Introduction to Nonparametric Learning of Kernels in Operators

Fei Lu

Department of Mathematics, Johns Hopkins University

Plan:
Lecture 1. Overview and a review of classical learning theory
Lecture 2. Learning interaction kernels in interacting particle systems
Lecture 3. Coercivity condition and minimax rate of convergence
Lecture 4. Learning interaction kernels in mean-field equations
Lecture 5. Data adaptive RKHS Tikhonov regularization
Lecture 6. Small noise analysis of RKHS regularizations

Lecture 4. Learning interaction kernels in mean-field equations

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

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

- N small: Data= M trajectories $\{\mathbf{X}_{t_1:t_L}^{(m)}\}_{m=1}^{M} \Rightarrow$ statistical learning
- $N=\infty$: Data= density $\{u_N(x,t_l) = \frac{1}{N}\sum_i \delta_{X_{t_l}^i}(x)\}$ or $\{u(x_m,t_l)\}_{m,l}$

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

Goal: algorithm, identifiability, convergence

1. Loss functional
2. Regression and numerical results
3. identifiability
4. Convergence rate

Inverse problem for Mean-field PDE

Goal: Identify from data ϕ in

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

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

- Two types of data:
  - low-D: discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$ with mesh $\{x_m\}$
  - high-D: (partial) particle samples $\{u_N(x, t_l) \approx M^{-1} \sum_{i=1}^{M} \delta(X_{t_l}^i - x)\}$
- Two types of equations: $\nu > 0$ or $\nu = 0$.

How (in a unified framework)? Computationally efficient?

Inverse problem for Mean-field PDE

Goal: Identify from data ϕ in

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

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

- Two types of data:
  - low-D: discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$ with mesh $\{x_m\}$
  - high-D: (partial) particle samples $\{u_N(x, t_l) \approx M^{-1} \sum_{i=1}^{M} \delta(X_{t_l}^i - x)\}$
- Two types of equations: $\nu > 0$ or $\nu = 0$.

How (in a unified framework)? Computationally efficient?

- Variational /regression: loss functional;
- Identifiability;
- Ill-posed: regularization

0

Outline

Loss functional in variational approach

$$\widehat{\phi} = \arg\min_{\phi \in \mathcal{H}} \mathcal{E}(\phi)$$

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

- Discrepancy: $\mathcal{E}(\phi) = \|\partial_t u - \nu \Delta u - \nabla.(u(K_\phi * u))\|^2$
  - derivatives approx. from discrete data
  - Weak SINDY [Bortz etc21,22], denoising+smoothing [Kang+Liao etc22]
- Wasserstein-2: $\mathcal{E}(\phi) = W_2(u^\phi, u)$

  costly: requires many PDE simulations in optimization
- **A probabilistic loss function** $\downarrow$
- A self-test loss function: simple, general

A probabilistic loss functional

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

- $= -\mathbb{E}[$ log-likelihood $]$: McKean–Vlasov 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
- Applicable to high-dimensional: $Z_t = \overline{X}_t - \overline{X}_t'$

$$\mathcal{E}(\phi) = \frac{1}{T}\int_0^T \Big(\mathbb{E}|\mathbb{E}[K_\phi(Z_t)|\overline{X}_t]|^2 - 2\nu \mathbb{E}[\nabla \cdot K_\phi(Z_t)] + \partial_t \mathbb{E}\Phi(Z_t) \Big)\, dt$$

Derivation of the loss function

Girsanov theorem [Oksendal-SDE]
$dX_t = -b(X_t)dt + \sigma dB_t \, 0 \leq t \leq T; \to \mathbb{P}_b$ on the path space
$dX_t = -0(X_t)dt + \sigma dB_t \, 0 \leq t \leq T; \to \mathbb{P}_0$
Under the Novikov condition (integrability), with

$$\frac{d\mathbb{P}_b}{d\mathbb{P}_0}(X_{[T_0,s]}) = \exp\Big(-\frac{1}{2\sigma^2}\int_{T_0}^{s} |b(X_t)|^2 dt + \int_{T_0}^{s} \langle b(X_t), dX_t\rangle \Big)$$

