Statistical learning and inverse problems from interacting particle systems

Fei Lu
Department of Mathematics, Johns Hopkins University

Jan. 9, 2022. CAM colloquium, PSU
What is the law of interaction?

Newton's law of gravity

Lennard-Jones potential:

flocking birds, bacteria/cells?

Infer the interaction kernel from data?

3. Mostch+Tadmor: Heterophilious Dynamics Enhances Consensus. 2014...
What is the **law of interaction**?

\[
m_i \ddot{X}_i = -\gamma \dot{X}_i + \frac{1}{N} \sum_{j=1,j\neq i}^{N} K_\phi(X_i, X_j),
\]

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

- **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?

\[ m_i \ddot{X}_i^t = -\gamma \dot{X}_i^t + \frac{1}{N} \sum_{j=1, j \neq i}^{N} K_\phi(X_i^t, X_j^t), \]

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

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

Infer the interaction kernel from data?

\( ^a \)

Part 0: statistical learning & inverse problem

- Part 1: statistical learning — Finitely many particles
- Part 2: inverse problem — infinitely many particles
- Part 3: Regularization for learning kernels in operators
Learning the interaction kernel

\[ dX^i_t = \frac{1}{N} \sum_{j=1}^{N} K_\phi(X^j_t, X^i_t) dt + \sqrt{2\nu} dB^i_t \]

\[ \Leftrightarrow \dot{X}_t = R_\phi(X_t) + \sqrt{2\nu} \dot{B}_t \]

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

Finite N:

- Data: M trajectories of particles \( \{X^{(m)}_{t_1:t_L}\}_{m=1}^{M} \)
- Statistical learning

![Graph showing trajectories](image-url)
Learning the interaction kernel

\[ dX^i_t = \frac{1}{N} \sum_{j=1}^{N} K_\phi(X^j_t, X^i_t)dt + \sqrt{2\nu} dB^i_t \quad \Leftrightarrow \quad \dot{X}_t = R_\phi(X_t) + \sqrt{2\nu} \dot{B}_t \]

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

Finite N:
- Data: M trajectories of particles \( \{X^{(m)}_{t_1:t_L}\}_{m=1} \)
- Statistical learning

Large N (\( \gg 1 \))
- Data: density of particles \( \{u(x_m, t_l) \approx N^{-1} \sum_i \delta(X^i_{t_l} - x_m)\}_{m,l} \)
  \[ \partial_t u = \nu \Delta u + \nabla \cdot [u(K_\phi \ast u)] \]
- Inverse problem for PDEs
What’s in common and what’s different?

What is new from

- classical learning \( \{(x_i, y_i)\}_{i=1}^{M} \Rightarrow y = \phi(x) \)?

- operator learning \( \{u_k, f_k\}_{k=1}^{M} \Rightarrow f = R[u] \)?
Learning kernels in operators:

\[
dX^i_t = \frac{1}{N} \sum_{j=1}^{N} K_{\phi}(X^i_t, X^j_t) dt + \sqrt{2\nu} dB^i_t \quad \Leftrightarrow \quad R_{\phi}(X_t) = \dot{X}_t - \sqrt{2\nu} \dot{B}_t
\]

\[
\partial_t u = \nu \Delta u + \nabla \cdot [u(K_{\phi} * u)] \quad \Leftrightarrow \quad R_{\phi}[u(\cdot, t)] = f(\cdot, t)
\]
Learning kernels in operators:

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

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

Classical learning
\[\{(x_i, \phi(x_i) + \epsilon_i)\}\]

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

Operator learning
\[\{(u_k, R[u_k] + \eta_k)\}\]

Local dependence
\[\{(x_i, \phi(x_i))\}\]

Nonlocal dependence
\[\{u_k, R_\phi[u_k]\}\]

Values are undetermined from data
Part 1: Finitely many particles

Statistical learning from sample trajectories
Finitely many particles

\[ R_\phi(X_t) = \dot{X}_t - \sqrt{2\nu} \dot{B}_t \quad \text{& Data } \{X_{t_1:t_L}^{(m)}\}_{m=1}^M \]

- Loss function (or log-likelihood for SDEs):

\[ \hat{\phi}_{n,M} = \arg \min_{\phi \in \mathcal{H}_n} \mathcal{E}_M(\phi) = \frac{1}{M} \sum_{m=1}^{M} \int_0^T |\dot{X}_t^m - R_\phi(X_t^m)|^2 dt \]

- Nonparametric Regression: \( \mathcal{H}_n = \text{span}\{\phi_i\}_{i=1}^n \), \( \phi = \sum_i c_i \phi_i \)

\[ \mathcal{E}_M(\phi) = c^\top Ac - 2b^\top c \quad \Rightarrow \quad \hat{\phi}_{n,M} = \sum_{i=1}^{n} \hat{c}_i \phi_i, \quad \hat{c} = A^{-1}b \]
Finitely many particles

\[ R_\phi(X_t) = \dot{X}_t - \sqrt{2\nu} \dot{B}_t \quad \& \text{Data } \{X^{(m)}_{t_1:t_L}\}_{m=1}^M \]

