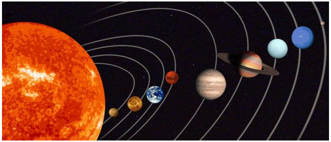
Nonparametric learning of interaction kernels in interacting particle systems

Fei Lu

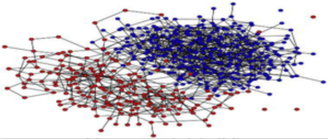
Department of Mathematics, Johns Hopkins University

September, 2022.
Third Symposium on Machine Learning and Dynamical Systems

What is the law of interaction ?

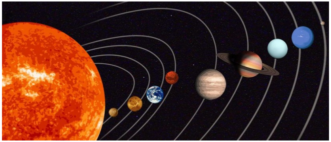


Popkin. Nature(2016)

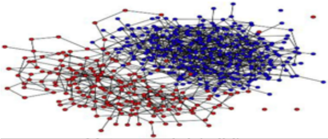


Voter model (wiki)

What is the law of interaction ?



Popkin. Nature(2016)



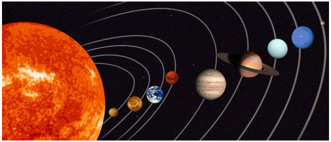
Voter model (wiki)

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

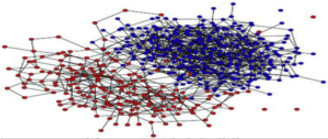
$$K_\phi(x, y) = \nabla_x [\Phi(|x - y|)] = \phi(|x - y|) \frac{x - y}{|x - y|}.$$

- Newton's law of gravity $\phi(r) = G \frac{m_1 m_2}{r^2}$
- Lennard-Jones potential: $\Phi(r) = \frac{c_1}{r^{12}} - \frac{c_2}{r^6}$.

What is the law of interaction ?



Popkin. Nature(2016)



Voter model (wiki)

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

$$K_\phi(x,y) = \nabla_x[\Phi(|x-y|)] = \phi(|x-y|)\frac{x-y}{|x-y|}.$$

- Newton's law of gravity $\phi(r) = G\frac{m_1 m_2}{r^2}$
- Lennard-Jones potential: $\Phi(r) = \frac{c_1}{r^{12}} - \frac{c_2}{r^6}$.

- flocking birds, bacteria/cells ?
- opinion/voter/multi-agent models, ...? [a]

Infer the interaction kernel from data?

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

Learn interaction kernel $K_\phi(x,y) = \phi(|x-y|)\frac{x-y}{|x-y|}$

$$dX_t^i = \frac{1}{N}\sum_{j=1}^{N} K_\phi(X_t^j, X_t^i)dt + \sqrt{2\nu}dB_t^i \quad \Leftrightarrow R_\phi(\boldsymbol{X}_t) = \dot{\boldsymbol{X}}_t - \sqrt{2\nu}\dot{\boldsymbol{B}}_t$$

Finite N: [a]

- Data: M trajectories of particles : $\{\boldsymbol{X}_{t_1:t_L}^{(m)}\}_{m=1}^{M}$
- Statistical learning
- ODE/SDEs: Opinion Dynamics, Lennard-Jones, Prey-Predator; 1st/2nd order

Learn interaction kernel $K_\phi(x,y) = \phi(|x-y|)\frac{x-y}{|x-y|}$

$$dX_t^i = \frac{1}{N}\sum_{j=1}^{N} K_\phi(X_t^j, X_t^i)dt + \sqrt{2\nu}dB_t^i \quad \Leftrightarrow R_\phi(\boldsymbol{X}_t) = \dot{\boldsymbol{X}}_t - \sqrt{2\nu}\dot{\boldsymbol{B}}_t$$

Finite N: [a]

- Data: M trajectories of particles : $\{\boldsymbol{X}_{t_1:t_L}^{(m)}\}_{m=1}^{M}$
- Statistical learning
- ODE/SDEs: Opinion Dynamics, Lennard-Jones, Prey-Predator; 1st/2nd order

Large N ($>> 1$)[b]

- Data: concentration density $\{u(x_m, t_l) \approx N^{-1}\sum_i \delta(X_{t_l}^i - x_m)\}_{m,l}$

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

- Inverse problem for PDE

[a] [Maggioni, Lu, Tang, Zhong, Miller, Li, Zhang: PNAS19, SPA20,FOC22,JMLR21] [b] [Lang-Lu 20,21]

