BSDE2022: Summer school on propagation of chaos, lecture 2

Sunday, June 26, 2022 9:36 AM

Quantitative propagation of chaos

Lec 1:
1. Trajectorial / coupling
2. ~~Compactness~~
3. Semigroup analysis
   - Mischler-Mouhot-Wennberg
   - Kolokoltsov
   - Chassagneux-Szpruch-Tse + Bencheik-Jourdain
   - Delarue-Tse
   - Tomasevic

   $\mathbb{E}F(\mu_t^n) - F(\mu_t)$

   MV equation PDE: $\partial_t \mu_t = \cdots$

   $S_t F(m) = F(\mu_t)$ when μ solves MV starting from $\mu_0 = m$

   (μ_t^n) perturbation
4. Entropy methods ☆
   - Jabin-Wang JFA '16, Inventiones '18
   - Lacker
   - Jabir

I. Basics of relative entropy

Def: For prob. measures ν, μ (on same space)

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu}\, d\mu = \int \log \frac{d\nu}{d\mu}\, d\nu & \text{if } \nu \ll \mu \\ \infty & \text{else.} \end{cases}$$

- $H \geq 0$ by Jensen, $= 0$ iff $\nu = \mu$
- Pinsker's inequality:

$$\underbrace{\sup_{\substack{\varphi \text{ meas} \\ |\varphi| \leq 1}} |\langle \mu, \varphi \rangle - \langle \nu, \varphi \rangle|^2}_{= \|\mu - \nu\|_{TV}^2} \leq 2 H(\nu|\mu)$$

- Gibbs variational formula: For φ "nice"

$$\log \int_{\mathcal{X}} e^{\varphi}\, d\mu = \sup_{\nu \in \mathcal{P}(\mathcal{X})} \Big(\langle \nu, \varphi \rangle - H(\nu|\mu) \Big).$$

- <u>Chain rule</u>: For measures on product spaces $\mathcal{X} \times \mathcal{Y}$,

$$H\Big(\mu(dx) K_x(dy) \,\Big|\, \mu'(dx) K'_x(dy)\Big)$$

$\log xy = \log x + \log y$

(Dupuis - Ellis book)

$$= H(\mu|\mu') + \int_{\mathcal{X}} H(K_x|K'_x)\, \mu(dx).$$

- <u>Subadditivity</u>: Let $\mu_1 \in \mathcal{P}(\mathcal{X}_1)$, $\mu_2 \in \mathcal{P}(\mathcal{X}_2)$, $P \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$ with $P(dx, dy) = P_1(dx) K_x(dy)$.

$$H(P|\mu_1 \times \mu_2) = H(P_1|\mu_1) + \int_{\mathcal{X}_1} H(K_x|\mu_2)\, P_1(dx)$$

$\boxed{H(\cdot|\mu_2) \text{ is convex, lsc}}$

$$\geq H(P_1|\mu_1) + H\Big(\underbrace{\int_{\mathcal{X}} K_x\, P_1(dx)}_{\text{2nd marginal of } P, \text{ say } P_2} \,\Big|\, \mu_2\Big)$$

$$= H(P_1|\mu_1) + H(P_2|\mu_2)$$

$$= 2H(P_1|\mu)$$

if $\mathcal{X}_1 = \mathcal{X}_2$, $\underbrace{P \text{ exchangeable}}_{P_1 = P_2}$, $\mu_1 = \mu_2 = \mu$

<u>Generally</u>: If $P \in \mathcal{P}(\mathcal{X}^n)$ is <u>exchangeable</u>, $P_k \in \mathcal{P}(\mathcal{X}^k)$ is k-particle marginal, and $\mu \in \mathcal{P}(\mathcal{X})$, then

$$H(P_k \mid \mu^{\otimes k}) \le \begin{cases} \frac{k}{n} H(P \mid \mu^{\otimes n}) & \text{if } \frac{n}{k} \in \mathbb{N} \\ \frac{1}{\lfloor \frac{n}{k} \rfloor} H(P/\mu^{\otimes n}) \le \frac{2k}{n} H(P/\mu^{\otimes n}) & \text{else.} \end{cases}$$

E.g. if $\frac{1}{n}\underbrace{H(P/\mu^{\otimes n})}_{O(1)} \to 0$ then $\underbrace{H(P_k/\mu^{\otimes k})}_{O(\frac{k}{n})} \to 0$ $\forall k$ fixed.

