Консервативні стохастичні диференціальних рівнянь з частинними похідними, як флуктуаційні границі середнього поля для стохастичного градієнтного спуску

Конаровський Віталій

Інститут Математики НАНУ – Київ

спільна робота з Benjamin Gess, Rishabh Gvalani та Sebastian Kassing

Зміст

Кероване навчання

- Маючи велику базу даних $\{(\theta_i, \gamma_i),\ i \in I\}$, де $\theta_i \sim P$ незалежні однаково розподілені випадкові величини, потрібно знайти функцію $f : \Theta \to \mathbb{R}$ таку, що $f(\theta_i) = \gamma_i$.

Кероване навчання

- Маючи велику базу даних $\{(\theta_i, \gamma_i),\ i \in I\}$, де $\theta_i \sim P$ незалежні однаково розподілені випадкові величини, потрібно знайти функцію $f : \Theta \to \mathbb{R}$ таку, що $f(\theta_i) = \gamma_i$.
- Будемо апроксимувати f за допомогою

$$f_n(\theta; x) = \frac{1}{n} \sum_{k=1}^{n} \Phi(\theta, x_k),$$

де $x_k \in \mathbb{R}^d$, $k \in \{1, \dots, n\}$, — параметри, які потрібно знайти.
Приклад: $\Phi(\theta, x_k) = c_k \cdot h(A_k\theta + b_k), \quad x_k = (A_k, b_k, c_k)$

Кероване навчання

- Маючи велику базу даних $\{(\theta_i, \gamma_i),\ i \in I\}$, де $\theta_i \sim P$ незалежні однаково розподілені випадкові величини, потрібно знайти функцію $f : \Theta \to \mathbb{R}$ таку, що $f(\theta_i) = \gamma_i$.
- Будемо апроксимувати f за допомогою

$$f_n(\theta; x) = \frac{1}{n} \sum_{k=1}^{n} \Phi(\theta, x_k),$$

де $x_k \in \mathbb{R}^d$, $k \in \{1, \ldots, n\}$, — параметри, які потрібно знайти.
Приклад: $\Phi(\theta, x_k) = c_k \cdot h(A_k \theta + b_k)$, $\quad x_k = (A_k, b_k, c_k)$

- Будемо виміряти відстань між f та f_n за допомогою **очікуваної похибки**

$$\mathcal{L}(x) := \frac{1}{2} \mathbb{E}_P |f(\theta) - f_n(\theta; x)|^2 = \frac{1}{2} \int_\Theta |f(\theta) - f_n(\theta; x)|^2 P(d\theta),$$

де P — розподіл θ_i.

Стохастичний ґрадієнтний спуск

Нехай $x_k(0)$ незалежні випадкові величини з розподілом μ_0

Стохастичний ґрадієнтний спуск

Нехай $x_k(0)$ незалежні випадкові величини з розподілом μ_0

Параметри x_k, $k \in \{1, \ldots, n\}$ можна визначити використовуючи стохастичний ґрадієнтний спуск

$$x_k(t_{i+1}) = x_k(t_i) - \nabla_{x_k}\left(\frac{1}{2}|f(\theta_i) - f_n(\theta_i; x)|^2\right)\Delta t$$

де Δt – **швидкість навчання**, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р.,

Стохастичний ґрадієнтний спуск

Нехай $x_k(0)$ незалежні випадкові величини з розподілом μ_0

Параметри x_k, $k \in \{1, \dots, n\}$ можна визначити використовуючи стохастичний ґрадієнтний спуск

$$\begin{aligned} x_k(t_{i+1}) &= x_k(t_i) - \nabla_{x_k} \left(\frac{1}{2} \left| f(\theta_i) - f_n(\theta_i; x) \right|^2 \right) \Delta t \\ &= x_k(t_i) - \left(f_n(\theta_i; x) - f(\theta_i) \right) \nabla_{x_k} \Phi(\theta_i, x_k(t_i)) \Delta t \end{aligned}$$

де Δt – **швидкість навчання**, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р.,