Learning kernels in operators: $R_\phi : \mathbb{X} \to \mathbb{Y}$

$$dX_t^i = \frac{1}{N}\sum_{j=1}^{N} K_\phi(X_t^j, X_t^i)dt + \sqrt{2\nu}dB_t^i \quad \Leftrightarrow R_\phi(\boldsymbol{X}_t) = \dot{\boldsymbol{X}}_t - \sqrt{2\nu}\dot{\boldsymbol{B}}_t$$

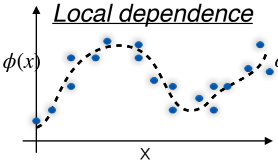
$$\partial_t u = \nu\Delta u + \nabla \cdot [u(K_\phi * u)] \quad \Leftrightarrow R_\phi[u(\cdot, t)] = f(\cdot, t)$$

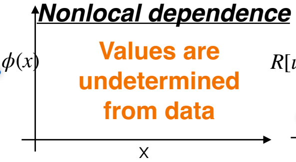
Learning kernels in operators: $R_\phi : \mathbb{X} \to \mathbb{Y}$

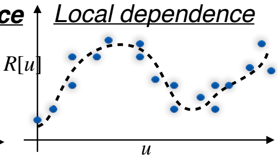
$$dX_t^i = \frac{1}{N}\sum_{j=1}^{N} K_\phi(X_t^j, X_t^i)dt + \sqrt{2\nu}dB_t^i \quad \Leftrightarrow R_\phi(\boldsymbol{X}_t) = \dot{\boldsymbol{X}}_t - \sqrt{2\nu}\dot{\boldsymbol{B}}_t$$

$$\partial_t u = \nu\Delta u + \nabla\cdot[u(K_\phi * u)] \quad \Leftrightarrow R_\phi[u(\cdot,t)] = f(\cdot,t)$$

Classical learning	**Learning kernel**	**Operator learning**
$\{(x_i, \phi(x_i) + \epsilon_i)\}$	$\{(u_k, R_\phi[u_k] + \eta_k)\}$	$\{(u_k, R[u_k] + \eta_k)\}$



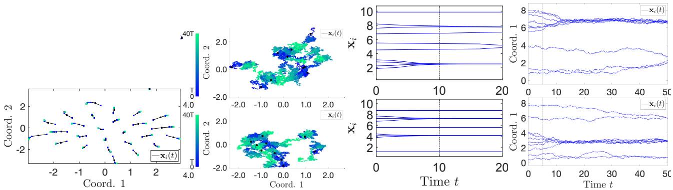




Nonparametric learning:
Loss function? Identifiability? Convergence?

Finite many particles

$$R_{\phi}(\boldsymbol{X}_t) = \dot{\boldsymbol{X}}_t - \sqrt{2\nu}\dot{\boldsymbol{B}}_t \quad \& \text{ Data} \qquad \Rightarrow \hat{\phi}_{n,M} = \arg\min_{\psi\in\mathcal{H}_n} \mathcal{E}_M(\psi)$$

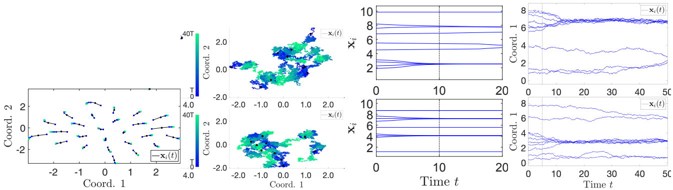


- Loss function (log-likelihood, or mse for ODE)
- Regression: with $\psi = \sum_i c_i\phi_i \in \mathcal{H}_n = \mathrm{span}\{\phi_i\}_{i=1}^n$:

$$\mathcal{E}(\psi) = c^\top A c - 2b^\top c \quad \Rightarrow \quad \widehat{\phi}_{n,M} = \sum_{i=1}^{n} \widehat{c}_i\phi_i, \quad \widehat{c} = A^{-1}b$$

Finite many particles

$$R_\phi(\boldsymbol{X}_t) = \dot{\boldsymbol{X}}_t - \sqrt{2\nu}\dot{\boldsymbol{B}}_t \quad \& \text{ Data} \qquad \Rightarrow \hat{\phi}_{n,M} = \arg\min_{\psi \in \mathcal{H}_n} \mathcal{E}_M(\psi)$$



- Loss function (log-likelihood, or mse for ODE)
- Regression: with $\psi = \sum_i c_i \phi_i \in \mathcal{H}_n = \mathrm{span}\{\phi_i\}_{i=1}^n$:

$$\mathcal{E}(\psi) = c^\top A c - 2b^\top c \quad \Rightarrow \quad \widehat{\phi}_{n,M} = \sum_{i=1}^n \widehat{c}_i \phi_i, \quad \widehat{c} = A^{-1} b$$

  - Choice of $\mathcal{H}_n$ & function space of learning?
  - Well-posed/ identifiability?
  - Convergence and rate?