- Loss function (or log-likelihood for SDEs):

\[ \hat{\phi}_{n,M} = \arg \min_{\phi \in \mathcal{H}_n} \mathcal{E}_M(\phi) = \frac{1}{M} \sum_{m=1}^M \int_0^T |\dot{X}_m^m - R_\phi(X_t^m)|^2 dt \]

- Nonparametric Regression: \( \mathcal{H}_n = \text{span}\{\phi_i\}_{i=1}^n, \phi = \sum_i c_i \phi_i \)

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

- Choice of \( \mathcal{H}_n \) \\& function space of learning?
- Well-posedness/ identifiability?
- Convergence and rate?
Classical learning in a nutshell

Data\(\{(x_m, y_m)\}_{m=1}^M \sim (X, Y) \Rightarrow \text{find } \phi \text{ s.t. } Y = \phi(X)\)

- Loss function: \(\hat{\phi}_{n,M} = \arg \min_{\phi \in \mathcal{H}_n} \mathcal{E}_M(\phi) = \frac{1}{M} \sum_{m=1}^M |Y_m - \phi(X_m)|^2\).
- Regression: with \(\psi = \sum_i c_i \phi_i \in \mathcal{H}_n = \text{span}\{\phi_i\}_{i=1}^n:\)

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

- Choice of \(\mathcal{H}_n \subset C^s\) in \(L^2(\rho_X)\): \(n_* = (M/\log M)^{1/(2s+d)}\)

- Well-posedness/identifiability: \(\phi_{optimal} = \mathbb{E}[Y|X = x]\)
  - minimax rate \(\mathbb{E}[\|\hat{\phi}_{n_*,M} - \phi_{optimal}\|_{L^2(\rho_X)}^2] \approx \left(\frac{\log M}{M}\right)^{s/(2s+d)}\)
**Classical learning theory**

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

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

$$\mathcal{E}(\phi) = \mathbb{E}\left|Y - \phi(X)\right|^2 = \|\phi - \phi_{true}\|_{L^2(\rho_X)}^2$$

**Learning kernel**

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

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

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

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

Goal: find \( \phi \) s.t. \( Y = \phi(X) \)

\[
\mathcal{E}(\phi) = \mathbb{E}|Y - \phi(X)|^2 = ||\phi - \phi_{true}||_{L^2(\rho_X)}^2
\]

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

Learning kernel

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

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

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

- Function space: \( L^2(\rho) \).
- Identifiability: \( \arg \min_{\phi \in L^2(\rho)} \mathcal{E}(\phi) \)??
- \( A \approx \mathbb{E}[R_{\phi_i}(X)R_{\phi_j}(X)] \geq c_H I_n \) Coercivity condition
Classical learning theory

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

Goal: find \( \phi \) s.t. \( Y = \phi(X) \)

\[
\mathcal{E}(\phi) = \mathbb{E}|Y - \phi(X)|^2 = \|\phi - \phi_{true}\|_{L^2(\rho_X)}^2
\]

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

Learning kernel

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

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

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

- Function space: \( L^2(\rho) \).
- Measure \( \rho \sim |X^i - X^j| \).
- Identifiability: \( \arg\min_{\phi \in L^2(\rho)} \mathcal{E}(\phi) \).
- \( A \approx \mathbb{E}[R_{\phi_i}(X)R_{\phi_j}(X)] \geq c_H I_n \) Coercivity condition

Error bounds for \( \hat{\phi}_{nM} \): asymptotic/non-asymptotic (CLT/concentration)

\[
\mathcal{E}(\hat{\phi}_{nM}) - \mathcal{E}(\phi_H) \geq c_H \|\hat{\phi}_{nM} - \phi_H\|^2
\]
Theorem (Convergence with minimax rate \cite{LZTM19,LMT21,LMT22})

Let \( \{ \mathcal{H}_n \} \) compact convex in \( L^\infty \) with \( \text{dist}(\phi_{\text{true}}, \mathcal{H}_n) \sim n^{-s} \). Assume the coercivity condition on \( \bigcup_n \mathcal{H}_n \). Set \( n_\ast = \left( \frac{M}{\log M} \right)^{\frac{1}{2s+1}} \). Then

\[
\mathbb{E}_{\mu_0}[\| \hat{\phi}_{n_\ast, M} - \phi_{\text{true}} \|_{L^2(\rho)}] \leq C \left( \frac{\log M}{M} \right)^{\frac{s}{2s+1}} .
\]
Lennard-Jones kernel estimators:

