Adaptive Kernel Learning in Interacting Particle Systems and their Simulations

James Guo and Tina Shen **Advisor:** Dr. Xiong Wang

Contact Information:
Department of Mathematics,
Department of Applied Mathematics and Statistics
Email: {sguo45,xshen43}@jhu.edu

JOHNS HOPKINS UNIVERSITY

Abstract

We introduce an adaptive coefficient learning strategy for the non-parametric estimation of the radial interaction kernels in interacting particle systems (IPS), which can be modeled by stochastic differential equations (SDEs). These systems are fundamental in various physical and biological fields, where we typically don't know the underlying interactive system. Here, we demonstrate the approach with Lennard-Jones kernel on particle system simulation, and the unknown kernel is projected onto orthogonal basis functions, with coefficients initially estimated using a Least Squares Estimator (LSE). Our adaptive learning procedure refines the basis by strategically eliminating less significant coefficients, optimizing selection for more influential basis functions. We present numerical results from 2D simulations, demonstrating the efficacy of this kernel learning approach and discussing its performance with various basis sets.

Introduction

The goal is to learn the interacting kernel function inside a system.

Interacting Particle Systems (IPS)

The key characterization of IPS is that the motion of each particle is *influenced by its interaction with other particles*, which can be described by a kernel function $K : \mathbb{R}^d \to \mathbb{R}^d$, which often radial, denoted $\phi : \mathbb{R}^+ \to \mathbb{R}$. [3]

With kernel as ϕ, the particle system of the motion of the N particles $X(t) = \{X_i(t)\}_{i=1}^N$ can be thought of as a stochastic differential equation:

$$dX(t) = R_\phi[X(t)]dt + \gamma dB(t),$$

where $B(t) := \big(B_1(t), \cdots, B_N(t)\big)$ is a N-dimensional i.i.d. Brownian motion, and $R_\phi[X(t)]$ is the sum of all the forces:

$$R_\phi[X]_i = \frac{1}{N}\sum_{j\neq i} \phi\big(d(X_i, X_j)\big)\frac{X_i - X_j}{|X_i - X_j|} \qquad \text{for } i \in [N],$$

where γ is the strength of the noise (assume to be a constant).

Lennard Jones Interaction

Physically, one of the most well known interaction kernel is the Lennard Jones kernel.

Let two particles at $x, y \in \mathbb{R}^d$, with distance $r = \|x - y\|_{\mathbb{R}^d}$, their Lennard Jones interaction force is computed as:

$$F(r) := 24\epsilon\left[\left(\frac{\sigma}{r}\right)^6 - \left(\frac{\sigma}{r}\right)^{12}\right],$$

where ϵ is scaling factor and σ is the equilibrium distance. [2]

Note that the potential can become very large for small r. We apply a cutoff (F_{cutoff}) for numerical stability. Specifically:

$$\tilde{F}(x, y) = \max\left\{F_{\text{cutoff}}, \min\{F, -F_{\text{cutoff}}\}\right\},$$

which can be visualized as follows:

Graph of the Lennard Jones Force with distances between x and y.

Basis in $L^2([0, 2\pi])$

A key motivation of the learning procedure is based on $L^2([0, 2\pi])$ as a Hilbert space of functions. Hence, we can think of the bases:

- **Fourier Basis**: Consider $\{\exp(inx)\}_{n\in}$, which is equivalently:
$$\{1\} \cup \{\sin(nx), \cos(nx)\}_{n=1}^{\infty}.$$
- **Legendre Basis**: Apply the Gram Schmidt process (without normalization) for the polynomial basis $\{1, x, x^2, \cdots, x^n, \cdots\}$.
- **Characteristic Basis**: Consider a partition of the $[0, 2\pi]$ interval into n pieces $I_1, \cdots, I_n$ with sharing endpoints of $\{\mathbb{1}_{I_1}(x), \cdots, \mathbb{1}_{I_n}(x)\}$ where:
$$I_k = \left[\frac{2(k-1)\pi}{n}, \frac{2k\pi}{n}\right] \qquad \text{for } k = 1, 2, \cdots, n.$$