Стохастичний ґрадієнтний спуск

Нехай $x_k(0)$ незалежні випадкові величини з розподілом μ_0

Параметри x_k, $k \in \{1, \dots, n\}$ можна визначити використовуючи стохастичний ґрадієнтний спуск

$$\begin{aligned} x_k(t_{i+1}) &= x_k(t_i) - \nabla_{x_k}\left(\frac{1}{2}|f(\theta_i) - f_n(\theta_i; x)|^2\right)\Delta t \\ &= x_k(t_i) - (f_n(\theta_i; x) - f(\theta_i))\,\nabla_{x_k}\Phi(\theta_i, x_k(t_i))\Delta t \\ &= x_k(t_i) + \left(\nabla F(x_k(t_i), \theta_i) - \frac{1}{n}\sum_{l=1}^{n}\nabla_{x_k} K(x_k(t_i), x_l(t_i), \theta_i)\right)\Delta t \end{aligned}$$

де Δt – **швидкість навчання**, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р.,
$F(x, \theta) = f(\theta)\Phi(\theta, x)$ та $K(x, y, \theta) = \Phi(\theta, x)\Phi(\theta, y)$.

Стохастичний ґрадієнтний спуск

Нехай $x_k(0)$ незалежні випадкові величини з розподілом μ_0

Параметри x_k, $k \in \{1, \ldots, n\}$ можна визначити використовуючи стохастичний ґрадієнтний спуск

$$\begin{aligned} x_k(t_{i+1}) &= x_k(t_i) - \nabla_{x_k}\left(\frac{1}{2}|f(\theta_i) - f_n(\theta_i; x)|^2\right)\Delta t \\ &= x_k(t_i) - (f_n(\theta_i; x) - f(\theta_i))\,\nabla_{x_k}\Phi(\theta_i, x_k(t_i))\Delta t \\ &= x_k(t_i) + \left(\nabla F(x_k(t_i), \theta_i) - \langle \nabla_x K(x_k(t_i), \cdot, \theta_i), \nu^n_{t_i}\rangle\right)\Delta t \end{aligned}$$

де Δt – **швидкість навчання**, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р., $\nu^n_t = \frac{1}{n}\sum_{l=1}^n \delta_{x_l(t)}$, $F(x, \theta) = f(\theta)\Phi(\theta, x)$ та $K(x, y, \theta) = \Phi(\theta, x)\Phi(\theta, y)$.

Стохастичний ґрадієнтний спуск

Нехай $x_k(0)$ незалежні випадкові величини з розподілом μ_0

Параметри x_k, $k \in \{1, \ldots, n\}$ можна визначити використовуючи стохастичний ґрадієнтний спуск

$$\begin{aligned} x_k(t_{i+1}) &= x_k(t_i) - \nabla_{x_k} \left(\frac{1}{2} |f(\theta_i) - f_n(\theta_i; x)|^2 \right) \Delta t \\ &= x_k(t_i) - (f_n(\theta_i; x) - f(\theta_i)) \nabla_{x_k} \Phi(\theta_i, x_k(t_i)) \Delta t \\ &= x_k(t_i) + \left(\nabla F(x_k(t_i), \theta_i) - \langle \nabla_x K(x_k(t_i), \cdot, \theta_i), \nu^n_{t_i} \rangle \right) \Delta t \\ &= x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i) \Delta t \end{aligned}$$

де Δt – **швидкість навчання**, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р., $\nu^n_t = \frac{1}{n} \sum_{l=1}^{n} \delta_{x_l(t)}$, $F(x, \theta) = f(\theta)\Phi(\theta, x)$ та $K(x, y, \theta) = \Phi(\theta, x)\Phi(\theta, y)$.

Збіжність до детермінованого ДРЧП

Нагадаємо, що $x_k(0) \sim \mu_0$ – н.о.р., Δt – швидкість навчання, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р.

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t, \quad k \in \{1, \dots, n\},$$

де $\nu^n_t = \frac{1}{n}\sum_{k=1}^n \delta_{x_k(t)}$.

Збіжність до детермінованого ДРЧП

Нагадаємо, що $x_k(0) \sim \mu_0$ – н.о.р., Δt – швидкість навчання, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р.

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t, \quad k \in \{1, \dots, n\},$$

де $\nu^n_t = \frac{1}{n}\sum_{k=1}^n \delta_{x_k(t)}$.

Для емпіричного розподілу $\nu^n = \frac{1}{n}\sum_{k=1}^n \delta_{x_k}$, маємо

$$f_n(\theta; x) = \frac{1}{n}\sum_{k=1}^n \Phi(\theta, x_k) = \langle \Phi(\theta, \cdot), \nu^n \rangle.$$

Збіжність до детермінованого ДРЧП

Нагадаємо, що $x_k(0) \sim \mu_0$ – н.о.р., Δt – швидкість навчання, $t_i = i\Delta t$, $\theta_i \sim P$ – н.о.р.