Mckean-Vlasov SDE: $d\overline{X}_t = -K_{\phi_{true}} * u(\overline{X}_t, t)dt + \sqrt{2\nu} dB_t$,

$$\mathcal{E}_{\overline{X}_{[0,T]}}(\phi) = -\frac{2}{T}\log\frac{d\mathbb{P}_\phi}{d\mathbb{P}_0} = \frac{1}{\nu T}\int_0^T \left(|K_\phi * u(\overline{X}_t)|^2 dt + 2\langle K_\phi * u(\overline{X}_t), d\overline{X}_t\rangle \right),$$

$$\mathbb{E}\mathcal{E}_{\overline{X}_{[0,T]}}(\phi) = \frac{1}{\nu T}\int_0^T \int_{\mathbb{R}^d} \left[\left|K_\phi * u\right|^2 u - 2u\langle K_{\phi_{true}} * u, K_\phi * u\rangle \right] dx\, dt,$$

Note: $\partial_t u = \nu\Delta u + \nabla\cdot[u(K_{\phi_{true}} * u)]$; $K_\phi = \nabla\Phi$; and
$\int_{\mathbb{R}^d} -u\langle K_{\phi_{true}} * u, K_\phi * u\rangle dx = \int_{\mathbb{R}^d} (\nabla\cdot[u(K_{\phi_{true}} * u)])\Phi * u dx$;

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

Derivation of the loss function

Girsanov theorem [Oksendal-SDE]
$dX_t = -b(X_t)dt + \sigma dB_t\, 0 \le t \le T; \to \mathbb{P}_b$ on the path space
$dX_t = -0(X_t)dt + \sigma dB_t\, 0 \le t \le T; \to \mathbb{P}_0$
Under the Novikov condition (integrability), with

$$\frac{d\mathbb{P}_b}{d\mathbb{P}_0}(X_{[T_0,s]}) = \exp\Big(-\frac{1}{2\sigma^2}\int_{T_0}^{s} |b(X_t)|^2 dt + \int_{T_0}^{s} \langle b(X_t), dX_t\rangle \Big)$$

Mckean-Vlasov SDE: $d\overline{X}_t = -K_{\phi_{true}} * u(\overline{X}_t, t)dt + \sqrt{2\nu}dB_t$,

$$\mathcal{E}_{\overline{X}_{[0,T]}}(\phi) = -\frac{2}{T}\log\frac{d\mathbb{P}_\phi}{d\mathbb{P}_0} = \frac{1}{\nu T}\int_0^T \Big(|K_\phi * u(\overline{X}_t)|^2 dt + 2\langle K_\phi * u(\overline{X}_t), d\overline{X}_t\rangle \Big),$$

$$\mathbb{E}\mathcal{E}_{\overline{X}_{[0,T]}}(\phi) = \frac{1}{\nu T}\int_0^T \int_{\mathbb{R}^d} \Big[\big|K_\phi * u\big|^2 u - 2u\langle K_{\phi_{true}} * u, K_\phi * u\rangle \Big]\, dx\, dt,$$

Note: $\partial_t u = \nu\Delta u + \nabla\cdot[u(K_{\phi_{true}} * u)]$; $K_\phi = \nabla\Phi$; and
$\int_{\mathbb{R}^d} -u\langle K_{\phi_{true}} * u, K_\phi * u\rangle dx = \int_{\mathbb{R}^d} (\nabla\cdot[u(K_{\phi_{true}} * u)])\Phi * u dx$;

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

What if $\nu = 0$?

A self-test loss function

Weak form of the equation $\partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi * u)]$

$$\begin{aligned} \langle \partial_t u, v \rangle &= \nu \langle \Delta u, v \rangle + \langle \nabla \cdot [u(K_\phi * u)], v \rangle \\ &= \nu \langle u, \Delta v \rangle - \langle u(K_\phi * u), \nabla v \rangle, \quad \forall v \in C_c^\infty \ldots \end{aligned}$$

"Take" $v = \Phi * u$ s.t. $\nabla \Phi(|x|) = K_\phi(x) = \phi(|x|)\frac{x}{|x|}$,

$$\langle \partial_t u, \Phi * u \rangle = \nu \langle u, \Delta \Phi * u \rangle - \langle u(K_\phi * u), K_\phi * u \rangle$$

We regain the loss function

$$\mathcal{E}(\phi) = \int_0^T [\langle \partial_t u, \Phi * u \rangle - \nu \langle u, \Delta \Phi * u \rangle + \langle u(K_\phi * u), K_\phi * u \rangle] dt$$

- regardless of $\nu = 0$ or > 0
- OK to data $\{u_N(x, t_l) \approx M^{-1} \sum_{i=1}^M \delta(X_{t_l}^i - x)\}$
- Applicable to other PDEs: self-test (a better name?)