"increasing prop. chaos"
Ben Arous - Zeitouni

- **Diffusions**: Let Z^i, $i=1,2$ in $\mathbb{R}^d$ solve SDEs

$$dZ^i_t = B_i(t, Z^i)dt + dW^i_t$$

where B_i prog. measurable, bounded.

see e.g.
Lacker: "Hierarchies..."
Lem 4.3/4

$$H\Big(L(Z^1[0,t]) \,\Big|\, L(Z^2[0,t]) \Big) = H\big(L(Z^1_0)/L(Z^2_0)\big) + \frac{1}{2}\mathbb{E}\int_0^t |B_1(s,Z^1) - B_2(s,Z^1)|^2 ds.$$

Time-marginal version: $H(L(Z^1_t) \mid L(Z^2_t))$
- uniform in time
- singular interaction

see e.g.
L.-Le Flem '22
"sharp uniform..."
Lemma 3.1

II. Global Entropy Method (Jabin-Wang)

Pairwise interactions; Particles $i=1,2,\dots,n$. $b: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$

Pairwise interactions: Particles $i = 1, 2, \dots, n$, $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ bded measurable

$$dX_t^i = \frac{1}{n-1} \sum_{\substack{j \neq i \\ j=1}}^n b(X_t^i, X_t^j)\, dt + dW_t^i.$$

Assume $(X_0^1, \dots, X_0^n)$ exchangeable.

- Let $P_{[0,t]}^n = \mathrm{Law}(X^1, \dots, X^n)$ on $C([0,t]; \mathbb{R}^d)$, and $P_{[0,t]}^{n,k} = \mathrm{Law}(X^1, \dots, X^k)$.
- MV: $dY_t^i = \langle \mu_t, b(Y_t^i, \cdot) \rangle dt + dW_t^i$; $\mu = \mathcal{L}(Y^i)$.

Let $H_t^{n,k} = H\left(P_{[0,t]}^{n,k} \,\middle|\, \mu_{[0,t]}^{\otimes k}\right)$, $1 \le k \le n$.

Propagation of chaos would follow if $H_t^{n,k} \to 0$ as $n \to \infty$, $\forall k$.

Goal of global approach: Bound $H_t^{n,n} = O(1)$.

Subadditivity $\Rightarrow H_t^{n,k} \le \frac{2k}{n} H_t^{n,n} = O\left(\frac{k}{n}\right)$.

cannot be improved!

later: improve to $O\left(\left(\frac{k}{n}\right)^2\right)$

Sketch: $\left[Z^1 = (X^1, \dots, X^n),\ Z^2 = (Y^1, \dots, Y^n) \right]$

$$\frac{d}{dt} H_t^n = \frac{1}{2} \mathbb{E} \sum_{i=1}^n \left| \frac{1}{n-1} \sum_{\substack{j \neq i \\ j=1}}^n b(X_t^i, X_t^j) - \langle \mu_t, b(X_t^i, \cdot) \rangle \right|^2$$

$$= \frac{n}{2} \mathbb{E} \left| \frac{1}{n-1} \sum_{j=2}^n \Big(\underbrace{b(X_t^1, X_t^j) - \langle \mu_t, b(X_t^1, \cdot) \rangle}_{\bar{b}_t(X_t^1, X_t^j)} \Big) \right|^2$$

$$= \frac{n}{2(n-1)^2} \left(\mathbb{E} \sum_{j=2}^n |\bar{b}_t(X_t^1, X_t^j)|^2 + \mathbb{E} \sum_{j=2}^n \sum_{\substack{k=2 \\ k \neq j}}^n \bar{b}_t(X_t^1, X_t^j) \cdot \bar{b}_t(X_t^1, X_t^k) \right)$$

$n \le 2(n-1)$

$\le O(1) + ($ ...