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t, \quad k \in \{1, \dots, n\},$$

де $\nu^n_t = \frac{1}{n}\sum_{k=1}^n \delta_{x_k(t)}$.

Для емпіричного розподілу $\nu^n = \frac{1}{n}\sum_{k=1}^n \delta_{x_k}$, маємо

$$f_n(\theta; x) = \frac{1}{n}\sum_{k=1}^n \Phi(\theta, x_k) = \langle \Phi(\theta, \cdot), \nu^n \rangle.$$

Якщо $x_k(0) \sim \mu_0$ н.о.р. тоді

$$d(\nu^n_t, \mu_t) = O(n^{-1/2}) + O(\Delta t^{1/2}) = O(n^{-1/2}), \quad \text{для } \Delta t = \frac{1}{n},$$

де μ_t є розв'язком рівняння

$$d\mu_t = -\nabla \left(V(\cdot, \mu_t)\mu_t \right) dt$$

з $V(x, \mu) = \mathbb{E}_P V(x, \mu, \theta)$. [Mei, Montanarib, Nguyen, 2018]

Main Goal

Проблема. Після переходу до границі рівняння

$$d\mu_t = -\nabla \left(V(\cdot, \mu_t)\mu_t \right) dt$$

втрачає інформацію про флуктуації в динаміці стохастичного ґрадієнтного спуску

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t, \quad \nu^n_t = \frac{1}{n}\sum_{l=1}^{n} \delta_{x_l(t)}.$$

Main Goal

Проблема. Після переходу до границі рівняння

$$d\mu_t = -\nabla \left(V(\cdot, \mu_t)\mu_t \right) dt$$

втрачає інформацію про флуктуації в динаміці стохастичного ґрадієнтного спуску

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t, \quad \nu^n_t = \frac{1}{n}\sum_{l=1}^{n} \delta_{x_l(t)}.$$

Мета: Запропонувати **стохастичне** ДРЧП, яке б враховувало флуктуації динаміки СҐС. Тоді, можливо, його розв'язки будуть краще апроксимувати динаміку СҐС при $n \to \infty$ та $\Delta t \to 0$.

Класичне СДР для динаміки СҐС

Стохастичний ґрадієнтний спуск

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t$$

Класичне СДР для динаміки СҐС

Стохастичний ґрадієнтний спуск

$$
\begin{aligned}
x_k(t_{i+1}) &= x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t \\
&= x_k(t_i) + \mathbb{E}_\theta V(\dots)\Delta t + \sqrt{\Delta t}\,(V(\dots) - \mathbb{E}_\theta V(\dots))\sqrt{\Delta t}
\end{aligned}
$$

Класичне СДР для динаміки СҐС

Стохастичний ґрадієнтний спуск

$$
\begin{aligned}
x_k(t_{i+1}) &= x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t \\
&= x_k(t_i) + \underbrace{\mathbb{E}_\theta V(\dots)}_{=V(x_k(t_i),\nu^n_{t_i})}\Delta t + \underbrace{\sqrt{\Delta t}}_{=\sqrt{\alpha}}\underbrace{(V(\dots) - \mathbb{E}_\theta V(\dots))}_{=G(x_k(t_i),\nu^n_{t_i},\theta_i)}\sqrt{\Delta t}
\end{aligned}
$$

Класичне СДР для динаміки СҐС

Стохастичний ґрадієнтний спуск

$$\begin{aligned} x_k(t_{i+1}) &= x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t \\ &= x_k(t_i) + \underbrace{\mathbb{E}_\theta V(\dots)}_{=V(x_k(t_i),\nu^n_{t_i})} \Delta t + \underbrace{\sqrt{\Delta t}}_{=\sqrt{\alpha}} \underbrace{(V(\dots) - \mathbb{E}_\theta V(\dots))}_{=G(x_k(t_i),\nu^n_{t_i},\theta_i)} \sqrt{\Delta t} \end{aligned}$$

є схемою Ейлера-Маруями для СДР

$$dX_k(t) = V(X_k(t), \mu^n_t)dt + \sqrt{\alpha}(\Sigma^{\frac{1}{2}})_k(X(t))dB(t), \quad k \in \{1, \dots, n\}$$

де $\mu^n_t = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, $\Sigma_{k,l}(x) = \mathbb{E}_\theta G(x_k, \mu, \theta) \otimes G(x_l, \mu, \theta)$ та B – n-вимірний броунівський рух.

