Propagation of chaos for Hölder continuous interaction kernels via Glivenko-Cantelli

Thomas Holding

University of Warwick

19th April 2017

Outline

Consider N diffusing particles $X^i \in \mathbb{R}^d$ interacting with kernel W.

$$\begin{cases} dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N}, \quad i = 1, \dots, N, \\ b_t^N(x) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} W(x - X_t^{j,N}), \\ (X_0^{i,N})_{i=1}^N \text{ i.i.d. with law } f_0. \end{cases}$$

We expect that, for large N, the probability density of one of the particles is well approximated by the solution to the non-linear *mean-field* equation

$$\begin{cases} \partial_t f_t + \nabla \cdot (b_t^\infty f_t) - \frac{1}{2}\Delta f_t = 0, \qquad (t,x) \in (0,T) \times \mathbb{R}^d, \\ b_t^\infty(x) = \displaystyle\int f_t(y) W(x-y)\, dy, \\ f_0(x) \text{ initial condition.} \end{cases}$$

Assumptions on the interaction kernel W will be made later in the talk.

Why might one expect this? Heuristic:

- Imagine $b^N = b$ were fixed and given. Then by Itô's Lemma the law of $X_t^{1,N}$ solves

$$\begin{cases} \partial_t f_t + \nabla \cdot (b_t f_t) - \frac{1}{2}\Delta f_t = 0, & (t,x) \in (0,T) \times \mathbb{R}^d, \\ f_0(x) \text{ initial condition.} \end{cases}$$

- *Idea:* Make the assumption that the particles are i.i.d. (*chaos*), so we can compute the force field $b_t^N(x)$ by

$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^{N} W(x - X_t^{j,N}) \underbrace{\approx}_{LLN} \mathbb{E} W(x - X_t^{1,N})$$
$$= \int W(x-y) f_t(y)\, dy = b_t^\infty(x)$$

where we have used the law of large numbers to approximate the actual field b^N by the *mean field* b^∞.

- The assumption of *chaos* is only true for $t = 0$. We must assume *propagation of chaos*, that particles remain (approximately) independent at later times if they are independent at the start.

Let d be a metric on the space of probability measures on $\mathbb{R}^d$, e.g. Bounded-Lipschitz metric.

Definition (Chaotic particle system (Sznitman))

A family of symmetric particle distributions $(X^{i,N})_{i=1}^N$ for $N = 1, 2, \dots,$ is *chaotic* if the empirical measure μ^N given by

$$\mu^N := \frac{1}{N}\sum_{j=1}^N \delta_{X^{i,N}}, \qquad \text{satisfies} \qquad d(\mu^N, f) \to 0 \text{ weakly as } N \to \infty$$

for some deterministic probability measure f.

Note that there is no mention of time in this definition.

We are interested in *quantitative* estimates of chaoticity. We want bounds on $d(\mu^N, f)$ that are *polynomial* in N, i.e.

$$\mathbb{E}d(\mu^N, f) \leq CN^{-\gamma}, \qquad \text{for some explicit } \gamma > 0.$$

Why is propagation of chaos important?

Establishing propagation of chaos is a central part of rigorously deriving *macroscopic* and *mesoscopic* continuum models from the *microscopic* laws governing the motion of particles. It is used in:

- The Vlasov-Poisson equation in kinetic theory, which models galaxies and plasmas.
- The vortex dynamics formulation of the 2D Euler equation.
- Swarming models for fish, birds, ...
- In the derivation of the homogeneous Boltzmann equation from Kac's model of randomly colliding particles.
- In the particles method in numerical integration of PDE.
- In the theory of particle filters in statistics.
- In mean-field models of biological neural networks.
- Many more ...

Prior (quantitative) work (locally Lipschitz interactions)

- Dobrushin '79 (also Braun-Hepp independently), proved propagation of chaos in the noiseless case (ODEs instead of SDEs) for *Lipschitz* interaction kernels W.
  Key observation: empirical measure is weak solution to limit PDE.
- Sznitman '91, did the same for SDEs (explained later in talk).
- Since then the main quantitative results have assumed that W is smooth except for a singularity at the origin, and work by estimating the distance between particles: Jabin, Hauray, Mischler, Fournier ... in the last decade.
- In these works the noise is a hindrance. It makes it harder to control the minimum distance between particles.

(Non-quantitative results using compactness are a whole other game.)

