Nonparametric inference of interaction laws in particle/agent systems

Fei Lu

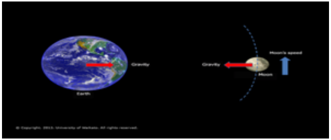
Department of Mathematics, Johns Hopkins University

Joint with:
Mauro Maggioni, Sui Tang and Ming Zhong

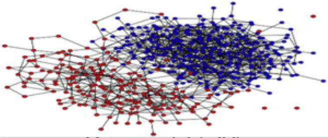
February 13, 2019
CSCAMM Seminar, UMD

Motivation

Q: What is the law of interaction between particles/agents?



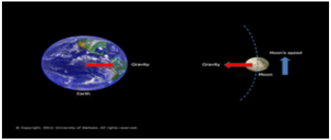
Popkin. Nature(2016)



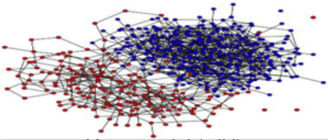
Voter model (wiki)

Motivation

Q: What is the law of interaction between particles/agents?



Popkin. Nature(2016)



Voter model (wiki)

$$m\ddot{x}_i(t) = -\nu \dot{x}_i(t) + \frac{1}{N}\sum_{j=1, j\neq i}^{N} K(x_i, x_j),$$

- Newton's law of gravitation:

$$K(x,y) = G\frac{m_1 m_2}{r^2}, r = |x-y|$$

- Molecular fluid: $K(x,y) = \nabla_x[\Phi(|x-y|)]$
  Lennard-Jones potential: $\Phi(r) = \frac{c_1}{r^{12}} - \frac{c_2}{r^6}$.

- flocking birds/school of fish

$$K(x,y) = \phi(|x-y|)\psi(\langle x, y\rangle)$$

- opinion/voter models, bacteria models ...[a]

---

[a] (1) Cucker+Smale: On the mathematics of emergence. 2007. (2) Vicsek+Zafeiris: Collective motion. 2012. (3) Mostch+Tadmor: Heterophilious Dynamics Enhances Consensus. 2014 ...

An inference problem:

Infer the rule of interaction in the system

$$m\ddot{x}_i(t) = -\nu\dot{x}_i(t) + \frac{1}{N}\sum_{j=1, j\neq i}^{N} K(x_i - x_j), \quad i = 1, \cdots, N,\ x_i(t) \in \mathbb{R}^d$$

from observations of trajectories.

- x_i is the position of the i-th particle/agent
- Data: many independent trajectories $\{\boldsymbol{x}^j(t) : t \in \mathcal{T}\}_{j=1}^{M}$
- **Goal**: infer ϕ in $K(x) = -\nabla\Phi(|x|) = -\phi(|x|)x$

$m = 0 \Rightarrow$ a first order system

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1, j\neq i}^{N} \phi_{true}(|x_i - x_j|)(x_j - x_i) := [\mathbf{f}_\phi(\boldsymbol{x}(t))]_i$$

Least squares regression: with $\mathcal{H}_n = \mathrm{span}\{e_i\}_{i=1}^n$,

$$\hat{\phi}_n = \arg\min_{\phi\in\mathcal{H}_n} \mathcal{E}_M(\phi) := \sum_{m=1}^{M} \|\dot{\boldsymbol{x}}^m - \mathbf{f}_\phi(\boldsymbol{x}^m)\|^2$$

- How to choose the hypothesis space $\mathcal{H}_n$?
- Inverse problem well-posed/ identifiability?
- Consistency and rate of “convergence”?