Класичне СДР для динаміки СҐС

Стохастичний ґрадієнтний спуск

$$\begin{aligned} x_k(t_{i+1}) &= x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t \\ &= x_k(t_i) + \underbrace{\mathbb{E}_\theta V(\dots)}_{=V(x_k(t_i),\nu^n_{t_i})}\Delta t + \underbrace{\sqrt{\Delta t}}_{=\sqrt{\alpha}}\underbrace{(V(\dots) - \mathbb{E}_\theta V(\dots))}_{=G(x_k(t_i),\nu^n_{t_i},\theta_i)}\sqrt{\Delta t} \end{aligned}$$

є схемою Ейлера-Маруями для СДР

$$dX_k(t) = V(X_k(t), \mu^n_t)dt + \sqrt{\alpha}(\Sigma^{\frac{1}{2}})_k(X(t))dB(t), \quad k \in \{1, \dots, n\}$$

де $\mu^n_t = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, $\Sigma_{k,l}(x) = \mathbb{E}_\theta G(x_k, \mu, \theta) \otimes G(x_l, \mu, \theta)$ та B – n-вимірний броунівський рух.

$\rightsquigarrow \quad \Sigma^{\frac{1}{2}}$ є $dn \times dn$ матрицею!

Проблема мартингалів для емпіричного розподілу

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}(\Sigma^{\frac{1}{2}})_k(X(t))dB(t), \quad k \in \{1, \dots, n\}$$

де $\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, $\Sigma_{k,l}(x) = \tilde{A}(x_k, x_l, \mu) := \mathbb{E}_\theta G(x_k, \mu, \theta) \otimes G(x_l, \mu, \theta)$

Проблема мартингалів для емпіричного розподілу

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}(\Sigma^{\frac{1}{2}})_k(X(t))dB(t), \quad k \in \{1, \ldots, n\}$$

де $\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, $\Sigma_{k,l}(x) = \tilde{A}(x_k, x_l, \mu) := \mathbb{E}_\theta G(x_k, \mu, \theta) \otimes G(x_l, \mu, \theta)$

Беручи $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d)$, ми отримаємо для емпіричної міри μ_t^n

$$\langle \varphi, \mu_t^n \rangle = \langle \varphi, \mu_0^n \rangle + \frac{\alpha}{2}\int_0^t \left\langle \nabla^2 \varphi : A(\cdot, \mu_s^n), \mu_s^n \right\rangle ds + \int_0^t \langle \nabla \varphi \cdot V(\cdot, \mu_s^n), \mu_s^n \rangle\, ds$$
$$+ \mathrm{Mart.},$$

де $A(x, \mu) = \tilde{A}(x, x, \mu)$

Проблема мартингалів для емпіричного розподілу

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}(\Sigma^{\frac{1}{2}})_k(X(t))dB(t), \quad k \in \{1, \dots, n\}$$

де $\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, $\Sigma_{k,l}(x) = \tilde{A}(x_k, x_l, \mu) := \mathbb{E}_\theta G(x_k, \mu, \theta) \otimes G(x_l, \mu, \theta)$

Беручи $\varphi \in \mathcal{C}_c^2(\mathbb{R}^d)$, ми отримаємо для емпіричної міри μ_t^n

$$\langle \varphi, \mu_t^n \rangle = \langle \varphi, \mu_0^n \rangle + \frac{\alpha}{2}\int_0^t \left\langle \nabla^2 \varphi : A(\cdot, \mu_s^n), \mu_s^n \right\rangle ds + \int_0^t \langle \nabla\varphi \cdot V(\cdot, \mu_s^n), \mu_s^n \rangle ds$$
$$+ \mathrm{Mart.},$$

де $A(x, \mu) = \tilde{A}(x, x, \mu)$ та

$$[\mathrm{Mart.}]_t = \alpha \int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\nabla\varphi(x) \otimes \nabla\varphi(y)) : \tilde{A}(x, y, \mu_s^n)\mu_s^n(dx)\mu_s^n(dy)ds$$

[Rotskoff, Vanden-Eijnden, CPAM, 2022]