A self-test loss function

Quiz: what test function to use for the system also with a kinetic potential?

$$\partial_t u = \nu \Delta u + \nabla \cdot [u(\nabla \Phi * u + \nabla V)]$$

A self-test loss function

Quiz: what test function to use for the system also with a kinetic potential?

$$\partial_t u = \nu \Delta u + \nabla \cdot [u(\nabla \Phi * u + \nabla V)]$$

Test function: $v = \Phi * u + V$

Outline

1. Loss functional
2. Regression and numerical results
3. identifiability
4. Convergence rate

Nonparametric regression

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

Data: $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$ with mesh $\{x_m\}$, $\Delta x = x_m - x_{m-1}$
Regression: $\phi = \sum_{i=1}^n c_i\phi_i \in \mathcal{H}_n$:

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

- Choice of $\mathcal{H}_n$ & function space of learning?
  - Exploration measure $\rho_T \leftarrow |\overline{X}_t - \overline{X}_t'|$

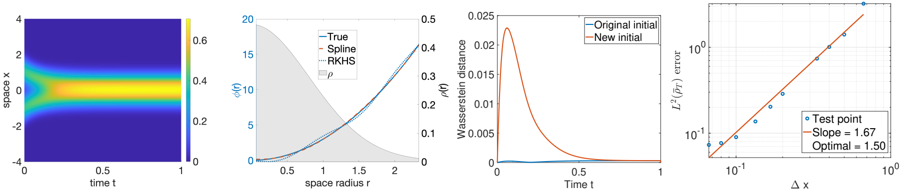
$$\rho_T(dr) := \frac{1}{T}\int_0^T \rho_t(dr)dt, \quad \rho_t(dr) := \mathbb{E}[\delta(|\overline{X}_t' - \overline{X}_t| \in dr)].$$

$$\rho_L^M(dr) = \frac{1}{L}\sum_{l=1}^L \sum_{m,m'=1}^{M,M} u(x_m, t_l)u(x_{m'}, t_l)\delta_{|x_m - x_{m'}|}(r)dr.$$

- Inverse problem well-posedness/ identifiability? $\arg\min_{\phi\in L^2(\rho)} \mathcal{E}(\phi)$
- Convergence and rate? $\Delta x \to 0$

Smooth kernel

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

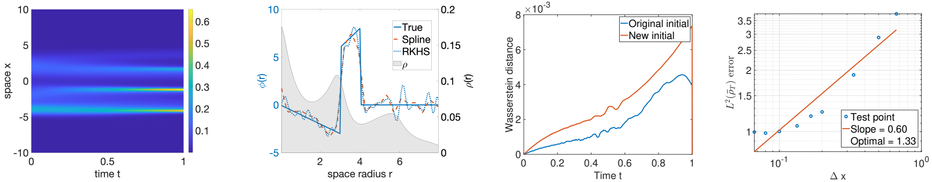


Data $u(x,t)$ Estimator Wasserstein-2 Rate

- Near optimal rate ($\phi \in W^{1,\infty}$)
- Other examples:
  - suboptimal when ϕ discontinuous,
  - low rate for singular ϕ

Discontinuous kernel

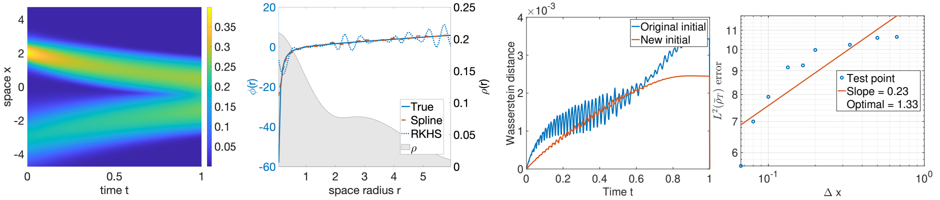
Example 2: Opinion dynamics $\phi(r)$ piecewise linear



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

Singular kernel

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



- low rate: open

Outline

1. Loss functional
2. Regression and numerical results
3. identifiability
4. Convergence rate

Identifiability

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

Uniqueness of the minimizer in $\mathcal{H} = L^2_\rho$?