Outline

1. Learning via nonparametric regression:
   - A regression measure and function space
   - Identifiability: a coercivity condition
   - Consistency and rate of convergence
2. Numerical examples
   - A general algorithm
   - Lennard-Jones model
   - Opinion dynamics and multiple-agent systems
3. Open problems

The dynamical system:

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1, j\neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i) := [\mathbf{f}_\phi(\boldsymbol{x}(t))]_i$$

Admissible set ($\approx$ globally Lipschitz):

$$\mathcal{K}_{R,S} := \{\varphi \in W^{1,\infty} : \text{ supp } \varphi \in [0,R], \sup_{r\in[0,R]} [|\varphi(r)| + |\varphi'(r)|] \leq S\}$$

Data: M-trajectories $\{\boldsymbol{x}^m(t) : t \in \mathcal{T}\}_{m=1}^{M}$

- $\boldsymbol{x}^m(0) \overset{i.i.d}{\sim} \mu_0 \in \mathcal{P}(\mathbb{R}^{dN})$
- $\mathcal{T} = [0, T]$ or $\{t_1, \cdots, t_L\}$ with $\dot{\mathbf{x}}(t_i)$

Goal: nonparametric inference[1] of ϕ

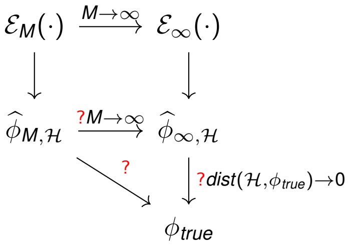
[1] (1) Bongini, Fornasier, Hansen, Maggioni: Inferring Interaction Rules for mean field equations, M3AS, 2017.
(2) Binev, Cohen, Dahmen, Devore and Temlyakov: Universal Algorithms for learning theory, JMLR 2005.
(3) Cucker, Smale: On the mathematical foundation of learning. Bulletin of AMS, 2001.

$$\hat{\phi}_{M,\mathcal{H}} = \underset{\phi \in \mathcal{H}}{\arg\min}\, \mathcal{E}_M(\phi) := \frac{1}{ML} \sum_{l,m=1}^{L,M} \|\mathbf{f}_{\phi}(\boldsymbol{X}^m(t_l)) - \dot{\boldsymbol{X}}^m(t_l)\|^2$$

- $\mathcal{E}_M(\phi)$ is quadratic in ϕ, and $\mathcal{E}_M(\phi) \geq \mathcal{E}_M(\phi_{true}) = 0$
- The minimizer exists for any $\mathcal{H} = \mathcal{H}_n = span\{\phi_1, \ldots, \phi_n\}$

Agenda

- a function space with metric $dist(\widehat{\phi}, \phi_{true})$;
- Learnability:
  - Convergence of estimators?
  - Convergence rate?



Review of classical nonparametric regression:

Estimate $\phi(z) = \mathbb{E}[Y|Z=z] : \mathbb{R}^D \to \mathbb{R}$ from data $\{z_i, y_i\}_{m=1}^M$.

- $\{z_i, y_j\}$ are iid samples;
- $\hat{\phi}_n := \arg\min_{f\in\mathcal{H}_n} \mathcal{E}_M(f) := \sum_{m=1}^M \|y_i - f(z_i)\|^2$
- Optimal rate: if $\mathrm{dist}(\mathcal{H}_n, \phi_{true}) \lesssim n^{-s}$ and $n_* = (M/\log M)^{\frac{1}{2s+1}}$,
$$\|\hat{\phi}_{n_*} - \phi\|_{L^2(\rho_Z)} \lesssim M^{-\frac{s}{2s+D}}$$

Review of classical nonparametric regression:

Estimate $\phi(z) = \mathbb{E}[Y|Z = z] : \mathbb{R}^D \to \mathbb{R}$ from data $\{z_i, y_i\}_{m=1}^M$.