Side note: $H_t^n = O(n)$ is easy but

$H_t^n = O(n)$ is easy but useless

$$\le O(1) + \underbrace{\int_{(\mathbb{R}^d)^n} \frac{1}{n-1} \sum_{\substack{j,k=2 \\ j\neq k}}^{n} b_t(x_1, x_j)\cdot b_t(x_1, x_k)\, P_t^n(dx)}_{\star}$$

Gibbs $\Rightarrow$ $\forall a>0$ TBD, $\forall \varphi$,

$$\int \varphi \, dP^n \le \frac{1}{a}\left(H(P|\mu^{\otimes n}) + \log \int e^{a\varphi} d\mu^{\otimes n} \right)$$

Then

$$\star \le \frac{1}{a} H_t^n + \frac{1}{a} \log \underbrace{\int_{(\mathbb{R}^d)^n} \exp\left(\frac{a}{n-1} \sum_{\substack{j,k=2 \\ j\neq k}}^{n} b_t(x_1,x_j)\cdot b_t(x_1,x_k)\right) \mu_t^{\otimes n}(dx)}_{\star\star}$$

Goal now: Show $\star\star = O(1)$, for $a>0$ small.

This would lead to $\frac{d}{dt} H_t^n = \frac{1}{a} H_t^n + O(1)$

$$\Rightarrow \sup_{t\in[0,T]} H_t^n = O(1) \quad \forall T>0.$$

$\boxed{\text{assume } H_0^n = H(P_0^n|\mu_0^{\otimes n}) = O(1)}$

Study $\star\star$ via large deviations ideas.

Fix x_1, t. Let $f(y,z) = b_t(x_1,y)\cdot b_t(x_1,z)$.

Study

$$\int_{(\mathbb{R}^d)^{n-1}} \exp\left(\frac{a}{n-1} \sum_{\substack{j,k=2 \\ j\neq k}}^{n} f(x_j, x_k)\right) \mu_t^{\otimes(n-1)}(dx_2, \dots, dx_n)$$

or

$$\mathbb{E}\left[\exp\left(\underbrace{\frac{a}{n-1} \sum_{\substack{j,k=2 \\ j\neq k}}^{n} f(Y_j, Y_k)}\right)\right] \qquad Y_j \sim \mu_t \text{ iid}.$$

$\simeq an \langle L_n(Y)^{\otimes 2}, f \rangle$

- If f continuous, can get $\log(\cancel{\quad}) = o(n)$ via Sanov/Varadhan
- If f just bounded, needs work.

Jabin-Wang / Lim-Lu-Nolen Lemma 4.3

III. Local entropy methods

L. '21 "Hierarchies..."

Idea: Develop $\frac{d}{dt} H_t^k = \frac{d}{dt} H(P^{n,k}_{[0,t]} \mid \mu^{\otimes k}[0,t])$, $k<n$.

Need self-contained dynamics for $X^1, \dots, X^k$,

i.e. drift depending only on $(t, X^1, \dots, X^k)$.

Filtering lemma: $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}) \xrightarrow{\varphi} W$

Let $dZ_t = \varphi_t dt + dW_t$ be some Itô process.

φ_t: $\mathbb{F}$-adapted but not nec. given as a function of (t, Z).

Let $\hat{\varphi}(t, Z) = \mathbb{E}[\varphi_t \mid \mathcal{F}_t^Z]$ be opt. projection. $\mathcal{F}_t^Z = \sigma(Z_s : s \le t)$.

Then $dZ_t = \hat{\varphi}(t, Z)dt + d\widehat{W}_t$ for some $\mathbb{F}^Z$-Brownian motion.

Proof: Show

$$\widehat{W}_t = Z_t - Z_0 - \int_0^t \hat{\varphi}(s, Z)ds$$
$$= W_t + \int_0^t (\varphi_s - \hat{\varphi}(s, Z))ds$$

is $\mathbb{F}^Z$-BM. using Levy's criterion.