Methods

The primary goal to estimate the unknown interaction kernel from the observed trajectories of the particle system in following steps:

1. **Basis Expansion**: Assume the kernel can be approximated by a linear combination of M chosen basis functions $\{\psi_k(r)\}_{k=1}^M$, *i.e.*:
$$\phi_M(r) = \sum_{k=1}^{M} c_k\psi_k(r),$$
turning the task into estimating $c = (c_1, \cdots, c_M)^\intercal$.
2. **Simulation**: Simulate the trajectories by the given law and noise to obtain sets of discrete-time observations for particle positions.
3. **Least Squares Formulation**: Estimated c_k's by minimizing a least-squares objective function, which often leads to a linear system of equations $Ac = b$, based on the expectations involving the basis functions and the observed dynamics:
$$\hat{A}_{kl} = \frac{1}{NL}\sum_{i=1}^{N}\sum_{t=l}^{L}\langle R_{\psi_k}[X_t]_i, R_{\psi_l}[X_t]_i\rangle_{\mathbb{R}^d},$$
$$\hat{b}_k = \frac{1}{NL}\sum_{i=1}^{N}\sum_{t=l}^{L}\langle R_{\psi_k}[X_t]_i, Y_i(t)\rangle_{\mathbb{R}^d},$$
where N is the total number of particles, L is the total number of sampled time instances, X_{t_j} denotes the location configuration of all particles at time t_j, and $Y_i(t)$ is the observed interaction-driven velocity component for particle i at time t.

We aim to obtain an estimated kernel $\hat{\phi}_M(r) = \sum_{k=1}^M \hat{c}_k\psi_k(r)$ that accurately approximates the true kernel $\phi(r)$ During which we apply adaptations such as mixed basis and learning basis that abandons certain insignificant functions from the basis leaving space for more basis functions.

Criterion on Goodness of Fit

Based on the SDE, we can apply Girsanov Theorem [1] to define $\mathbb{Q}$ as another probability measure such that Radon–Nikodym derivative is:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left[-\int_0^T \frac{1}{\sigma}R_\phi[X(s)]dB(s) - \frac{1}{2}\int_0^T \frac{1}{\sigma^2}R_\phi^2[X(s)]ds\right].$$

Therefore, the Kullback-Leibler divergence ($\mathbb{KLD}$) using the change of measure formula is:

$$\mathbb{KLD}(\mathbb{Q} \parallel \mathbb{P}) = \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{2}\int_0^T \frac{1}{\sigma^2}R_\phi^2[X(s)]ds\right].$$

Eventually, we consider the trajectories, and obtain that:

$$\mathcal{E} \approx \frac{1}{MT}\sum_{m=1}^{M}\sum_{n=0}^{\lfloor T/\Delta t\rfloor}\|R_\phi[X(n\Delta t)] - Y(n\Delta t)\|^2\Delta t.$$

In particular, we think about the above as the approximation of the error especially when the ground truth is unknown. When we know the ground truth, we can think about the L^2 norm of the difference of the approximation and truth.

Results

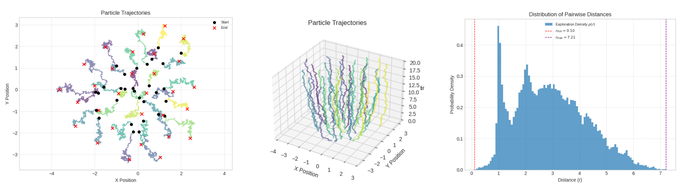
We elaborate on the qualitative and quantitative observations during the scope of the experiment.

