Nonparametric online kernel inference for interacting particle systems

Fabio Nobile[1], **Grigorios A. Pavliotis**[2], and **Eliott Van Dieren**[1,2]

CSQI, Institute of Mathematics, EPFL, 1015 Lausanne, CH[1]
Department of Mathematics, Imperial College London, SW7 2AZ, UK[2]
{fabio.nobile, eliott.vandieren}@epfl.ch, g.pavliotis@imperial.ac.uk

1. Interacting Particle System setting and Goals

Let $X_t = (X_t^{(i)})_{i=1}^N \in \mathbb{R}^N$ be a set of N particles, with $t \in [0, T]$. Their interacting dynamics follow the SDEs

$$dX_t^{(i)} = v\left(X_t^{(i)}\right)dt + \frac{1}{N}\sum_{n=1}^{N} \phi\left(X_t^{(i)} - X_t^{(n)}\right)dt + \sigma dB_t^{(i)} \quad i = 1, ..., N, \tag{1}$$

where $X_0^{(i)} \sim \mu_0$ for all $i \in \{1, ..., N\}$, the function $\phi : \mathbb{R} \to \mathbb{R}$ is the interaction kernel and $v : \mathbb{R} \to \mathbb{R}$ is the drift function. We also write $\phi = -W'$, where W is the interaction potential. $(B_t^{(i)})_{t\geq 0}$ is a standard Brownian motion for $i \in \{1, ..., N\}$, and $\sigma \in \mathbb{R}$ is constant.

The main goal is to infer ϕ from (1) nonparametrically using data $(X_t)_{t\geq 0}$.
We want to generalise kernel inference to obtain an **online framework** given as

Trajectories of X_t ⟶ $\boxed{\text{Algorithm}}$ ⟶ Approximated function ϕ

2. Stochastic Gradient Descent in Continuous Time

The Stochastic Gradient Descent in Continuous Time (SGDCT) is an online statistical inference methodology [2]. Let $X_t \in \mathbb{R}^d$ be a given diffusion process following the stochastic differential equation

$$dX_t = f^*(X_t)dt + \sigma dB_t, \quad t \in [0, T], \tag{2}$$

where $B_t \in \mathbb{R}^d$ is a standard Brownian motion, $\sigma \in \mathbb{R}^{d\times d}$. The role of the SGDCT is to find $\theta \in \mathbb{R}^J$ s.t. $f(x, \theta) \approx f^*(x)$. The SDE describing the dynamics of the parameter estimates is given as

$$d\theta_t = l_t \nabla_\theta f(X_t, \theta_t)(\sigma\sigma^\top)^{-1}[dX_t - f(X_t, \theta_t)dt], \tag{3}$$

where l_t is the learning rate and $\nabla_\theta f(X_t, \theta_t) \in \mathbb{R}^{J\times d}$. It is shown in [3] that under some smoothness assumptions

$$\mathbb{E}[\|\theta_t - \theta^*\|^p] \leq \frac{\mathcal{K}}{(C_0 + t)^{p/2}}.$$

3. Methodology

One can rewrite (1) as (2) by considering X_t as a N-dimensional diffusion process and setting

$$f^*(X_t) = \begin{pmatrix} v(X_t^{(1)}) + \frac{1}{N}\sum_{n=1}^N \phi(X_t^{(1)} - X_t^{(n)}) \\ \vdots \\ v(X_t^{(N)}) + \frac{1}{N}\sum_{n=1}^N \phi(X_t^{(N)} - X_t^{(n)}) \end{pmatrix}. \tag{4}$$

We have access to discretized measurements of $(X_t^{(i)})_{t\geq 0}$ for $i = 1, ..., N$. We work with domains $[0, 2\pi]$ and $\mathbb{R}$. The Fourier series representation is an orthonormal basis for the circle, while we use Hermite expansions for the real line.

On the circle: We will infer ϕ using the **truncated Fourier series representation of its related potential W**, such that

$$W(x) \approx K + \sum_{j=1}^{J} w_j \cos(jx).$$

On the real line: Here, we use the **truncated Hermite expansion representation of its related potential W**, such that

$$W(x) \approx K + \sum_{j=1}^{J} w_j H_j(x),$$

where H_j is the j-th order Hermite polynomial.

3.1 Inference algorithm

Hence, we will use the SGDCT to infer the weights $\theta := \{w_j\}_{j=1}^J$ for both cases.

Algorithm 1 Stochastic Gradient Descent for Interacting Particle Systems

Input: $((X_k)_{k=0}^K)_{n=1}^{n_{\text{steps}}}$, μ_0^θ, C_0, C **Output:** θ_t for $t \in \{0, \Delta t, ..., K\}$.