The result

$$dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N},$$
$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^N W(x - X_t^{j,N}).$$

$$\partial_t f_t + \nabla \cdot (b_t^\infty f_t) - \tfrac{1}{2}\Delta f_t = 0,$$
$$b_t^\infty(x) = \int f_t(y) W(x-y)\,dy.$$

Theorem (H. 2016)

Let W be α-Hölder continuous for some $\alpha \in (0,1)$. Then there exists an explicit $\gamma > 0$ depending only on α, such that

$$\mathbb{E} \sup_{t\in[0,T]} d(\mu_t^N, f_t) \leq CN^{-\gamma}.$$

In particular, this holds for interaction kernels W that are nowhere Lipschitz.

This result implies the result of Sznitman (which requires W Lipschitz, i.e. $\alpha = 1$), but the exponent γ obtained is worse.

Coupling methods

$$dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N},$$

$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^{N} W(x - X_t^{j,N}).$$

$d(\mu_t^N, f_t)$

$$\partial_t f_t + \nabla \cdot (b_t^\infty f_t) - \tfrac{1}{2}\Delta f_t = 0,$$

$$b_t^\infty(x) = \int f_t(y) W(x - y)\, dy.$$

Coupling methods

$$dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N},$$

$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^N W(x - X_t^{j,N}).$$

Stability of SDE

$d(\mu_t^N, \mu_t^{b^\infty,N})$

$$dX_t^{b^\infty,i,N} = b_t^\infty(X_t^{b^\infty,i,N})dt + dB_t^{i,N}.$$

$d(\mu_t^N, f_t)$

$d(\mu_t^{b^\infty,N}, f_t)$

Law of Large numbers approximation

$$\partial_t f_t + \nabla\cdot(b_t^\infty f_t) - \tfrac{1}{2}\Delta f_t = 0,$$

$$b_t^\infty(x) = \int f_t(y)W(x-y)\,dy.$$

$$d(\mu_t^N, f_t) \leq d(\mu_t^N, \mu_t^{b^\infty,N}) + d(\mu_t^{b^\infty,N}, f_t) \qquad \text{(Sznitman)}$$

Coupling methods

$$dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N},$$
$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^N W(x - X_t^{j,N}).$$

Stability of SDE — $d(\mu_t^N, \mu_t^{b^\infty,N})$

$$dX_t^{b^\infty,i,N} = b_t^\infty(X_t^{b^\infty,i,N})dt + dB_t^{i,N}.$$

Law of Large numbers approximation — $d(\mu_t^N, f_t^{b^N})$

$d(\mu_t^N, f_t)$

$d(\mu_t^{b^\infty,N}, f_t)$ — Law of Large numbers approximation

$$\partial_t f_t^{b^N} + \nabla\cdot(b_t^N f_t^{b^N}) - \tfrac{1}{2}\Delta f_t^{b^N} = 0.$$

$d(f_t^{b^N}, f_t)$ — Stability of PDE

$$\partial_t f_t + \nabla\cdot(b_t^\infty f_t) - \tfrac{1}{2}\Delta f_t = 0,$$
$$b_t^\infty(x) = \int f_t(y) W(x-y)\, dy.$$

$$d(\mu_t^N, f_t) \le d(\mu_t^N, \mu_t^{b^\infty,N}) + d(\mu_t^{b^\infty,N}, f_t) \qquad \text{(Sznitman)}$$

OR

$$d(\mu_t^N, f_t) \le d(\mu_t^N, f_t^{b^N}) + d(f_t^{b^N}, f_t)$$

Estimating the hard term

$$dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N},$$
$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^{N} W(x - X_t^{j,N}).$$

Law of Large numbers?

$\longleftrightarrow$

$d(\mu_t^N, f_t^{b^N})$

$$\partial_t f_t^{b^N} + \nabla \cdot (b_t^N f_t^{b^N}) - \tfrac{1}{2}\Delta f_t^{b^N} = 0.$$

- If b^N were deterministic this would be an easy application of the law of large numbers.
- But b^N depends on all the particles.
- What *do* we know about b^N?

Estimating the hard term

$$dX_t^{i,N} = b_t^N(X_t^{i,N})dt + dB_t^{i,N},$$
$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^{N} W(x - X_t^{j,N}).$$

Law of Large numbers?