Simulation Stage

The SDE was discretized using the Euler-Maruyama scheme [3]:

$$X_n(t_{j+1}) = X_n(t_j) + R_\phi[X(t_j)]_n\Delta t + \gamma\sqrt{\Delta t}Z_j,$$

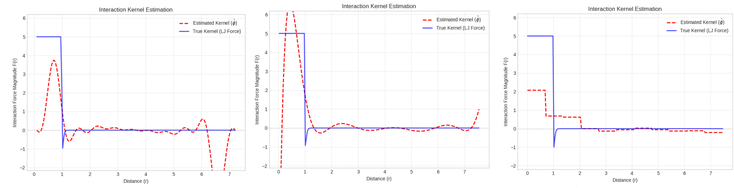
The simulation results lie as follows:



Sample paths and distance distributions of particles in a simulated system with parameters $\epsilon = 1$, $\sigma = 1$, $F_{\text{cutoff}} = 5$, $N = 30$, $T = 20$, and $\Delta t = 0.05$.

Estimation with 11 Basis

With the system, we used 11 basis for a sigle basis class, as follows:



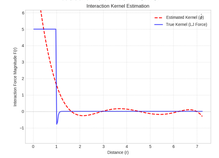
Fourier
Loss: 2.242197

Legendre
Loss: 1.913352

Characteristic
Loss: 1.644453

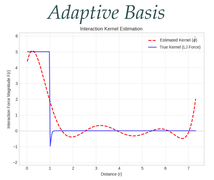
Adaptative Methods

Mixed Basis



Loss: 0.828474

Adaptive Basis



Loss: 0.465331

Basis results:

Here are the remaining bases left from each class of the mixed bases:

Type	Remainders
Fourier	[1,2,3,4,5]
Legendre	[1,2,3,4,5]
Char.	[1]

Discussion

For the simulation, we can observe certain remarks on it:

- The distribution of the distances between points has a large peak around $d(x, y) = 1$ since it an equilibrium between a pair of the points. There would be a smaller peak around the distance of 2 and gradually become less popular getting to two ends.
- An issue with the simulation is that it can only be done in terms of discrete time steps, not continuous time, so there were cases when the points suddenly got closer and exploded in terms of repulsion. We could only minimize the step size.
- One shall anticipate some sort of clustering in general, but with $\sigma = 1$ and all the points were not very spread out initially, the pattern is not very obvious. We shall test out more specific spread out parameters to check in future.

For the fitting, we shall initially observe that both bases have made a good attempt to correspond to the true kernel graphically. In particularity, the Lennard Jones force is exhibiting very bad shape, hence this set of basis is providing a relatively good result. Another note is that the learning itself is impacted by the noise:

- Recall that the SDE contains the noise part, the kernel was still relatively accurately captured, which was an indication that the model is relatively working well.
- From the result, we can see that the mixed basis exhibits a relatively good result, and the adaptive basis effectively truncated certain bases that are not efficient.
- For the LSE, we may also attempt to use Tamed LSE idea in [3], since Tamed LSE can address for the condition when the matrix in the LSE stage is ill-posed.

In terms of improvements, we can also anticipate to implement more:

- Since we have already lifted the normality, we can potentially lift some conditions with orthogonality by having some mixed basis to learn the model, as we want to achieve higher accuracy level.
- Given that the assumption here is that the noise is constant, the system could be extended to a more general one, *i.e.*, we may use $\gamma\big(X(t)\big)$ instead of a constant γ.

In general, this approach of learning kernel exhibits high potentials of learning various different systems, and could come out to be effective especially when we consider the kernel as interactions in actual physical models. It is a direct application of stochastic differential equations to systems that interact with radial kernels.

References

[1] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6 edition, 2003.

[2] Peter Schwerdtfeger, Antony Burrows, and Odile R. Smits. The lennard-jones potential revisited: Analytical expressions for vibrational effects in cubic and hexagonal close-packed lattices. *The Journal of Physical Chemistry A*, 125(14):3037–3057, 2021. PMID: 33787272.

[3] Xiong Wang, Inbar Seroussi, and Fei Lu. Optimal minimax rate of learning nonlocal interaction kernels, 2025.