The quadratic term

$$\langle L_{\overline{G}}\phi, \phi\rangle = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left|K_\phi * u\right|^2 u dx = \int_0^\infty \int_0^\infty \phi(r)\phi(s)G(r,s)\, drds$$

$$= \int_0^\infty \int_0^\infty \phi(r)\phi(s)\frac{G(r,s)}{\rho(r)\rho(s)}\rho(r)\rho(s)\, drds$$

$$G(r,s) = \frac{1}{T}\int_0^T \int_{\mathbb{S}^d}\int_{\mathbb{S}^d}\int_{\mathbb{R}^d} \langle \xi,\eta\rangle\, (rs)^{d-1} u(x-r\xi,t)u(x-s\eta,t)u(x,t)dxd\xi d\eta dt.$$

$$\overline{G}(r,s) = \frac{G(r,s)}{\rho(r)\rho(s)}$$

$$\mathcal{E}(\phi) = \langle L_{\overline{G}}\phi,\phi\rangle - 2\langle \phi^D, \phi\rangle + const.$$

$$\nabla\mathcal{E}(\phi) = L_{\overline{G}}\phi - \phi^D = 0 \quad \Rightarrow \widehat{\phi} = L_{\overline{G}}^{-1}\phi^D$$

Identifiability

- Identifiability: $A^{-1}b \leftrightarrow L_{\overline{G}}^{-1}\phi^D \in L^2_\rho$
  - $L_{\overline{G}}$: positive compact operator
  - Function space of identifiability (FSOI) = $\overline{\mathrm{span}\{\psi_i\}_{\lambda_i>0}}$
  - Ill-posed $\to$ regularization: truncated SVD/L-curve.
- Coercivity condition on $\mathcal{H}$ (not L^2_ρ)

$$c_{\mathcal{H}} = \inf_{\phi\in\mathcal{H}, \|\phi\|_{L^2_\rho}=1} \langle L_{\overline{G}}\phi, \phi\rangle > 0$$

  - Finite-D $\mathcal{H}$
  - Infinite-D compact set (not a linear subspace in L^2_ρ)

Revisit the numerical results [Page **??**]

- The data are smooth
- The true kernel in/not in FSOI

Exploration measure v.s. Uniform measure

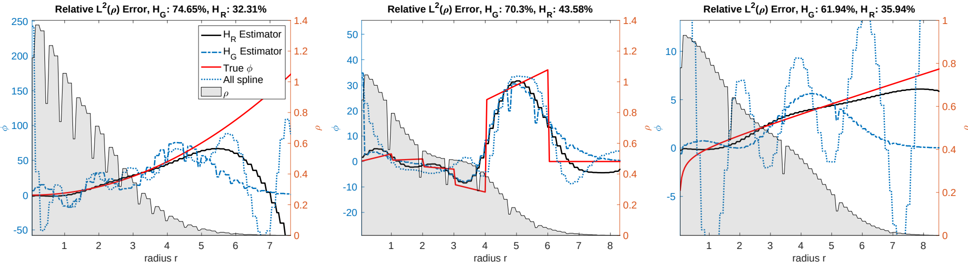
Can we use a uniform measure on $\mathrm{supp}(\rho)$?
The quadratic term

$$\langle L_{\overline{G}}\phi, \phi\rangle_{L^2_\rho} = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left|K_\phi * u\right|^2 u dx = \int_0^\infty \int_0^\infty \phi(r)\phi(s)G(r,s)\, drds$$
$$= \int_0^\infty \int_0^\infty \phi(r)\phi(s) \underbrace{\frac{G(r,s)}{\rho(r)\rho(s)}}_{\overline{G}(r,s)} \rho(r)\rho(s)\, drds = \langle L_G\phi, \phi\rangle_{L^2}$$