1: **for** $n = 1, ..., n_{\text{steps}}$ **do**
2: Generate $\theta_0^n \sim \mu_0^\theta$
3: Set $X = ((X_k)_{k=0}^K)_n$, the n-th stream of observations
4: **for** $k = \Delta t, ..., T$ **do**
5: Set $l_k = C/(C_0 + k)$
6: Set $dX_k = X_k - X_{k-1}$
7: Compute the update step

$$\theta_k^n = \theta_{k-1}^n + l_k \sigma^{-2} \nabla_\theta f(X_{k-1}, \theta_{k-1}^n)[dX_k - f(X_{k-1}, \theta_{k-1}^n)dt]$$

8: **end for**
9: **end for**
10: **returns** $\theta_t = \frac{1}{n_{\text{steps}}}\sum_{n=1}^{n_{\text{steps}}} \theta_t^n$

The algorithm has a time complexity of $\mathcal{O}(N^2 J K n_{\text{steps}})$ where $K = T/dt$.

3.2 Adaptive SGDCT

An adaptive algorithm is required if one does not know J beforehand. One starts with a given truncation number J and increases it until the relative L^2 error between the two approximations is small enough.

Algorithm 2 Adaptive SGDCT algorithm

Input: $((X_k)_{k=0}^K)_{n=1}^{n_{\text{steps}}}$, tol, J **Output:** $\theta_t^{J^*}$ for $t \in \{0, \Delta t, ..., K\}$.

1: Compute W^J and related θ_t^J with SGDCT (3)
2: err = ∞
3: **while** err $\geq$ tol **do**
4: $J = J + 1$
5: Compute W^J and related θ_t^J with SGDCT (3)
6: err $= \|W^J - W^{J-1}\|_2$
7: **end while**
8: $J^* = J - 1$
9: **returns** θ^{J^*}

4. Numerical results

We define the Curie-Weiss quadratic interaction kernel [1] as

$$W(x, \kappa) = \frac{\kappa}{2}x^2,$$

and couple it with a quadratic Ornstein-Uhlenbeck confining potential, which yields the system

$$dX_t^{(i)} = -X_t^{(i)} - \kappa\left(X_t^{(i)} - \overline{X}_t\right)dt + dB_t^{(i)}, \quad i = 1, ..., N, \tag{5}$$

where $\overline{(\cdot)}$ is the empirical mean operator. For the SGDCT method, we have the estimated function

$$f(X_t, \theta) = \begin{pmatrix} -X_t^{(1)} + \frac{1}{N}\sum_{n=1}^N \sum_{j=1}^J 2jw_j H_{j-1}\left(X_t^{(1)} - X_t^{(n)}\right) \\ \vdots \\ -X_t^{(N)} + \frac{1}{N}\sum_{n=1}^N \sum_{j=1}^J 2jw_j H_{j-1}\left(X_t^{(N)} - X_t^{(n)}\right) \end{pmatrix}, \tag{6}$$

and the matrix

$$\nabla_\theta f(X_t, \theta) = \begin{pmatrix} \frac{1}{N}\sum_{n=1}^N 2H_0\left(X_t^{(1)} - X_t^{(n)}\right) & \dots & \frac{1}{N}\sum_{n=1}^N 2H_0\left(X_t^{(N)} - X_t^{(n)}\right) \\ \vdots & \ddots & \vdots \\ \frac{1}{N}\sum_{n=1}^N 2JH_{J-1}\left(X_t^{(1)} - X_t^{(n)}\right) & \dots & \frac{1}{N}\sum_{n=1}^N 2JH_{J-1}\left(X_t^{(N)} - X_t^{(n)}\right) \end{pmatrix}. \tag{7}$$

We know that $w = \{0, \frac{\kappa}{8}\}$, with $T = 5 * 10^4$, $\Delta t = 0.01$, $C = 10$, $C_0 = 100$, $J = 2$, $N = 200$, $n_{\text{steps}} = 1$, we observe convergence to the true weights as shown in Figure 1. Only two particles have been used out of the 200, leveraging the propagation of chaos phenomenon.