СДР із нескінченновимірним шумом для СҐС

Стохастичний ґрадієнтний спуск

$$x_k(t_{i+1}) = x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t$$
$$= x_k(t_i) + \underbrace{\mathbb{E}_\theta V(\dots)}_{=V(x_k(t_i),\nu^n_{t_i})} \Delta t + \underbrace{\sqrt{\Delta t}}_{=\sqrt{\alpha}} \underbrace{(V(\dots) - \mathbb{E}_\theta V(\dots))}_{=G(x_k(t_i),\nu^n_{t_i},\theta_i)} \sqrt{\Delta t}$$

є схемою Ейлера-Маруями для СДР

$$dX_k(t) = V(X_k(t), \mu^n_t)dt + \sqrt{\alpha}(\Sigma^{\frac{1}{2}})_k(X(t))dB(t), \quad k \in \{1, \dots, n\}$$

де $\mu^n_t = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, $\Sigma_{k,l}(x) = \mathbb{E}_\theta G(x_k, \mu, \theta) \otimes G(x_l, \mu, \theta)$ та B – n-вимірний броунівський рух.

СДР із нескінченновимірним шумом для СҐС

Стохастичний ґрадієнтний спуск

$$
\begin{aligned}
x_k(t_{i+1}) &= x_k(t_i) + V(x_k(t_i), \nu^n_{t_i}, \theta_i)\Delta t \\
&= x_k(t_i) + \underbrace{\mathbb{E}_\theta V(\dots)}_{=V(x_k(t_i),\nu^n_{t_i})}\Delta t + \underbrace{\sqrt{\Delta t}}_{=\sqrt{\alpha}}\underbrace{(V(\dots) - \mathbb{E}_\theta V(\dots))}_{=G(x_k(t_i),\nu^n_{t_i},\theta_i)}\sqrt{\Delta t}
\end{aligned}
$$

є схемою Ейлера-Маруями для СДР

$$
dX_k(t) = V(X_k(t), \mu^n_t)dt + \sqrt{\alpha}\int_\Theta G(X_k(t), \mu^n_t, \theta)W(d\theta, dt), \quad k \in \{1, \dots, n\}
$$

де $\mu^n_t = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, W – білий шум на $L_2(\Theta, P)$ (P – розподіл θ).

[Gess, Kassing, K. Journal of Machine Learning Research '24]

Стохастичне рівняння середнього поля

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}\int_\Theta G(X_k(t), \mu_t^n, \theta)W(d\theta, dt), \quad k \in \{1, \dots, n\}$$

де $\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, W – білий шум на $L_2(\Theta, P)$.

Стохастичне рівняння середнього поля

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha} \int_\Theta G(X_k(t), \mu_t^n, \theta) W(d\theta, dt), \quad k \in \{1, \dots, n\}$$

де $\mu_t^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(t)}$, W – білий шум на $L_2(\Theta, P)$.

Використовуючи формулу Іто, отримаємо **Стохастичне диференціальне рівняння середнього поля:**

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)dt$$

Стохастичне рівняння середнього поля

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}\int_\Theta G(X_k(t), \mu_t^n, \theta)W(d\theta, dt), \quad k \in \{1, \ldots, n\}$$

де $\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, W – білий шум на $L_2(\Theta, P)$.

Використовуючи формулу Іто, отримаємо **Стохастичне диференціальне рівняння середнього поля:**

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)dt + \sqrt{\alpha}\nabla \cdot \int_\Theta G(\cdot, \mu_t, \theta)\mu_t W(d\theta, dt)$$

де $A(x_k, \mu) = \mathbb{E}_\theta G(x_k, \mu) \otimes G(x_k, \mu)$.

Стохастичне рівняння середнього поля

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}\int_\Theta G(X_k(t), \mu_t^n, \theta)W(d\theta, dt), \quad k \in \{1, \dots, n\}$$

де $\mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}$, W – білий шум на $L_2(\Theta, P)$.

Використовуючи формулу Іто, отримаємо **Стохастичне диференціальне рівняння середнього поля:**

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)dt + \sqrt{\alpha}\nabla \cdot \int_\Theta G(\cdot, \mu_t, \theta)\mu_t \, W(d\theta, dt)$$

де $A(x_k, \mu) = \mathbb{E}_\theta G(x_k, \mu) \otimes G(x_k, \mu)$.

$\rightsquigarrow$ Проблема мартингалів для цього рівняння та ж сама, що і в [Rotskoff, Vanden-Eijnden, CPAM, '22]

Пов'язані роботи

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)\, dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)\, dt - \sqrt{\alpha}\nabla \cdot \int_\Theta (G(\cdot, \mu_t, \theta)\mu_t)\, W(d\theta, dt),$$

Існування та єдиність розв'язків для схожих СДРЧП

- **Рівняння неперервності в гідродинаміці та оптимальному транспортуванні** [Ambrosio, Trevisan, Crippa...]. Тут $A = G = 0$.