- $\{z_i, y_j\}$ are iid samples;
- $\hat{\phi}_n := \arg\min_{f \in \mathcal{H}_n} \mathcal{E}_M(f) := \sum_{m=1}^M \|y_i - f(z_i)\|^2$
- Optimal rate: if $\mathrm{dist}(\mathcal{H}_n, \phi_{true}) \lesssim n^{-s}$ and $n_* = (M/\log M)^{\frac{1}{2s+1}}$,
$$\|\hat{\phi}_{n_*} - \phi\|_{L^2(\rho_Z)} \lesssim M^{-\frac{s}{2s+D}}$$

Our learning of kernel $\phi : \mathbb{R}_+ \to \mathbb{R}$ from data $\{\boldsymbol{x}^m(t)\}$

$$\dot{x}_i(t) = \frac{1}{N} \sum_{j=1, j \neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i)$$

- $\{r_{ijt}^m := |x_i^m(t) - x_j^m(t)|\}$ not iid
- The values of $\phi(r_{ijt}^m)$ unknown

Regression measure

Distribution of pairwise-distances $\rho : \mathbb{R}_+ \to \mathbb{R}$

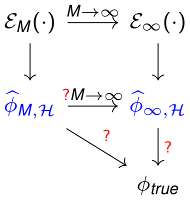
$$\rho_T(r) = \frac{1}{\binom{N}{2} L} \sum_{l,i,i'=1,i<i'}^{L,N} \mathbb{E}_{\mu_0} \delta_{r_{ii'}(t_l)}(r)$$

- unknown, estimated by empirical distribution $\rho_T^M \xrightarrow{M\to\infty} \rho_T$ (LLN)
- intrinsic to the dynamics

Regression function space $L^2(\rho_T)$

- the admissible set $\subset L^2(\rho_T)$
- $\mathcal{H} =$ piecewise polynomials $\subset L^2(\rho_T)$
- singular kernels $\subset L^2(\rho_T)$

Identifiability: a coercivity condition



$$\hat{\phi}_{M,\mathcal{H}} = \arg\min_{\phi\in\mathcal{H}} \mathcal{E}_M(\phi)$$

$$\mathcal{E}_\infty(\hat{\phi}) - \mathcal{E}_\infty(\phi) = \frac{1}{NT}\int_0^T \mathbb{E}_{\mu_0}\|\mathbf{f}_{\hat{\phi}-\phi}(\boldsymbol{X}(t))\|^2 dt \geq c\|(\hat{\phi}-\phi)(\cdot)\cdot\|^2_{L^2(\rho_T)}$$

Coercivity condition. There exists $c_T > 0$ s.t. for all $\varphi(\cdot)\cdot \in L^2(\rho_T)$

$$\frac{1}{NT}\int_0^T \mathbb{E}_{\mu_0}\|\mathbf{f}_\varphi(\boldsymbol{x}(t))\|^2 dt = \langle\langle\varphi,\varphi\rangle\rangle \geq c_T\|\varphi(\cdot)\cdot\|^2_{L^2(\rho_T)}$$

- coercivity: bilinear functional $\langle\langle\varphi,\psi\rangle\rangle := \frac{1}{NT}\int_0^T \mathbb{E}_{\mu_0}\langle\mathbf{f}_\varphi,\mathbf{f}_\psi\rangle(\boldsymbol{x}(t))dt$
- controls condition number of regression matrix

Consistency of estimator

Theorem (L. Maggioni, Tang, Zhong)

Assume the coercivity condition. Let $\{\mathcal{H}_n\}$ be a sequence of compact convex subsets of $L^\infty([0,R])$ such that $\inf_{\varphi\in\mathcal{H}_n}\|\varphi-\phi_{true}\|_\infty \to 0$ *as $n\to\infty$. Then*

$$\lim_{n\to\infty}\lim_{M\to\infty}\|\widehat{\phi}_{M,\mathcal{H}_n}(\cdot)\cdot-\phi_{true}(\cdot)\cdot\|_{L^2(\rho_T)}=0,\ \textit{almost surely.}$$

- For each n, compactness of $\{\widehat{\phi}_{M,\mathcal{H}_n}\}$ and coercivity implies that $\widehat{\phi}_{M,\mathcal{H}_n}\to\widehat{\phi}_{\infty,\mathcal{H}_n}$ in L^2
- Increasing $\mathcal{H}_n$ and coercivity implies consistency.
- In general, truncation to make $\mathcal{H}_n$ compact