$$\mathcal{E}(\phi) = \langle L_{\overline{G}}\phi, \phi\rangle_{L^2_\rho} - 2\langle \phi^D, \phi\rangle_{L^2_\rho} + const.$$
$$\nabla\mathcal{E}(\phi) = L_{\overline{G}}\phi - \phi^D = 0 \Rightarrow \widehat{\phi}^{\overline{G}} = L_{\overline{G}}^{-1}\phi^D$$

$$\mathcal{E}(\phi) = \langle L_G\phi, \phi\rangle_{L^2} - 2\langle \phi^D, \phi\rangle_{L^2} + const.$$
$$\nabla\mathcal{E}(\phi) = L_G\phi - \phi^D = 0 \Rightarrow \widehat{\phi}^G = L_G^{-1}\phi^D$$

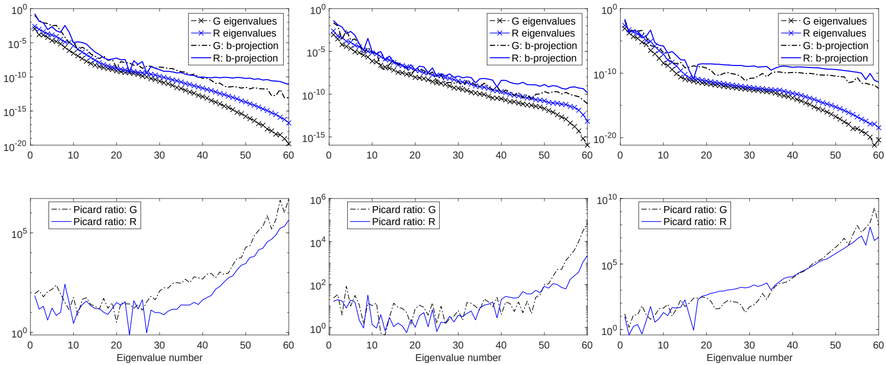
Exploration measure v.s. Uniform measure:



(a) Cubic potential (b) Opinion dynamics (c) Attraction-repulsion

Figure: Regularized estimators via truncated SVD. (R represents $\overline{G}$)

Exploration measure v.s. Uniform measure



(a) Cubic potential (b) Opinion dynamics (c) Attraction-repulsion

Figure: SVD analysis of the regression in three examples. (R represents $\overline{G}$ with L^2_ρ). The weighted SVD has larger eigenvalues than those of the unweighted SVD, and it has slightly smaller Picard ratios $\frac{u_i^\top b}{\sigma_i}$.

Outline

Convergence rate

Assume $\Delta t = 0$.

Theorem (Error bound [Lang-Lu22sisc])

Let $\mathcal{H}_n = \mathrm{span}\{\phi_i\}_{i=1}^n$ s.t. $\|\phi_{\mathcal{H}_n} - \phi\|_{L^2_\rho} \lessapprox n^{-s}$. Assume the coercivity condition on $\cup\mathcal{H}_n$. Then, with $n \approx (\Delta x)^{-\alpha/(s+1)}$, we have:

$$\|\widehat{\phi}_{n,M} - \phi\|_{L^2_\rho} \lessapprox c_{\mathcal{H}}^{-1} (\Delta x)^{\alpha s/(s+1)}$$

- Δx^α comes from numerical integrator (e.g.,Riemann sum)
  - In statistical learning: $\alpha = 1/2$ (Monte Carlo, CLT)
- Trade-off: numerical error v.s. approximation error

Skitch of proof:

$$\|\widehat{\phi}_{n,M} - \phi\|_{L^2_\rho} \leq \|\widehat{\phi}_{n,M} - \widehat{\phi}_n\|_{L^2_\rho} + \|\widehat{\phi}_n - \phi\|_{L^2_\rho}.$$

- $\widehat{\phi}_n$ the optimal estimator in $\mathcal{H}_n$
- Numerical error from Riemann sum:

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

$$\|\widehat{c}_{n,M,L} - \widehat{c}\| = \left\|A^{-1}_{n,M,L} b_{n,M,L} - A^{-1}b\right\|$$

$$\|A - A_{n,M,L}\| \leq nc^A(\Delta x + \Delta t),$$
$$\|b - b_{n,M,L}\| \leq \sqrt{n}c^b(\Delta x + \Delta t),$$

- Approximation error: $\|\widehat{\phi}_n - \phi\|_{L^2_\rho} \sim n^{-s}$.