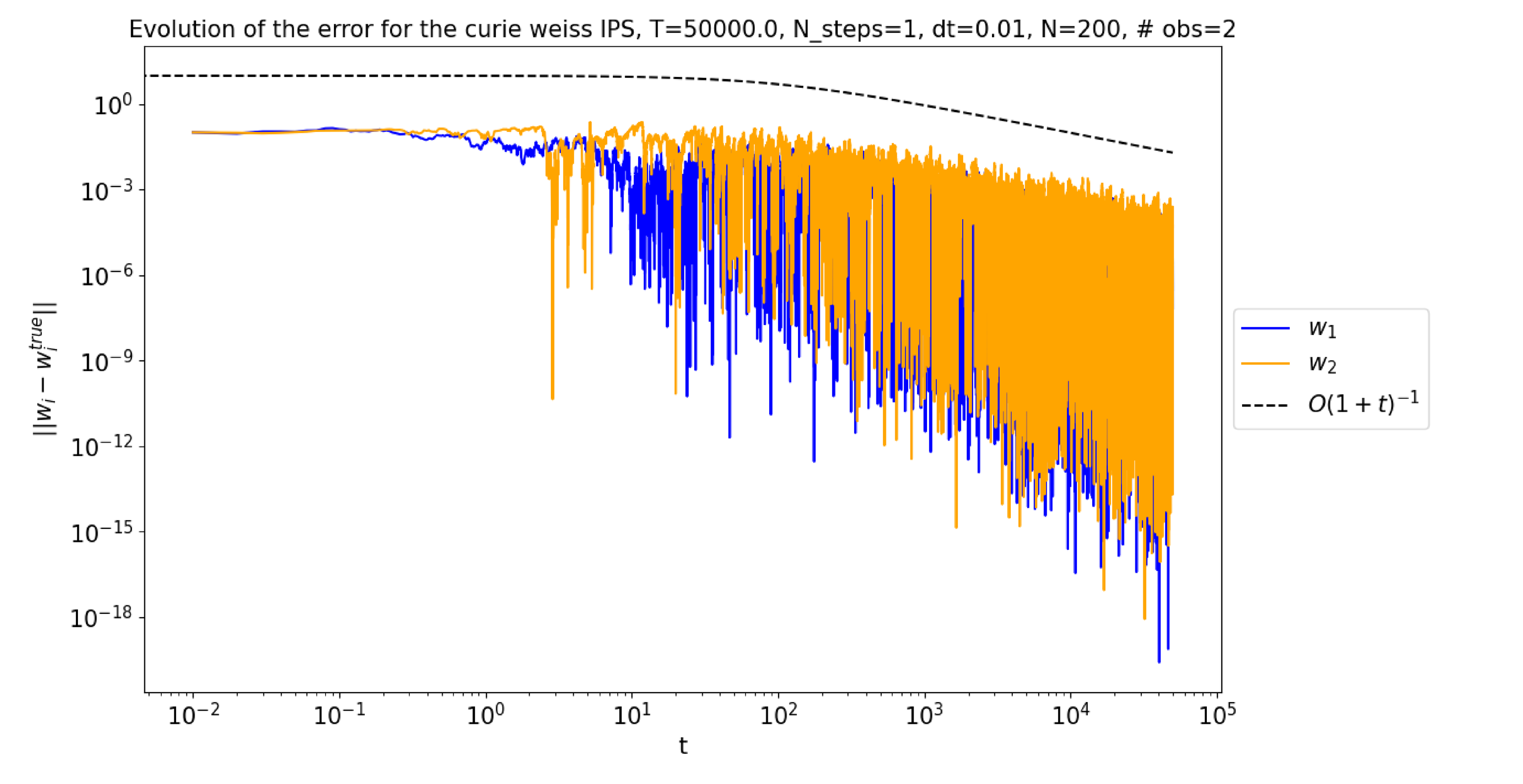

Figure 1: Error rates for the Curie-Weiss interaction potential. The errors follow the theoretical rate of $\mathcal{O}(1+t)^{-1}$. We use the MSE for error computation. Only two particles are observed out of 200.

5. Future directions

- Explore how the methodology behaves for high-dimensional particles.
- Infer both the interaction and confining potential in parallel.
- Make the method more robust to measurement noise.

References

[1] D. Dawson. "Critical dynamics and fluctuations for a mean-field model of cooperative behavior". In: *Journal of Statistical Physics* 31.1 (1983), pp. 29–85.

[2] J. Sirignano and K. Spiliopoulos. "Stochastic gradient descent in continuous time". In: *SIAM Journal on Financial Mathematics* 8.1 (2017), pp. 933–961.

[3] J. Sirignano and K. Spiliopoulos. "Stochastic gradient descent in continuous time: A central limit theorem". In: *Stochastic Systems* 10.2 (2020), pp. 124–151.

Notably presented at Numerical Methods for Random Differential Models, EPFL, June 11-15th, 2024.