L^2(\rho_T)$
  - $\overline{G}_T \to$ strictly positive integral operator?
- Singular kernels?
- Real data applications: learning cell-dynamics



References

- Quanjun Lang, F.Lu. Learning interaction kernels in mean-field equations of 1st-order systems of interacting particles. In preparation. 2020
- F. Lu, M.Maggioni, and S. Tang. Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories. arXiv2007
- Zehong Zhang and F. Lu, Cluster prediction for opinion dynamics from partial observations. arXiv2007
- Z. Li, F. Lu, S. Tang, C. Zhang, and M. Maggioni. On the identifiability of interaction functions of particle systems. to appear on SPA20
- F. Lu, M. Maggioni, and S. Tang. Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. arXiv1912
- F. Lu, M. Maggioni, S. Tang and M. Zhong. Nonparametric inference of interaction laws in systems of agents from trajectory data. PNAS, 2019
- M. Bongini, M. Fornasier, M. Maggioni and M. Hansen. Inferring Interaction Rules From Observations of Evolutive Systems I: The Variational Approach. M3AS, 27(05), 909-951, 2017

Thank you!