Tradeoff: $g(n) = n(\Delta x)^{-\alpha} + n^{-s}$. $\Rightarrow$
$n_* \approx s^{-1/(s+1)}(\Delta x)^{\alpha/(s+1)}$, $g(n_*) \approx (\Delta x)^{-\alpha s/(s+1)}$.

Coercivity condition, ill-posedness, regularization

A good-looking estimator is relatively easy. Convergence is difficult.

- No coercivity condition on $\mathcal{H} = L^2_\rho$.
- Ill-posed: $\widehat{\phi} = L_{\overline{G}}^{-1}\phi^D$
  - $L_{\overline{G}}^{-1}$ is unbounded ($L_{\overline{G}}$ is compact with $\lambda_k \to 0$)
  - ϕ^D has error: $\phi^D = L_{\overline{G}}\phi_* + \xi$
  - Well-posed if $L_{\overline{G}}$ is finite rank

Coercivity condition, ill-posedness, regularization

A good-looking estimator is relatively easy. Convergence is difficult.

- No coercivity condition on $\mathcal{H} = L^2_\rho$.
- Ill-posed: $\widehat{\phi} = L_{\overline{G}}^{-1}\phi^D$
  - $L_{\overline{G}}^{-1}$ is unbounded ($L_{\overline{G}}$ is compact with $\lambda_k \to 0$)
  - ϕ^D has error: $\phi^D = L_{\overline{G}}\phi_* + \xi$
  - Well-posed if $L_{\overline{G}}$ is finite rank
- Regularization: direct and iterative methods
  - truncated SVD
  - Tikhonov:
  - Iterative methods

Summary

Inverse problems for mean-field PDE of interacting particles

- Construction of loss functions
- Nonparametric regression
- Identifiability
- Convergence of estimator
- Ill-posedness and Regularization

Learning kernels in operators:

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

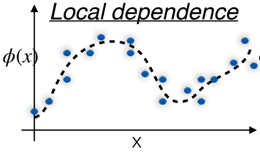
$$\partial_t u = \nu\Delta u + \nabla\cdot[u(K_\phi * u)] \quad \Leftrightarrow R_\phi[u(\cdot,t)] = f(\cdot,t)$$

Infer ϕ in $R_\phi[u] = f$ from data $\mathcal{D} = \{(u_k, f_k)\}_{k=1}^M$

- R_ϕ linear/nonlinear in u, but linear in ϕ
- Estimator: $\widehat{\phi} = L_{\overline{G}}^{-1}\phi^D$
  - Finite N: coercivity $\lambda_{min}(L_{\overline{G}}) \geq c_{\mathcal{H}}$. $L_{\overline{G}} = c_N L_G + \frac{N-1}{N^2}I$
  - Infinite N: ill-posed, $\lambda_k(L_{\overline{G}}) \to 0 \Rightarrow$ Regularization necessary

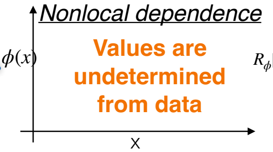
Classical learning

$$\{(x_i, \phi(x_i) + \epsilon_i)\}$$



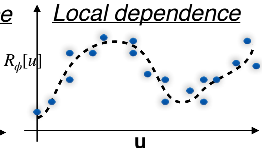
Learning kernel

$$\{(u_k, R_\phi[u_k] + \eta_k)\}$$

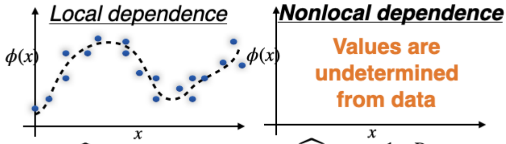


Operator learning

$$\{(u_k, R_\phi[u_k] + \eta_k)\}$$



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



	Classical learning	Learning kernel
Inversion	$\widehat{\phi} = I^{-1}\phi^D$	$\widehat{\phi} = L_G^{-1}\phi^D$
Regularization	$\widehat{\phi} = (I + \lambda Q)^{-1}\phi^D$	$\widehat{\phi} = (L_G + \lambda L_G^{-1})^{-1}\phi^D$