Optimal rate of convergence

Theorem (L. Maggioni, Tang, Zhong)

Let $\{\mathcal{H}_n\}$ be a seq. of compact convex subspaces of $L^\infty[0,R]$ s.t.

$$\dim(\mathcal{H}_n) \leq c_0 n, \text{ and } \inf_{\varphi\in\mathcal{H}_n} \|\varphi - \phi_{true}\|_\infty \leq c_1 n^{-s}.$$

Assume the coercivity condition. Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then

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

- The 2nd condition is about regularity: $\phi \in C^s$
- Choose $\mathcal{H}_n$ according to s and M

Prediction of future evolution

Theorem (L. Maggioni, Tang, Zhong)

Denote by $\widehat{\boldsymbol{X}}(t)$ and $\boldsymbol{X}(t)$ the solutions of the systems with kernels $\widehat{\phi}$ and ϕ respectively, starting from the same initial conditions that are drawn i.i.d from μ_0. Then we have

$$\mathbb{E}_{\mu_0}[\sup_{t\in[0,T]} \|\widehat{\boldsymbol{X}}(t) - \boldsymbol{X}(t)\|^2] \lesssim \sqrt{N}\|\widehat{\phi}(\cdot)\cdot - \phi(\cdot)\cdot\|^2_{L^2(\rho_T)},$$

- Follows from Grownwall's inequality

Outline

1. Learning via nonparametric regression:
   - A regression measure and function space
   - Learnability: a coercivity condition
   - Consistency and rate of convergence
2. Numerical examples
   - A general algorithm
   - Lennard-Jones model
   - Opinion dynamics and multiple-agent systems
3. Open problems

The regression algorithm

$$\mathcal{E}_M(\varphi) = \frac{1}{LMN} \sum_{l,m,i=1}^{L,M,N} \left\| \dot{\boldsymbol{x}}_i^{(m)}(t_l) - \sum_{i'=1}^{N} \frac{1}{N} \varphi(r_{i,i'}^m(t_l)) \boldsymbol{r}_{i,i'}^m(t_l) \right\|^2 ,$$

$$\mathcal{H}_n := \{\varphi = \sum_{p=1}^{n} a_p \psi_p(r) : \mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n\},$$

$$\mathcal{E}_{L,M}(\varphi) = \mathcal{E}_{L,M}(\mathbf{a}) = \frac{1}{M} \sum_{m=1}^{M} \|\mathbf{d}^m - \Psi_L^m \mathbf{a}\|_{\mathbb{R}^{LNd}}^2 .$$

$$\frac{1}{M} \sum_{m=1}^{M} A_L^m \mathbf{a} = \frac{1}{M} \sum_{m=1}^{M} b_L^m , \text{ rewrite as } A_M \mathbf{a} = b_M$$

- can be computed parallelly
- Caution: choice of $\{\psi_p\}$ affects condi(A_M)

Assume coercivity condition: $\langle\langle \varphi, \varphi \rangle\rangle \geq c_T \|\varphi(\cdot) \cdot \|^2_{L^2(\rho_T)}$.

Proposition (Lower bound on smallest singular value of A_M)

Let $\{\psi_1, \cdots, \psi_n\}$ be a basis of $\mathcal{H}_n$ s.t.

$$\langle \psi_p(\cdot)\cdot, \psi_{p'}(\cdot)\cdot \rangle_{L^2(\rho_T^L)} = \delta_{p,p'}, \|\psi_p\|_\infty \leq S_0.$$

Let $A_\infty = \left(\langle\langle \psi_p, \psi_{p'} \rangle\rangle\right)_{p,p'} \in \mathbb{R}^{n \times n}$. Then $\sigma_{\min}(A_\infty) \geq c_L$.
Moreover, A_∞ is the a.s. limit of A_M. Therefore, for large M, the smallest singular value of A_M satisfies with a high probability that

$$\sigma_{\min}(A_M) \geq (1 - \epsilon) c_L$$

- Choose $\{\psi_p(\cdot)\cdot\}$ linearly independent in $L^2(\rho_T)$
- Piecewise polynomials: on a partition of support(ρ_T)
- Finite difference $\approx$ derivatives $\Rightarrow$ an $O(\Delta t)$ error to estimator