Пов'язані роботи

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)\, dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)\, dt - \sqrt{\alpha}\nabla \cdot \int_\Theta (G(\cdot, \mu_t, \theta)\mu_t)\, W(d\theta, dt),$$

Існування та єдиність розв'язків для схожих СДРЧП

- **Рівняння неперервності в гідродинаміці та оптимальному транспортуванні** [Ambrosio, Trevisan, Crippa...]. Тут $A = G = 0$.
- **Стохастичне нелінійне рівняння Фоккера-Планка** [Coghi, Gess '19]. Коваріація A має більш загальну структуру (тобто $A - \mathbb{E}G \otimes G \geq 0$) але шум – скінченновимірний.

Пов'язані роботи

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)\, dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)\, dt - \sqrt{\alpha}\nabla \cdot \int_\Theta (G(\cdot, \mu_t, \theta)\mu_t)\, W(d\theta, dt),$$

Існування та єдиність розв'язків для схожих СДРЧП

- **Рівняння неперервності в гідродинаміці та оптимальному транспортуванні** [Ambrosio, Trevisan, Crippa...]. Тут $A = G = 0$.
- **Стохастичне нелінійне рівняння Фоккера-Планка** [Coghi, Gess '19]. Коваріація A має більш загальну структуру (тобто $A - \mathbb{E}G \otimes G \geq 0$) але шум – скінченновимірний.

Пов'язані роботи

$$d\mu_t = -\nabla \cdot (V(\cdot,\mu_t)\mu_t)\,dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot,\mu_t)\mu_t)\,dt - \sqrt{\alpha}\nabla \cdot \int_\Theta (G(\cdot,\mu_t,\theta)\mu_t)\,W(d\theta,dt),$$

Існування та єдиність розв'язків для схожих СДРЧП

- **Рівняння неперервності в гідродинаміці та оптимальному транспортуванні** [Ambrosio, Trevisan, Crippa...]. Тут $A = G = 0$.
- **Стохастичне нелінійне рівняння Фоккера-Планка** [Coghi, Gess '19]. Коваріація A має більш загальну структуру (тобто $A - \mathbb{E}G \otimes G \geq 0$) але шум – скінченновимірний.
- **Нелінійне СДРЧП для системи частинок** [Kurtz, Xiong '99]. Рівняння має більш загальну форму але початкова умова μ_0 мусить володіти L_2-щільністю відносно міри Лебеґа.

Пов'язані роботи

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)\, dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)\, dt - \sqrt{\alpha}\nabla \cdot \int_\Theta (G(\cdot, \mu_t, \theta)\mu_t)\, W(d\theta, dt),$$

Існування та єдиність розв'язків для схожих СДРЧП

- **Рівняння неперервності в гідродинаміці та оптимальному транспортуванні** [Ambrosio, Trevisan, Crippa. . .]. Тут $A = G = 0$.
- **Стохастичне нелінійне рівняння Фоккера-Планка** [Coghi, Gess '19]. Коваріація A має більш загальну структуру (тобто $A - \mathbb{E}G \otimes G \geq 0$) але шум – скінченновимірний.
- **Нелінійне СДРЧП для системи частинок** [Kurtz, Xiong '99]. Рівняння має більш загальну форму але початкова умова μ_0 мусить володіти L_2-щільністю відносно міри Лебеґа.

Результат з [Kurtz, Xiong] не може бути застосованим до нашого рівняння якщо μ_0 не має L_2-щільності!

Зміст

Рівняння неперервності

$$d\mu_t = -\nabla \cdot (V\mu_t)dt$$

Рівняння неперервності

$$d\mu_t = -\nabla \cdot (V\mu_t)dt$$

$$\implies \mu_t = \mu_0 \circ X(\cdot, t),$$

де

$$dX(u,t) = V(X(u,t))dt, \quad X(u,0) = u.$$

[Ambrosio, Trevisan, Lions,...]

Рівняння неперервності

$$d\mu_t = -\nabla \cdot (V\mu_t)dt$$

$$\implies \mu_t = \mu_0 \circ X(\cdot, t),$$

де

$$dX(u,t) = V(X(u,t))dt, \quad X(u,0) = u.$$

[Ambrosio, Trevisan, Lions,...]