$\longleftrightarrow$

$d(\mu_t^N, f_t^{b^N})$

$$\partial_t f_t^{b^N} + \nabla \cdot (b_t^N f_t^{b^N}) - \tfrac{1}{2}\Delta f_t^{b^N} = 0.$$

- If b^N were deterministic this would be an easy application of the law of large numbers.
- But b^N depends on all the particles.
- What *do* we know about b^N?
- As W is α-Hölder continuous, b^N is α-Hölder in x, and $(\alpha/2 - \varepsilon)$-Hölder in t almost surely.
- We will estimate $\mathbb{E}\sup_{t\in[0,T]} d(\mu_t^N, f_t^{b^N})$ using *only* this information.

Estimating the hard term

- When you don't know something, bound with the worst case!
- As, $\mu^N = \mu^{b^N,N}$ by definition, we have

$$\mathbb{E} \sup_{t\in[0,T]} d(\mu_t^N, f_t^{b^N}) \leq \mathbb{E} \sup_{b\in\mathcal{C}} \sup_{t\in[0,T]} d(\mu_t^{b,N}, f_t^b).$$

$$dX_t^{b,i,N} = b_t(X_t^{b,i,N})dt + dB_t^{i,N},$$

Uniform Law of Large numbers

$$\longleftrightarrow \quad \sup_{b\in\mathcal{C}} d(\mu_t^{b,N}, f_t^b)$$

$$\partial_t f_t^b + \nabla\cdot(b_t f_t^b) - \frac{1}{2}\Delta f_t^b = 0.$$

$$\mathcal{C} = \{b : \|b\|_{C^{0,\alpha}([0,T]\times\mathbb{R}^d)} \leq C\}.$$

Estimating the hard term

- When you don't know something, bound with the worst case!
- As, $\mu^N = \mu^{b^N,N}$ by definition, we have

$$\mathbb{E} \sup_{t\in[0,T]} d(\mu_t^N, f_t^{b^N}) \leq \mathbb{E} \sup_{b\in\mathcal{C}} \sup_{t\in[0,T]} d(\mu_t^{b,N}, f_t^b).$$

$$dX_t^{b,i,N} = b_t(X_t^{b,i,N})dt + dB_t^{i,N},$$

Uniform Law of Large numbers

$$\longleftrightarrow$$

$$\sup_{b\in\mathcal{C}} d(\mu_t^{b,N}, f_t^b)$$

$$\partial_t f_t^b + \nabla\cdot(b_t f_t^b) - \tfrac{1}{2}\Delta f_t^b = 0.$$

$$\mathcal{C} = \{b : \|b\|_{C^{0,\alpha}([0,T]\times\mathbb{R}^d)} \leq C\}.$$

Theorem (Glivenko-Cantelli result for SDEs (H. 2016))

There exists explicit γ depending only on α such that

$$\mathbb{E} \sup_{b\in\mathcal{C}} \sup_{t\in[0,T]} d(\mu_t^{b,N}, f_t^b) \leq CN^{-\gamma}.$$

Proving the Glivenko-Cantelli result

Want to prove: $$\mathbb{E} \sup_{b\in\mathcal{C}} \sup_{t\in[0,T]} d(\mu_t^{b,N}, f_t^b) \leq CN^{-\gamma}.$$

- Uniform law of large numbers of a large set $\mathcal{C}$.
- Study one-particle stochastic process $(X_t^b)_{b\in\mathcal{C}}$.
- First step: Continuity with respect to $b \in \mathcal{C}$.
- $\mathcal{C}$ is too large to apply the Kolmogorov continuity theorem (or the chaining method in general!).

Proving the Glivenko-Cantelli result

Want to prove: $$\mathbb{E} \sup_{b \in \mathcal{C}} \sup_{t \in [0,T]} d(\mu_t^{b,N}, f_t^b) \leq C N^{-\gamma}.$$

- Uniform law of large numbers of a large set $\mathcal{C}$.
- Study one-particle stochastic process $(X_t^b)_{b \in \mathcal{C}}$.
- First step: Continuity with respect to $b \in \mathcal{C}$.
- $\mathcal{C}$ is too large to apply the Kolmogorov continuity theorem (or the chaining method in general!).
- Fortunately, we have a very strong continuity property:

Lemma

For each $b \in \mathcal{C}$, then there exists a random variable J (different for each b) in L^p such that for all $\tilde{b} \in \mathcal{C}$, we have,

$$\sup_{t \in [0,T]} |X_t^b - X_t^{\tilde{b}}| \leq J \int_0^T \left\| b_t - \tilde{b}_t \right\|_{L^\infty} dt.$$