Classical learning theory

Given: Data $\{(x_m, y_m)\}_{m=1}^{M} \sim (X, Y)$
Goal: find f s.t. $Y = f(X)$

$$\mathcal{E}(f) = \mathbb{E}|Y - f(X)|^2 = \|f - f_{true}\|^2_{L^2(\rho_X)}$$

Learning kernel

Given: Data $\{\boldsymbol{X}^{(m)}_{[0,T]}\}_{m=1}^{M}$
Goal: find ϕ s.t. $\dot{\boldsymbol{X}}_t = R_\phi(\boldsymbol{X}_t)$

$$\mathcal{E}(\phi) = \mathbb{E}|\dot{\boldsymbol{X}} - R_\phi(\boldsymbol{X})|^2 \neq \|\phi - \phi_{true}\|^2_{L^2(\rho)}$$

Minimization: $f = \sum_{i=1}^{n} c_i\phi_i \in \mathcal{H}_n$, $\nabla\mathcal{E}_M = 0 \Rightarrow \quad \widehat{f}_{n,M} = \sum_i \widehat{c}_i\phi_i$.

Classical learning theory

Given: Data$\{(x_m, y_m)\}_{m=1}^M \sim (X, Y)$

Goal: find f s.t. $Y = f(X)$

$$\mathcal{E}(f) = \mathbb{E}|Y - f(X)|^2 = \|f - f_{true}\|^2_{L^2(\rho_X)}$$

Learning kernel

Given: Data$\{\boldsymbol{X}^{(m)}_{[0,T]}\}_{m=1}^M$

Goal: find ϕ s.t. $\dot{\boldsymbol{X}}_t = R_\phi(\boldsymbol{X}_t)$

$$\mathcal{E}(\phi) = \mathbb{E}|\dot{\boldsymbol{X}} - R_\phi(\boldsymbol{X})|^2 \neq \|\phi - \phi_{true}\|^2_{L^2(\rho)}$$

Minimization: $f = \sum_{i=1}^n c_i\phi_i \in \mathcal{H}_n$, $\nabla\mathcal{E}_M = 0 \Rightarrow \quad \widehat{f}_{n,M} = \sum_i \widehat{c}_i\phi_i$.

- Function space: $L^2(\rho_X)$.
- Identifiability: $\mathbb{E}[Y|X = x] = \arg\min_{f\in L^2(\rho_X)} \mathcal{E}(f)$.
- $A \approx \mathbb{E}[\phi_i(X)\phi_j(X)] = I_n$ by setting $\{\phi_i\}$ ONB in $L^2(\rho_X)$.
- Error bounds for $\widehat{f}_{n_M}$

- Function space: $L^2(\rho)$. measure $\rho \sim |X^i - X^j|$
- Identifiability: $\arg\min_{\phi\in L^2(\rho)} \mathcal{E}(\phi)$??
- $A \approx \mathbb{E}[R_{\phi_i}(\boldsymbol{X})R_{\phi_j}(\boldsymbol{X})] \approx I_n$?? A Coercivity condition
- **Error bounds for $\widehat{\phi}_{n_M}$?**

Assume a coercivity condition on $\mathcal{H}$

$$\langle\langle \phi, \phi \rangle\rangle = \mathbb{E}[R_\phi(\boldsymbol{X}) R_\phi(\boldsymbol{X})] \geq c_{\mathcal{H}} \|\phi\|^2_{L^2(\rho)}, \quad \forall \phi \in \mathcal{H}$$

- $c_{\mathcal{H}} = \frac{1}{N-2}$ for $\mathcal{H} = L^2(\rho)$ for some (LLMTZ21); open