Стохастичне рівняння середнього поля отримане з:

$$dX_k(t) = V(X_k(t), \mu_t^n)dt + \sqrt{\alpha}\int_\Theta G(X_k(t), \mu_t^n, \theta)W(d\theta, dt),$$

$$X_k(0) = x_k(0), \quad \mu_t^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}.$$

Існування та єдиність розв'язків

Теорема 1 (Gess, Gvalani, K. 2022)

Нехай коефіцієнти рівняння V, G – ліпшицеві та достатньо гладкі відносно просторової змінної. Тоді рівняння

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)\, dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)\, dt \\ - \sqrt{\alpha}\nabla \cdot \int_\Theta G(\cdot, \mu_t, \theta)\mu_t W(d\theta, dt)$$

має єдиний розв'язок. Крім того, μ_t задовольняє принцип суперпозиції, тобто,

$$\mu_t = \mu_0 \circ X^{-1}(\cdot, t), \quad t \geq 0,$$

де X – розв'язок рівняння

$$dX(u,t) = V(X(u,t), \mu_t)dt + \sqrt{\alpha}\int_\Theta G(X(u,t), \mu_t, \theta)W(d\theta, dt)$$

$$X(u,0) = u, \quad u \in \mathbb{R}^d.$$ [Дороговцев, Kotelenez, Пилипенко '97-...]

Зміст

Відстань Вассерштейна

Нехай (E, d) – повний сепарабельний метричний простір та $p \geq 1$ $\mathcal{P}_p(E)$ – простір ймовірнісних мір ρ на E з

$$\int_E d^p(x, o)\rho(dx) < \infty.$$

Відстань Вассерштейна

Нехай (E, d) – повний сепарабельний метричний простір та $p \geq 1$ $\mathcal{P}_p(E)$ – простір ймовірнісних мір ρ на E з

$$\int_E d^p(x, o)\rho(dx) < \infty.$$

Для $\rho_1, \rho_2 \in \mathcal{P}_p(E)$ визначимо **метрику Вассерштейна** за допомогою

$$\mathcal{W}_p^p(\rho_1, \rho_2) = \inf \left\{ \mathbb{E} d^p(\xi_1, \xi_2) : \quad \xi_i \sim \rho_i \right\}$$

Наближення вищого порядку для СҐС

Стохастичне диференціальне рівняння середнього поля:

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)dt + \sqrt{\alpha}\nabla \cdot \int_\Theta G(\cdot, \mu_t, \theta)\mu_t \, W(d\theta, dt)$$

де $A(x_k, \mu) = \mathbb{E}_\theta G(x_k, \mu) \otimes G(x_k, \mu)$.

Наближення вищого порядку для СҐС

Стохастичне диференціальне рівняння середнього поля:

$$d\mu_t = -\nabla \cdot (V(\cdot, \mu_t)\mu_t)dt + \frac{\alpha}{2}\nabla^2 : (A(\cdot, \mu_t)\mu_t)dt + \sqrt{\alpha}\nabla \cdot \int_\Theta G(\cdot, \mu_t, \theta)\mu_t \, W(d\theta, dt)$$

де $A(x_k, \mu) = \mathbb{E}_\theta G(x_k, \mu) \otimes G(x_k, \mu)$.

Теорема 2 (Gess, Gvalani, K. 2022)

- V, G – ліпшицеві і диф. відносно просторової змінної з обмеженими похідними;
- ν_t^n – емпіричний розподіл асоційований з динамікою СҐС з $\alpha = \frac{1}{n}$;
- μ_t^n – (єдиний) розв'язок рівняння середнього поля з

$$\mu_0^n = \nu_0^n = \frac{1}{n}\sum_{k=1}^{n} \delta_{x_k(0)}$$

та $x_k(0) \sim \mu_0$ – н.о.р.

Тоді для всіх $p \in [1, 2)$ $\mathcal{W}_p(\mathrm{Law}\, \mu^n, \mathrm{Law}\, \nu^n) = o(n^{-1/2})$.