- \( r \) (pairwise distances)
- \( \log_{10}(M) \)
- \( \log_{10}(\text{Abs Err}) = 0.05 \)
- Slope = -0.39
- Optimal decay = 0.25
- Slope = -0.41

Opinion dynamics kernel estimators:

- \( r \) (pairwise distances)
- \( \log_{10}(M) \)
- \( \log_{10}(\text{Abs Err}) = 0.5 \)
- Slope = -0.35
- Optimal decay = 0.1
- Slope = -0.33
Coercivity condition on $\mathcal{H}$

$$\langle \phi, \phi \rangle = \frac{1}{T} \int_0^T \mathbb{E}[R_\phi(X_t)R_\phi(X_t)]dt \geq c_\mathcal{H} \| \phi \|^2_{L^2(\rho)}, \quad \forall \phi \in \mathcal{H}$$

- Partial results: $c_\mathcal{H} = \frac{1}{N-2}$ for $\mathcal{H} = L^2(\rho)$
  - Gaussian or $\Phi(r) = r^{2\beta}$ stationary \cite{LLMTZ21spa,LL20}
  - Harmonic analysis: strictly positive definite integral kernel

$$\mathbb{E}[\phi(|X - Y|)\phi(|X - Z|)\frac{\langle X - Y, X - Z \rangle}{|X - Y||X - Z|}] \geq 0, \forall \phi \in L^2(\rho)$$

- Open: non-stationary? A compact $\mathcal{H} \subset C(\text{supp}(\rho))$?
- No coercivity on $L^2(\rho)$ when $N \to \infty$ since $c_\mathcal{H} \to 0$
Part 2: Infinitely many particles

Inverse problem for mean-field PDEs
Goal: Identify $\phi$ from discrete data $\{u(x_m, t_l)\}_{m,l=1}^{M,L}$ of

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

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: \( \mathcal{E}(\phi) = \| \partial_t u - \nu \Delta u - \nabla \cdot (u(K_\phi * u)) \|^2 \)
  - 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 functional
A probabilistic loss functional

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

\[ = -\mathbb{E}\left[ \right. \mathrm{log-likelihood} \left. \right]: \text{McKean–Vlasov process} \]

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

- Derivative free
- Suitable for high dimension: \( Z_t = \bar{X}_t - \overline{X'}_t \)

\[ \mathcal{E}(\phi) = \frac{1}{T} \int_0^T \left( \mathbb{E}\left[ \mathbb{E}[K_{\phi}(Z_t)|\overline{X}_t]|^2 - 2\nu \mathbb{E}[\nabla \cdot K_{\phi}(Z_t)] + \partial_t \mathbb{E}[\Phi(Z_t)] \right] \right) dt \]
Nonparametric regression $\phi = \sum_{i=1}^{n} c_i \phi_i \in \mathcal{H}_n$:

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

- Choice of $\mathcal{H}_n$ & function space of learning?
  - Exploration measure $\rho_T \leftarrow |X_t - X'_t|$ 
- Inverse problem well-posedness/identifiability?
  - $\arg \min_{\phi \in L^2(\rho)} \mathcal{E}(\phi)$ 
- Convergence and rate? $\Delta x = M^{-1/d} \to 0$
Identifiability

\[ \mathcal{E}(\phi) = \langle L_G \phi, \phi \rangle - 2\langle \phi^D, \phi \rangle + \text{const}. \]

\[ \nabla \mathcal{E}(\phi) = L_G \phi - \phi^D = 0 \quad \Rightarrow \hat{\phi} = L_G^{-1} \phi^D \]

- **Identifiability**: \( A^{-1} b \leftrightarrow L_G^{-1} \phi^D \)
  - \( L_G \): positive compact operator
  - Function space of identifiability (FSOI): \( \text{span}\{\psi_i\}_{\lambda_i > 0} \)

- Coercivity condition on \( \mathcal{H} \) (not \( L^2(\rho) \))

\[ c_{\mathcal{H}} = \inf_{\phi \in \mathcal{H}, \|\phi\|_{L^2(\rho_T)} = 1} \langle L_G \phi, \phi \rangle > 0 \]
Convergence rate

Theorem (Numerical error bound [Lang-Lu20])

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

$$\|\hat{\phi}_{n,M} - \phi\|_{L^2(\rho_T)} \lesssim (\Delta x)^{\alpha s/(s+1)}$$

- $\Delta x^\alpha$ 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
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 $\phi$ discontinuous,
  - low rate for singular $\phi$

Other examples:
- suboptimal when $\phi$ discontinuous,
- low rate for singular $\phi$
Part 3: Learning kernels in operators

Regularization
Learning kernels in operators

Learn the kernel $\phi$:

$$R_\phi[u] = f$$

from data:

$$\mathcal{D} = \{(u_k, f_k)\}_{k=1}^{N}, \quad (u_k, f_k) \in \mathbb{X} \times \mathbb{Y}$$