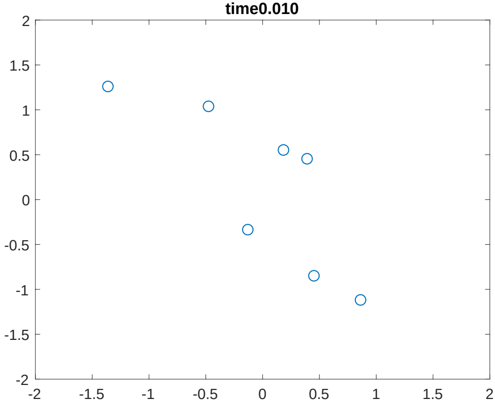
Implementation

1. Approximate regression measure
   - Estimate the ρ_T with large datasets
   - Partition on support(ρ_T)
2. Construct hypothesis space $\mathcal{H}$:
   - degree of piecewise polynomials
   - set dimension of $\mathcal{H}$ according to sample size
3. Regression:
   - Assemble the arrays (in parallel)
   - Solve the normal equation

The Lennard-Jones potential

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

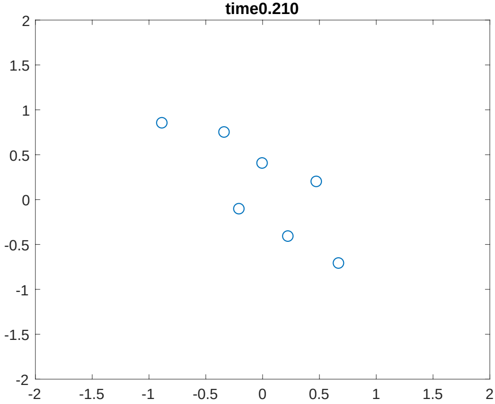
$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1, j\neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i)$$



The Lennard-Jones potential

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

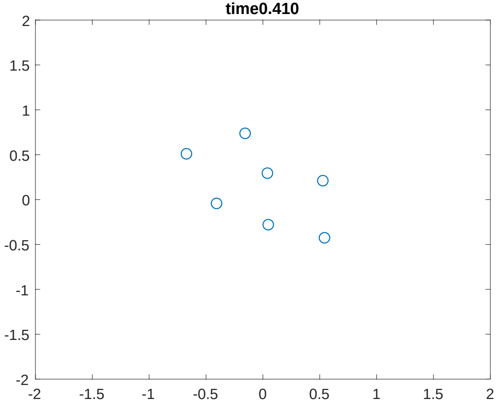
$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1,j\neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i)$$



The Lennard-Jones potential

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

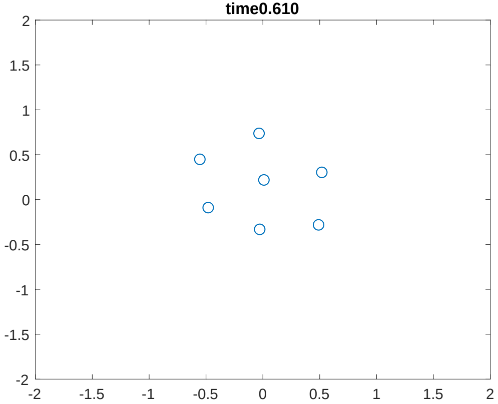
$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1, j\neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i)$$