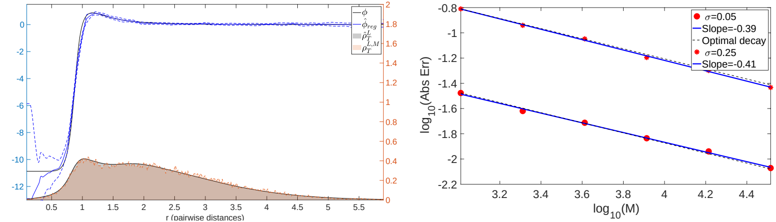
Theorem (LZTM19,LMT22)

Let $\{\mathcal{H}_n\}$ compact convex in L^∞ with $\mathrm{dist}(\phi_{true}, \mathcal{H}_n) \sim n^{-s}$. Assume the coercivity condition $\cup_n \mathcal{H}_n$. Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$. Then

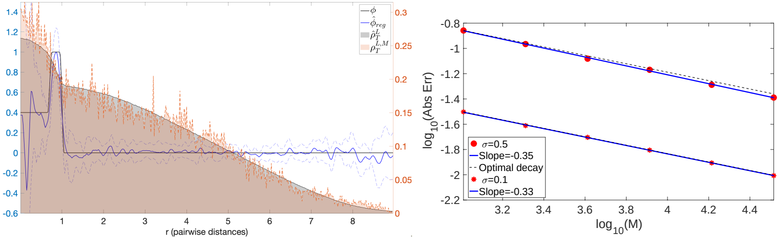
$$\mathbb{E}_{\mu_0}[\|\widehat{\phi}_{M,\mathcal{H}_{n_*}} - \phi_{true}\|_{L^2(\rho)}] \leq C \left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}}.$$

- Concentration for r.v. or martingale
- $\dim(\mathcal{H}_n)$ adaptive to s ($\phi \in C^s$) and M:

Lennard-Jones kernel estimators:



Opinion dynamics kernel estimators:



Inverse problem for Mean-field PDE

Goal: Identify ϕ from discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$ of

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

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

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

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

- $= -\mathbb{E}[$ 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 free
- Suitable for high dimension

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

Nonparametric regression

$$\mathcal{E}(\psi) = \langle\!\langle \psi, \psi \rangle\!\rangle - 2 \langle\!\langle \psi, \phi \rangle\!\rangle ,$$

LS-regression $\psi = \sum_{i=1}^{n} c_i \phi_i \in \mathcal{H}_n$:

$$\mathcal{E}(\psi) = c^\top A c - 2 b^\top c \quad \Rightarrow \quad \widehat{\phi}_{n,M} = \sum_{i=1}^{n} \widehat{c}_i \phi_i, \quad \widehat{c} = A^{-1} b$$

- Choice of $\mathcal{H}_n$ & function space of learning?
- Inverse problem well-posed/ identifiability?
- Convergence and rate? $\Delta x = M^{-1/d} \to 0$

Identifiability

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

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

Convergence rate

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

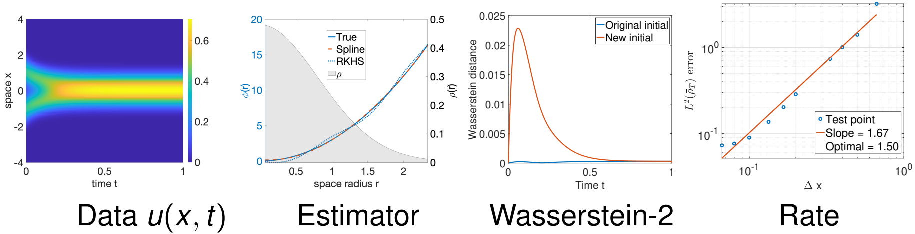
Theorem (Numerical error bound [Lang-Lu20])

Let $\mathcal{H} = \mathrm{span}\{\phi_i\}_{i=1}^n$ *s.t.* $\|\widehat{\phi}_n - \phi\|_{\mathbb{H}} \lessapprox n^{-s}$ *. Assume the coercivity condition on* $\cup\mathcal{H}_n$*. Then, with dimension* $n \approx (\Delta x)^{-\alpha/(s+1)}$*, we have:*