With this lemma we can prove the Glivenko-Cantelli theorem using standard tools from *empirical process theory*:

Definition (Covering Number)

$\mathcal{A}$ a set, d a metric on $\mathcal{A}$:

$$N(\varepsilon, \mathcal{A}, d) = \text{Size of smallest } \varepsilon\text{-cover of } \mathcal{A}$$

where an ε-cover (ε-net) is a finite set $(a_k)_{k=1}^m \subset \mathcal{A}$ such that for each $a \in \mathcal{A}$ there is a k with $d(a, a_k) \leq \varepsilon$.

Then $N(\varepsilon, \mathcal{C}, \|\cdot\|_{L^\infty}) \leq \exp(C\varepsilon^{-(d+2)/\alpha})$ which is sharp (and huge!), but:

Lemma (Orlicz maximal inequality for sub-Gaussian r.v.s)

Let $Z_1, \ldots Z_m$ be real-valued sub-Gaussian r.v.s (not necessarily independent), and $\|\cdot\|_{\psi_2}$ be the sub-Gaussian Orlicz norm, then

$$\left\| \max_{k=1}^m |Z_k| \right\|_{\psi_2} \leq C\sqrt{\log(1+m)} \max_{k=1}^m \|Z_k\|_{\psi_2}.$$

Proving the lemma

Lemma

For each $b \in \mathcal{C}$, then there exists a random variable J (different for each b) in L^p such that for all $\tilde{b} \in \mathcal{C}$, the following holds almost surely,

$$\sup_{t\in[0,T]} |X_t^b - X_t^{\tilde{b}}| \leq J \int_0^T \left\| b_t - \tilde{b}_t \right\|_{L^\infty} dt.$$

- $\frac{d}{dt}|X_t^b - X_t^{\tilde{b}}| \leq |b(X_t^b) - \tilde{b}(X_t^{\tilde{b}})|$, but b not Lipschitz, so can't close estimate!

Proving the lemma

Lemma

For each $b \in \mathcal{C}$, then there exists a random variable J (different for each b) in L^p such that for all $\tilde{b} \in \mathcal{C}$, the following holds almost surely,

$$\sup_{t\in[0,T]} |X_t^b - X_t^{\tilde{b}}| \leq J \int_0^T \left\| b_t - \tilde{b}_t \right\|_{L^\infty} dt.$$

- $\frac{d}{dt}|X_t^b - X_t^{\tilde{b}}| \leq |b(X_t^b) - \tilde{b}(X_t^{\tilde{b}})|$, but b not Lipschitz, so can't close estimate!
- But *noise is regularising!*

Theorem (Stochastic flow. (Flandoli, Gubinelli, Priola '10))

The solution map $\phi : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $X_0 \mapsto X_t$ of the SDE $dX_t = b_t(X_t)dt + dB_t$ where $b \in \mathcal{C}$ (i.e. α-Hölder), is $C^{1,\beta}$ in x almost surely.

- Use this to replace the (absent) Lipschitz property of b in the above estimate.

Second order SDEs and hypoellipticity

Can do the same for second order particle system, but the degeneracy of the noise means we need 2/3-Hölder continuity.

$$dX_t^{i,N} = V_t^{i,N} dt$$
$$dV_t^{i,N} = b_t^N(X_t^{i,N})dt - V_t^{i,N} dt + dB_t^{i,N}$$
$$b_t^N(x) = \frac{1}{N}\sum_{j=1}^{N} W(x - X_t^{j,N})$$

(f_t solves a non-linear kinetic Fokker-Planck equation (omitted for space)).

Theorem (H. 2016)

Let W be α-Hölder continuous for some $\alpha \in (2/3, 1)$. Then there exists an explicit $\gamma > 0$ depending only on α, such that

$$\mathbb{E} \sup_{t\in[0,T]} d(\mu_t^N, f_t) \leq CN^{-\gamma}.$$

This holds for interaction kernels W that are nowhere Lipschitz.

Related work and conclusions

- The coupling method can also be used to prove propagation of chaos for networks of *leaky integrate and fire neurons*, which is a commonly used model in computational neuroscience. (Article still in preparation).
- The dynamics of a LIF neuron are not even well-posed in the sense of Hadamard: no continuous dependence on initial data! (and no regularisation by noise!)