The Lennard-Jones potential

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

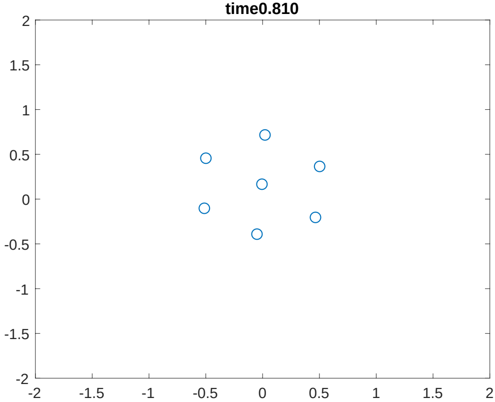
$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1,j\neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i)$$



The Lennard-Jones potential

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

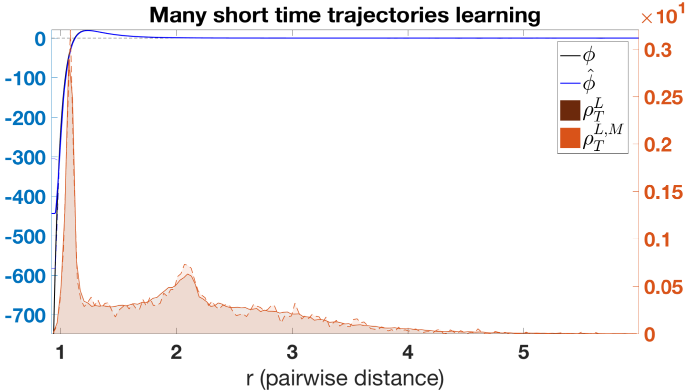
$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1, j\neq i}^{N} \phi(|x_i - x_j|)(x_j - x_i)$$



The Lennard-Jones potential

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

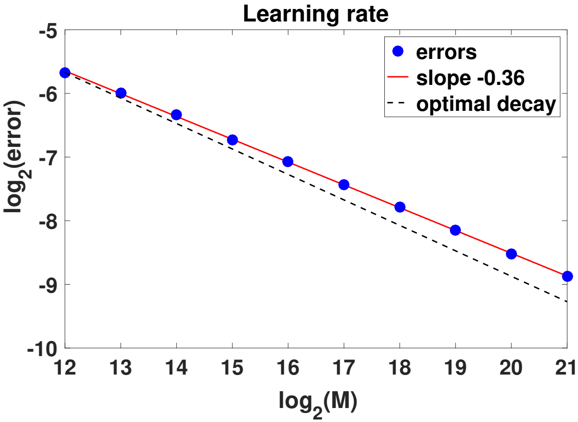
- piecewise linear estimator; Gaussian initial conditions.



Optimal rate

$$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \Rightarrow \quad \phi(r)r = V_{LJ'}(r)$$

- V_{LJ} is highly singular, yet we get close to optimal rate (-0.4).



Example: Opinion Dynamics

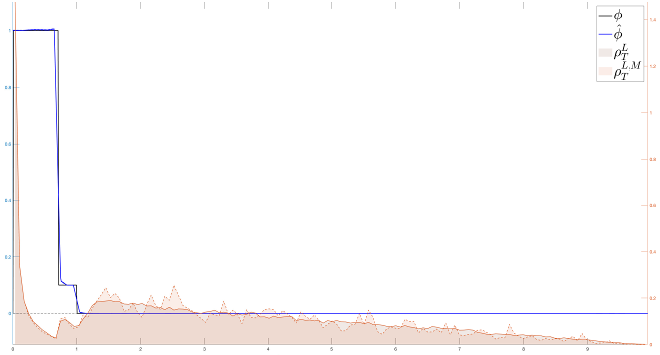
$$\phi(r) = \begin{cases} 1, & 0 \le r < \frac{1}{\sqrt{2}}, \\ 0.1, & \frac{1}{\sqrt{2}} \le r < 1, \\ 0, & 1 \le r. \end{cases}$$