- $R_\phi$ linear/nonlinear in $u$, but linear in $\phi$

- Examples:
  - interaction kernel: $R_\phi[u] = \nabla \cdot [u(K_\phi \ast u)] = \partial_t u - \nu \Delta u$
  - Toeplitz/Hankel matrix
  - integral/nonlocal operators,...
Ill-posed inverse problem

\[ \mathcal{E}(\phi) = \| R_\phi [u] - f \|_Y^2 \]

\[ \nabla \mathcal{E}(\phi) = L_G \phi - \phi^D = 0 \quad \Rightarrow \hat{\phi} = L_G^{-1} \phi^D \]

Regularization

\[ \mathcal{E}_\lambda(\phi) = \mathcal{E}(\phi) + \lambda \| \psi \|_Q^2 \quad \rightarrow \hat{\phi} = (L_G + \lambda Q)^{-1} \phi^D \]

- \( \lambda \) by the L-curve method [Hansen00]
- Regularization norm \( \| \cdot \|_Q? \) \( Q = Id, Q = RKHS? \)
Ill-posed inverse problem

\[ \mathcal{E}(\phi) = \| R_\phi[u] - f \|_Y^2 \]

\[ \nabla \mathcal{E}(\phi) = L_G \phi - \phi^D = 0 \quad \Rightarrow \quad \hat{\phi} = L_G^{-1} \phi^D \]

Regularization

\[ \mathcal{E}_\lambda(\phi) = \mathcal{E}(\phi) + \lambda \| \psi \|_Q^2 \rightarrow \hat{\phi} = (L_G + \lambda Q)^{-1} \phi^D \]

- \( \lambda \) by the L-curve method [Hansen00]
- Regularization norm \( \| \cdot \|_Q \)? \( Q = Id, Q = RKHS? \)

Data Adaptive RKHS Tikhonov Regularization [Lu+Lang+An22]

- norm of RKHS \( H_G = L_G^{1/2} L^2(\rho) \leftrightarrow Q = L_G^{-1} \)
- \( L_G \) is data dependent
- Computation: \( \hat{\phi} = (L_G + \lambda L_G^{-1})^{-1} \phi^D = (L_G^2 + \lambda I)^{-1} L_G \phi^D \)
DARTR: Data Adaptive RKHS Tikhonov Regularization

\[ R_\phi[u] = \nabla \cdot [u(K_\phi \ast u)] = f \]

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

**Convergence Rates**

Typical estimators, \( \Delta x = 0.05 \)

**Convergence of Estimators, nsr = 0.1 & 1**

**Loss value**

Sine kernel

Gaussian kernel

**Convergence Rates**

\( \Delta x = 0.0125 \times \{1,2,4,8,16\} \)
Small noise limit:

- $Q = I$: divergent estimator
- $Q = L_G^{-1}$: stable/convergent

### Discretization
- $L^2(\eta)$ error, continuous A

### Model error
- $L^2(\eta)$ error, continuous A

### Partial observation
- $L^2(\eta)$ error, continuous A

### Wrong noise
- $L^2(\eta)$ error, continuous A

---

$\sigma_\eta$

- Fixed prior, IQR
- Fixed prior, median
- DA prior, IQR
- DA prior, median
Summary and future directions

Nonparametric regression for interaction kernels
- Finite N (ODEs/SDEs): statistical learning
- $N = \infty$ (Mean-field PDEs): inverse problem

Learning kernels in operators:
- Probabilistic loss functionals
- Identifiability: $\hat{\phi} = L_G^{-1} \phi^D$
- Coercivity condition
  - yes: convergence
  - no: regularization — DARTR (ill-posed inverse problem)
Learning with nonlocal dependence: a new direction?
- Coercivity condition, spectrum decay
- Regularization for NN in function space?
- Convergence (minimax rate)

Classical learning
\[ \{(x_i, \phi(x_i) + \epsilon_i)\} \]

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

**Local dependence**
- Inversion: \( \hat{\phi} = I^{-1}\phi^D \)
- Regularization: \( \hat{\phi} = (I + \lambda Q)^{-1}\phi^D \)

**Nonlocal dependence**
- Values are undetermined from data
- \( \hat{\phi} = L_G^{-1}\phi^D \)
- \( \hat{\phi} = (L_G + \lambda L_G^{-1})^{-1}\phi^D \)
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
- Chada, Lang, Lu, Wang: A data-adaptive prior for Bayesian learning of kernels in operators. arXiv2212
- F.Lu, Q .Lang and Q. An. Data adaptive RKHS Tikhonov regularization for learning kernels in operators. MSML22
- F. Lu, M. Maggioni and S. Tang: Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. JMLR21