Statistical methods for interacting particle systems

Mark Podolskij (University of Luxembourg)

join work with D. Belomestny and S.-Y. Zhou

Outline of the talk

- A high dimensional system and propagation of chaos
- Nonparametric estimation of a McKean-Vlasov SDE
- Minimax lower bounds

Model

- In this talk we consider a high dimensional interacting particle system described by the SDE

$$X_t^i = X_0^i + \Big(\frac{1}{N}\sum_{j=1}^{N}\int_0^t \varphi(X_s^i - X_s^j)\Big)ds + W_t^i, \qquad i = 1, \dots, N$$

$$= X_0^i + \int_0^t \varphi \star \mu_s^N(X_s^i)ds + W_t^i, \qquad \mu_t^N := \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^j},$$

  where (W^i) are independent Brownian motions and $X_0 \sim \mu_0^{\otimes N}$. The main goal is to estimate the *interaction function* φ given complete path observations $(X_t^i)_{t\in[0,T]}$ with $N \to \infty$ and $T > 0$ being fixed.

- Note that we face an **inverse problem** with **unknown** noise distribution.

Propagation of chaos in a nutshell

- When the interaction function φ is bounded and Lipschitz:

$$\mu_t^N \to \mu_t \text{ as } N \to \infty \qquad \text{(in Wasserstein distance)}$$

  for any $t \in [0, T]$.

- As a consequence, $X_t^i \to \overline{X}_t^i$ in $L^2(\Omega)$ and the corresponding *mean field system* is given by

$$\overline{X}_t^i = X_0^i + \int_0^t (\varphi \star \mu_s)(\overline{X}_s^i) ds + W_t^i \qquad \mu_t = \text{law}(\overline{X}_t^i),$$

  with $\varphi \star \mu_t$ denoting the convolution.

Applications

Interacting particle systems and mean field limits find numerous applications in sciences:

- Interaction of gas particles in statistical physics.
- Opinion dynamics in social systems.
- Mean field games in mathematical finance.
- Dynamics of gradient descent in wide neural networks.
- Attention modelling in large language models.

Method: Empirical risk minimisation

- Assume that we observe the data $(X_t^i)_{t\in[0,T]}$, $i = 1, \dots, N$, with $N \to \infty$ and $T > 0$ being fixed. We suppose that

$$|\varphi(x) - \varphi(y)| \leq L|x - y|, \qquad |\varphi(x)| \leq K.$$

- We consider a sequence of **compact** spaces $(S_m)_{m\geq 1}$ wrt $\|\cdot\|_\infty$. For a collection of functions $(\psi_j)_{j\in\mathbb{Z}}$ bounded by 1, a typical example is

$$S_m := \left\{ \sum_{|j|\leq m} a_j\psi_j : \ . \sum_{|j|\leq m} |a_j| \leq K \right\}$$

- Our main idea is based upon the minimisation problem:

$$\min_{f\in S_m} \frac{1}{NT} \sum_{i=1}^{N} \int_0^T \left(f \star \mu_t^N(X_t^i) - \varphi \star \mu_t^N(X_t^i) \right)^2 dt, \quad m = m(N)$$

  However, the above quantity is not feasible and needs to be approximated.

The estimation method

- The estimation procedure is based on the minimisation of the functional

$$\gamma_N(f) := \frac{1}{NT}\sum_{i=1}^{N}\left(\int_0^T f\star\mu_t^N(X_t^i)^2dt - 2\int_0^T f\star\mu_t^N(X_t^i)dX_t^i\right)$$

over S_m. In other words,

$$\varphi_N := \operatorname{argmin}_{f\in S_m}\gamma_N(f).$$

- We introduce bilinear forms:

$$\langle f,g\rangle_N := \frac{1}{NT}\sum_{i=1}^{N}\int_0^T (f\star\mu_t^N)(X_t^i)(g\star\mu_t^N)(X_t^i)\,dt,$$

$$\langle f,g\rangle_\star := \frac{1}{T}\int_0^T\int_{\mathbb{R}}(f\star\mu_t)(x)(g\star\mu_t)(x)\mu_t(x)\,dx\,dt.$$

where μ_t is the density of the limit $\overline{X}_t$.

Contractivity property of $\|\cdot\|_\star$

- It holds that

$$\|f\|_\star \leq \|f\|_{L^2(\mathbb{R})}$$

but the two norms are far from being equivalent.

- Example: Consider the function

$$f(x) = 1_{[M,2M]}(x)$$

and let μ_t be the $\mathcal{N}(0,1)$-distribution. Then $\|f\|^2_{L^2(\mathbb{R})} = M$ while

$$\|f\|^2_\star = \frac{1}{T}\int_0^T \int_{\mathbb{R}} f \star \mu_t(x)^2 \mu_t(x)\, dx\, dt \leq c_1 \exp(-c_2 M^2).$$

In particular, $\|f\|_{L^2(\mathbb{R})} \to \infty$ whereas $\|f\|_\star \to 0$ as $M \to \infty$.