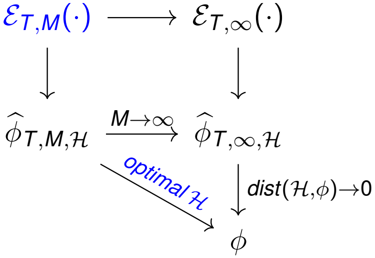
$N = 10, d = 1.$

$M = 250, \mu_0 = Unif[0, 10]^{10}$

$\mathcal{T} = [0, 10]$, 200 discrete instances

$\mathcal{H} =$ piecewise constant functions





Learning theory

- extended the classical regression theory
- a coercivity condition for identifiability

Theory guided regression algorithms

- Selection of $\mathcal{H}$ (basis functions & dimension)
- Measurement of error of estimators
- Optimal learning rate
- Model selection

Open directions

- 1st- and 2nd-order heterogeneous agent dynamics
  - Multiple type of agents (leader-follower, predator-prey)
  - Angle and/or alignment based interactions
- Stochastic systems, mean field equations
- Partial and noisy observations
- The coercivity condition

The coercivity condition

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1}^{N} \phi(|x_i - x_j|)(x_j(t) - x_i(t)), \quad x_i \in \mathbb{R}^d$$

- Exchangeability: particles indistinguishable.
- $U = x_1(t) - x_2(t)$, $V = x_1(t) - x_3(t)$ correlated.

The coercivity condition is equivalent to

$$c_t := \inf_{g\in\mathcal{H}, \|g\|_2=1} \mathbb{E}_{p_t}\left[g(|U|)g(|V|)\frac{\langle U, V\rangle}{|U||V|}\right] \geq C > 0$$

where $g(r) = \phi(r)r$, and p_t is the joint distribution of (U, V).

Conjecture

The coercivity condition holds for compact $\mathcal{H}$ if $c_0 > 0$.

Current status:

- Proved $c_0 > 0$ if initial distributions with iid particles, or Gaussian with $\mathrm{cov}(U) - \mathrm{cov}(U, V) = \lambda I_d$.
- Basic idea: positive definite kernel

$$\mathbb{E}\left[g(|U|)g(|V|)\frac{\langle U, V\rangle}{|U||V|}\right] = \int g(r)g(s)\mathcal{K}(r, s)drds$$

  - The kernel $\mathcal{K}$ is a positive definite kernel:

$$\mathcal{K}(r, s) = \sum_i \lambda_i e_i(r)e_i(s).$$

- When $t > 0$: no proof yet, verified by numerical tests

A route in plan:

$$dX(t) = -\nabla J(X(t))dt + \sqrt{2/\beta}dW(t), \quad X(t) \in \mathbb{R}^{dN}$$

$$J(X) = \frac{1}{2N}\sum_{i,j}\Phi(|X_i - X_j|)$$

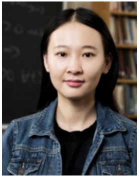
- Ergodic with invariant meas.: $p_\infty(x) = Z^{-1}e^{-\beta J(x)}$
- Need: positive definite kernel for all t

$$p_t(u,v) := \int p_t(x_1, x_1 - u, x_1 - v, x_{4:N})dx_1 dx_{4:N}.$$

- Uniform in $t \in [0,\infty]$ for lower bound of the spectrum on $\mathcal{H}$

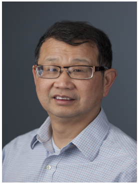
Thanks to my collaborators


Mauro Maggioni


Sui Tang


Ming Zhong

Helpful discussions with


Yaozhong Hu


Cheng Zhang


Yulong Lu

References

- F. Lu, M. Maggioni, and S. Tang. Learning interaction rules in deterministic interacting particle systems: a Monte Carlo Approach. Preprint.
- F. Lu, M. Maggioni, S. Tang and M. Zhong. Discovering governing laws of interaction in heterogeneous agents dynamics from observations. arXiv
- M. Bongini, M. Fornasier, M. Maggioni and M. Hansen. Inferring Interaction Rules From Observations of Evolutive Systems I: The Variational Approach, Mathematical Models and Methods in Applied Sciences, 27(05), 909-951, 2017

Thank you!