Кількісна гранична теорема для ДРСП

Теорема 3 (Gess, Gvalani, K. 2022)

За припущень попередньої теореми, $\eta_t^n := \sqrt{n}\left(\mu_t^n - \mu_t^0\right) \to \eta_t$
де η_t – гаусівський процес, який є розв'язком

$$d\eta_t = -\nabla \cdot \left(V(\cdot, \mu_t^0)\eta_t + \langle \nabla K(x, \cdot), \eta_t \rangle \mu_t^0(dx)\right) dt - \nabla \cdot \int_\Theta G(\cdot, \mu_t^0, \theta)\mu_t^0 W(d\theta, dt).$$

Крім того, $\mathbb{E} \sup_{t \in [0,T]} \|\eta_t^n - \eta_t\|_{-J}^2 \leq \frac{C}{n}$.

Кількісна гранична теорема для ДРСП

Теорема 3 (Gess, Gvalani, K. 2022)

За припущень попередньої теореми, $\eta_t^n := \sqrt{n}\left(\mu_t^n - \mu_t^0\right) \to \eta_t$
де η_t – гаусівський процес, який є розв'язком

$$d\eta_t = -\nabla \cdot \left(V(\cdot, \mu_t^0)\eta_t + \langle \nabla K(x, \cdot), \eta_t \rangle \mu_t^0(dx)\right) dt - \nabla \cdot \int_\Theta G(\cdot, \mu_t^0, \theta)\mu_t^0 W(d\theta, dt).$$

Крім того, $\mathbb{E} \sup\limits_{t\in[0,T]} \|\eta_t^n - \eta_t\|_{-J}^2 \leq \frac{C}{n}$.

Зауваження. [Sirignano, Spiliopoulos, '20]

Для $\tilde{\eta}_t^n := \sqrt{n}(\nu_t^n - \mu_t^0)$

$$\mathbb{E} \sup_{t\in[0,T]} \|\tilde{\eta}_t^n\|_{-J}^2 \leq C \quad \text{та} \quad \tilde{\eta}^n \to \eta.$$

ЦГТ для ДРСП + ЦГТ для СҐС $\implies$ Набл. вищого порядку

Зауважимо, що

$$\mu_t^n = \mu_t^0 + n^{-1/2}\eta + O(n^{-1}).$$

ЦГТ для ДРСП + ЦГТ для СҐС $\implies$ Набл. вищого порядку

Зауважимо, що

$$\mu_t^n = \mu_t^0 + n^{-1/2}\eta + O(n^{-1}).$$
$$\nu_t^n = \mu_t^0 + n^{-1/2}\eta + o(n^{-1/2}).$$

ЦГТ для ДРСП + ЦГТ для СҐС $\implies$ Набл. вищого порядку

Зауважимо, що

$$\mu_t^n = \mu_t^0 + n^{-1/2}\eta + O(n^{-1}).$$
$$\nu_t^n = \mu_t^0 + n^{-1/2}\eta + o(n^{-1/2}).$$

Тому, $\mu^n - \nu^n = o(n^{-1/2})$.

ЦГТ для ДРСП + ЦГТ для СҐС $\implies$ Набл. вищого порядку

Зауважимо, що

$$\mu_t^n = \mu_t^0 + n^{-1/2}\eta + O(n^{-1}).$$
$$\nu_t^n = \mu_t^0 + n^{-1/2}\eta + o(n^{-1/2}).$$

Тому, $\mu^n - \nu^n = o(n^{-1/2})$.

$$\sqrt{n^p}\mathcal{W}_p^p\left(\mathrm{Law}(\mu^n), \mathrm{Law}(\nu^n)\right) = \sqrt{n^p}\inf \mathbb{E}\left[\sup_{t\in[0,T]} \|\mu_t^n - \nu_t^n\|_{-J}^p\right]$$
$$= \inf \mathbb{E}\left[\sup_{t\in[0,T]} \|\sqrt{n}(\mu_t^n - \mu_t^0) - \sqrt{n}(\nu_t^n - \mu_t^0)\|_{-J}^p\right]$$
$$= \mathcal{W}_p^p\left(\mathrm{Law}(\eta^n), \mathrm{Law}(\tilde{\eta}^n)\right) \to 0.$$

Reference

Gess, Kassing, Konarovskyi,
Stochastic Modified Flows, Mean-Field Limits and Dynamics of Stochastic Gradient Descent
Journal of Machine Learning Research (2024)

Gess, Gvalani, Konarovskyi,
Conservative SPDEs as fluctuating mean field limits of stochastic gradient descent
(arXiv:2207.05705)

Konarovskyi, Lehmann, von Renesse,
Dean-Kawasaki Dynamics with smooth Drift Potentials
Journal of Statistical Physics (2020)

Дякую за увагу