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

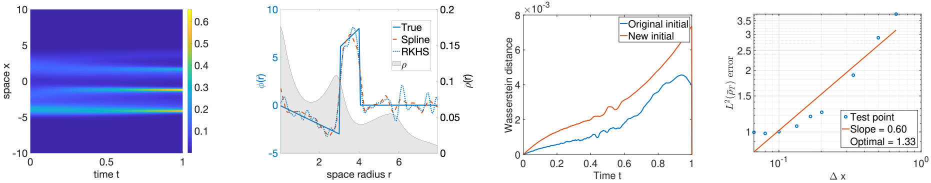
- Δx^α comes from numerical integrator (e.g.,Riemann sum)
- Trade-off: numerical error v.s. approximation error

Example 1: granular media $\phi(r) = 3r^2$



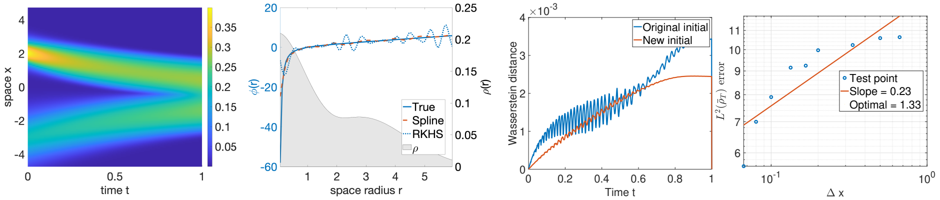
- near optimal rate ($\phi \in W^{1,\infty}$)

Example 2: Opinion dynamics $\phi(r)$ piecewise linear



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

Example 3: repulsion-attraction $\phi(r) = r - r^{-1.5}$ (singular)



- low rate: theory does not apply

Learning kernels in operators: regularization

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:
$$R_{\phi}[u] = \nabla \cdot [u(K_{\phi} * u)] = \partial_t u - \nu \Delta u$$
- integral/nonlocal operators,... linear inverse problems

Regularization

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

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

Regularization norm $\|\cdot\|_*$?

$$\mathcal{E}_\lambda(\psi) = \mathcal{E}(\psi) + \lambda\|\psi\|_*^2$$

Regularization

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

Regularization norm $\|\cdot\|_*$?

$$\mathcal{E}_\lambda(\psi) = \mathcal{E}(\psi) + \lambda\|\psi\|_*^2$$

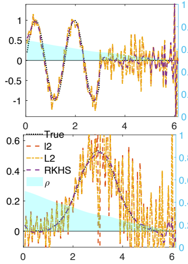
ANSWER: norm of the RKHS $H_G = L_G^{1/2}L^2(\rho)$ [Lu+Lang+An22]:

- search in the correct fun.space
- Data Adaptive RKHS Tikhonov Regularization

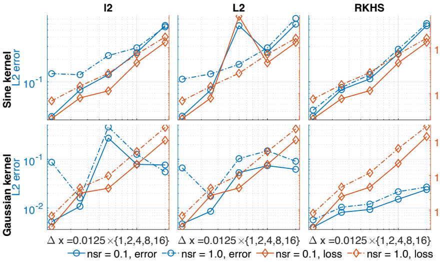
DARTR: Data Adaptive RKHS Tikhonov Regularization

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

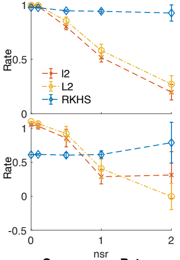
- Recover kernel from discrete noisy data
- Consistent convergence as mesh refines



Typical estimators, $\Delta x = 0.05$



Convergence of Estimators, nsr = 0.1 & 1



Convergence Rates

Summary and future directions

Nonparametric learning of interaction kernels

- Finite N: ode/sde
- Mean-field equation

Learning kernel in operators via regression:

- probabilistic loss functionals
- Identifiability
- Convergence

DARTR: regularization for ill-posed linear inverse problems

Future directions/open questions

- Coercivity condition
- General IPS settings:
  - Aggression equations (inviscid MFE)
  - High-D, non-radial kernels (Monte Carlo)
  - Learning from stationary distributions
  - Multiple MFE solutions
  - Systems on graph
- kernels in operator
  - Convergence and Minimax rate?
  - DARTR in Bayesian inverse p
  - Applications: 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
- F. Lu, M. Maggioni and S. Tang. Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories. FoCM21.
- F. Lu, M. Maggioni and S. Tang: Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. JMLR21
- Z. Li, F. Lu, M. Maggioni, S. Tang and C. Zhang: On the identifiability of interaction functions in systems of interacting particles. SPA21
- F. Lu, M Zhong, S Tang and M Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. PNAS19