A decomposition and a crucial inequality

We observe the decomposition

$$\|f-\varphi\|_N^2 = \gamma_N(f) + \gamma'_N(f) + \|\varphi\|_N^2,$$

where $\gamma'_N(f)$ represents the martingale term

$$\gamma'_N(f) := \frac{2}{NT}\sum_{i=1}^{N}\int_0^T (f \star \mu_t^N)(X_t^i)\, dW_t^i.$$

Lemma

It holds that

$$\|\varphi_N-\varphi\|_\star^2 \le \inf_{f\in S_m}\|f-\varphi\|_\star^2 + 2\left(\sup_{f\in S_m}\left|\|f-\varphi\|_N^2 - \|f-\varphi\|_\star^2\right| + \sup_{f\in S_m}|\gamma'_N(f)|\right).$$

Main statement

Theorem

Suppose that $\mathcal{N}\big(S_m, \|\cdot\|_\infty, \varepsilon\big) \lesssim \varepsilon^{-D_N}$ *and* $D_N^2 N^{-1/2} \to 0$ *as* $N \to \infty$. *Then it holds for any* $p \geq 2$*:*

$$\mathbb{E}\Big[\|\varphi_N - \varphi\|_\star^{2p}\Big]^{1/p} \leq \inf_{f \in S_m} \|f - \varphi\|_\star^2 + C_p \sqrt{\frac{D_N}{N}}.$$

Key ingredients:

(i) Use concentration inequalities for **U-statistics** to obtain convergence rates for **i.i.d. observations** $\overline{X}^i$.

(ii) Apply a **change of measure** device from Della Maestra and Hoffmann (22) to transfer the results to particle observations $X^{i,N}$

How large is the typical approximation error in $\|\cdot\|_\star$?

Approximation error

Consider the simple setting $\psi_j = 1_{[j,j+1)}$, $j \in \mathbb{Z}$, and define

$$S_m := \left\{ \sum_{|j| \le m} a_j \psi_j : \ . \sum_{|j| \le m} |a_j| \le K \right\}.$$

Lemma

Assume that

$$\varphi \in S := \left\{ \sum_{j \in \mathbb{Z}} a_j \psi_j : \ \sum_{j \in \mathbb{Z}} |a_j| \le K \right\}.$$

Consider the functional class S_m with $m \sim \sqrt{\log N}$. Then we deduce

$$\inf_{f \in S_m} \|f - \varphi\|_\star^2 \lesssim \frac{1}{N}$$

In particular,

$$\mathbb{E}\left[\|\varphi_N - \varphi\|_\star^{2p} \right]^{1/p} \lesssim \sqrt{\frac{\log N}{N}}$$

What about convergence rates in $\|\cdot\|_{L^2(\mathbb{R})}$?

An upper bound for $\|\cdot\|_{L^2(\mathbb{R})}$

Theorem

Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function in $L^1(\mathbb{R})$. There exist constants $m_R, \widehat{m}_R$ satisfying

$$m_R \geq c_1 \exp(-c_2 R^2), \qquad \widehat{m}_R \geq c_1 \exp(-c_2 R^2)$$

for $R > 0$, such that

$$\|f\|^2_{L^2(\mathbb{R})} \leq (m_{R_1} \widehat{m}_{R_2})^{-1}$$

$$\times \left(\|f\|^2_\star + \frac{m_{R_1}}{T} \int_0^T \int_{[-R_1,R_1]^c} \|f \star \mu_t(x)\|^2 dx dt \right)$$

$$+ \int_{[-R_2,R_2]^c} \|\widehat{f}(x)\|^2 dx,$$

for any $R_1, R_2 > 0$.

Convergence rate in $L^2(\mathbb{R})$

Theorem

Let S, S_m defined as before such that for all $f \in S \cup S_m$:

$$|f(x)| \leq C(1 \wedge |x|)^{-r_1}, \qquad |\widehat{f}(x)| \leq C(1 \wedge |x|)^{-r_2}.$$

Define $\log^{(2)}(x) := \log(\log(x))$. The it holds

$$\sup_{\varphi \in S} \mathbb{E}\left[\|\varphi_N - \varphi\|^2_{L^2(\mathbb{R})}\right] \lesssim \log(N)^{-r_1+1/2} + \left(\frac{\log(N)}{\log^{(2)}(N)}\right)^{-r_2+1/2}.$$

Minimax lower bound

Theorem

Let $\mathbb{P}_\varphi$ denote the law of the mean field equation $\overline{X}$ associated with the interaction function φ. Assume that $K \leq (32T)^{-1/2}$. Then there exists a constant $c > 0$ such that

$$\inf_{\widehat{\varphi}_N} \sup_{\varphi \in S} \mathbb{P}_\varphi^{\otimes N} \left(\|\widehat{\varphi}_N - \varphi\|^2_{L^2(\mathbb{R})} > c \left(\log(N)^{-r_1+1/2} + \log(N)^{-r_2+1/2} \right) \right) > 0$$

Happy Birthday, Yury!