Recall: $dX_t^i = \frac{1}{n-1} \sum_{j=1}^{n} b(X_t^i, X_t^j)dt + dW_t^i$ $\quad i=1,\dots,n$

Recall: $dX_t^i = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n} b(X_t^i, X_t^j)\,dt + dW_t^i \qquad i=1,\dots,n$

Idea: Use $Z=(X^1,\dots,X^k)$ in lemma.

Let $1\le k\le n$. Let $\mathcal{F}_t^k = \sigma(X_s^1,\dots,X_s^k : s\le t)$.

Then, for $1\le i\le k$,

$$\hat{b}_i^k(t,x^1,\dots,x^k) := \mathbb{E}\left[\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n} b(X_t^i,X_t^j)\,\Big|\,\mathcal{F}_t^k\right]$$

$$= \frac{1}{n-1}\sum_{\substack{j\le k\\ j\neq i}} b(X_t^i,X_t^j) + \frac{n-k}{n-1}\mathbb{E}\left[b(X_t^i,X_t^{k+1})\,\big|\,\mathcal{F}_t^k\right].$$

Then

$$dX_t^i = \hat{b}_i^k(t,X^1,\dots,X^k)\,dt + d\hat{W}_t^i \quad \dots \qquad \begin{matrix} i=1,\dots,k \\ k=1,\dots,n\end{matrix}$$

Compare to

$$dY_t^i = \langle \mu_t, b(Y_t^i,\cdot)\rangle dt + d\tilde{W}_t^i. \qquad \underline{\underline{\text{BBGKY}}}$$

Then

$$\frac{d}{dt}H_t^k = \frac{1}{2}\sum_{i=1}^{k}\mathbb{E}\left|\hat{b}_i^k(t,X^1,\dots,X^k) - \langle\mu_t, b(X_t^i,\cdot)\rangle\right|^2$$

$$= \frac{k}{2}\mathbb{E}\left|\hat{b}_1^k(t,X^1,\dots,X^k) - \langle\mu_t, b(X_t^1,\cdot)\rangle\right|^2$$

$$= k\,\mathbb{E}\left|\frac{1}{n-1}\sum_{\substack{j\le k\\ j\neq i}}\left(b(X_t^1,X_t^j) - \langle\mu_t, b(X_t^1,\cdot)\rangle\right)\right|^2$$

$$+ k\left(\frac{n-k}{n-1}\right)^2\mathbb{E}\left|\mathbb{E}\left[b(X_t^1,X_t^{k+1})\,\big|\,\mathcal{F}_t^k\right] - \langle\mu_t, b(X_t^1,\cdot)\rangle\right|^2$$

$$=: A_k + B_k$$

(Note: For $k=n$, $B_n = 0$, $A_n = O(n)$.)

• A ... $O(k^3)$ ...

- $A_k = O\left(\frac{k^3}{n^2}\right)$ *(ignores correlation)

- $B_k \leq k\,\mathbb{E}\left|\left\langle \mathrm{Law}(x^{k+1}|x^1,\dots,x^k) - \mu_t,\ b(x_t^1,\cdot)\right\rangle\right|^2$

$$\leq 2\|b\|_\infty^2\, k\,\mathbb{E}\, H\left(L(x^{k+1}|x^1,\dots,x^k)\,\middle|\,\mu_t\right) \quad \text{(Pinsker)}$$

$$= Ck\left(H_t^{k+1} - H_t^k\right)$$

Chain rule:
$H(\text{joint}) = H(\text{marg}) + \int H(\text{cond})$
joint $= [k+1]$
marg $= [k]$
cond $= [k+1/k]$

... Key inequality:

For $1 \leq i < n$: $\frac{d}{dt} H_t^k \leq C\boxed{\frac{k^3}{n^2}} + ck\left(H_t^{k+1} - H_t^k\right)$

Gronwall: $H_t^k \leq e^{-ckt} H_0^k + c\int_0^t \left(\frac{k^3}{n^2} + kH_s^{k+1}\right) e^{-ck(t-s)}\,ds$

Iterate $k, k+1, \dots, n$.

Use crude global bound $H^n = O(n)$ to "close."

(Note: $O\left(\left(\frac{k}{n}\right)^2\right)$ is optimal,
Ex: b linear (Gaussian case))