目录

第 1 章　$\mathscr{P}_2$ 上的函数的 Lions 导数

本章内容主要参考引用自 (Carmona et al., 2018) 的 5.2 节.

记 $\mathscr{P}$ 表示 $\mathbb{R}^d$ 上全体概率测度,$\mathscr{P}_p$ 表示全体 p 阶矩有限概率测度, 即

$$\mathscr{P}_p = \{\mu \in \mathscr{P} : \int_{\mathbb{R}^d} |x|^p \mu(\mathrm{d}x) < \infty\}.$$

$\mathscr{P}_p$ 在 $p-$Wasserstein 距离 $W_p(\cdot,\cdot)$ 下是一个完备度量空间,$W_p(\cdot,\cdot)$ 定义为

$$W_p(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \Big(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \pi(\mathrm{d}x,\mathrm{d}y)\Big)^{\frac{1}{p}},$$

其中 $\Pi(\mu,\nu)$ 表示所有边缘分布为 μ,ν 的联合分布. 本章中如无特别说明, 总默认 $\mathscr{P}_2$ 配备 $W_2(\cdot,\cdot)$ 距离及其诱导的度量拓扑. 称 $\mathscr{P}_2$ 的子集 K 是有界的, 是指: 存在 $R>0$, 使得 $\forall \mu \in K$, $\int_{\mathbb{R}^d} |x|^2 \mathrm{d}\mu(x) \le R$.

对一个函数 $f : \mathscr{P}_2 \to \mathbb{R}$, 我们希望考察 $\mu \in \mathscr{P}_2$ 处的局部性态, 但由于 $\mathscr{P}_2$ 不是 Banach 空间, 甚至不是向量空间, 所以不能直接使用第一章中的 Frechet 导数和 Gateaux 导数.Lions 导数的想法是不直接考察概率测度 μ 附近函数的性态, 而是考虑将 f 先" 提升" 为定义在 $L^2(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^d)$ 上的函数 $\tilde{f}$, 考察分布为 μ 的随机变量 $X \in L^2(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^d)$ 附近 $\tilde{f}$ 的性态, 而 $L^2(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^d)$ 不仅是 Banach 空间, 也是 Hilbert 空间, 所以可以应用 Frechet 可微性的概念. 一个自然的提升是令 $\tilde{f} = f \circ \mathscr{L}$, 其中 $\mathscr{L}(X)$ 表示 $\mathbb{R}^d$ 上由随机变量 X 诱导的概率分布, 也即 $\mathscr{L}(X)(A) = \mathbb{P}(X^{-1}(A)), \forall A \in \mathscr{B}(\mathbb{R}^d)$.

$$\begin{tikzcd}
X \arrow[rr, mapsto] & & \mu = \mathscr{L}(X) \arrow[dd, mapsto] \\
L^2(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^d) \arrow[r, "\mathscr{L}"] \arrow[dr, dashed, "\tilde{f}=f\circ\mathscr{L}"'] & B \arrow[d, "f"] & \\
& \mathbb{R} & f(\mu)
\end{tikzcd}$$

由 (Bogachev, 2007) 的命题 9.1.11, 对于任意的无原子的概率空间, 任意分布 $\mu \in \mathscr{P}$, 总存在一个随机变量 X 使得 $\mathscr{L}(X) = \mu$, 这一事实保证了提升的合理性. 本章讨论提升时, 无原子的概率空间 $(\Omega,\mathscr{F},\mathbb{P})$ 总是固定的, 简记 $L^2(\Omega;\mathbb{R}^d) = L^2(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^d)$. 对任意 $A \in \mathscr{F}$, 记 $\mathscr{F}|_A := \{E \cap A : E \in \mathscr{F}\}$. 另外, 在本章中, 对于一个 Banach 空间 B 上的函数 F, 不加区分的使用 $F'(a)$ 和 $DF(a)$ 表示 F 在

$a \in B$ 处的 Frechet 导数, 同样地, 若有定义, 不加区分地使用 F' 和 DF 表示 F 的导数映射.

1.1　L-可微性

1.1.1　定义

定义 1.1　给定一个函数 $f : \mathscr{P}_2 \to \mathbb{R}$, 称 f 在 $\mu_0 \in \mathscr{P}_2$ 是 L-可微的, 是指: 存在一个分布为 μ_0 的随机变量 X_0, 使得 f 的提升 $\tilde{f} = f \circ \mathscr{L} : L^2(\Omega; \mathbb{R}^d) \to \mathbb{R}$ 在 X_0 处是 Frechet 可微的.

这个定义的表述是自然的, 但良定性并不显然, 具体地说, 对于同一个概率测度, 可能有多个随机变量以其为分布, 那么就需要说明这些随机变量和其 Frechet 导数有某种不变量或者在某种等价关系下属于同一等价类, 使得定义在一定的标准下具有唯一性. 为此, 我们先陈述以下引理.

引理 1.1　设 $X, Y \in L^2(\Omega; \mathbb{R}^d)$ 分布相同. 则对任意 $\epsilon > 0$, 存在保测映射[①] $S, T : \Omega \to \Omega$ 使得

$$\mathbb{P}(\{\omega \in \Omega : (S \circ T)(\omega) = (T \circ S)(\omega) = \omega\}) = 1$$

且

$$\mathbb{P}(|Y - X \circ T| \leq \epsilon) = 1.$$

证明梗概　记 $\mathscr{F}^*, \mathbb{P}^*$ 分别为 $\mathscr{F}, \mathbb{P}$ 的完备化扩张.

对给定的 ϵ, 设 $\{A_n\}_{n=1}^{\infty} \subset \mathscr{B}(\mathbb{R}^d)$ 是 $\mathbb{R}^d$ 的一个剖分, 且任意 A_n 直径不超过 ϵ(例如边长为 $\frac{\epsilon}{\sqrt{d}}$ 的半开半闭的方体). 令 $B_n = X^{-1}(A_n), C_n = Y^{-1}(A_n)$, 由于 X, Y 同分布, 所以 $\mathbb{P}(B_n) = \mathbb{P}(C_n)$. 不妨设 $\mathbb{P}(B_n) > 0$. 由于 $(\Omega, \mathscr{F}, \mathbb{P})$ 无原子, 所以存在 $M_n \subset \tilde{M}_n \subset B_n, N_n \subset \tilde{N}_n \subset C_n$,$\mathbb{P}(\tilde{M}_n) = \mathbb{P}(\tilde{N}_n) = 0$.

记 $\Omega_n^X = B_n - M_n, \mathscr{F}_n^X = \{A \cap \Omega_n^X : A \in \mathscr{F}^*\}, \mathbb{P}_n^X(E) = \frac{\mathbb{P}^*(E)}{\mathbb{P}(B_n)}, E \in \mathscr{F}_n^X$ 以及 $\Omega_n^Y = C_n - N_n, \mathscr{F}_n^Y = \{A \cap \Omega_n^Y : A \in \mathscr{F}^*\}, \mathbb{P}_n^Y(E) = \frac{\mathbb{P}^*(F)}{\mathbb{P}(C_n)}, F \in \mathscr{F}_n^Y$.

由 (Bogachev, 2007) 的推论 6.6.7 和定理 9.2.2 可知, 存在保测的双射

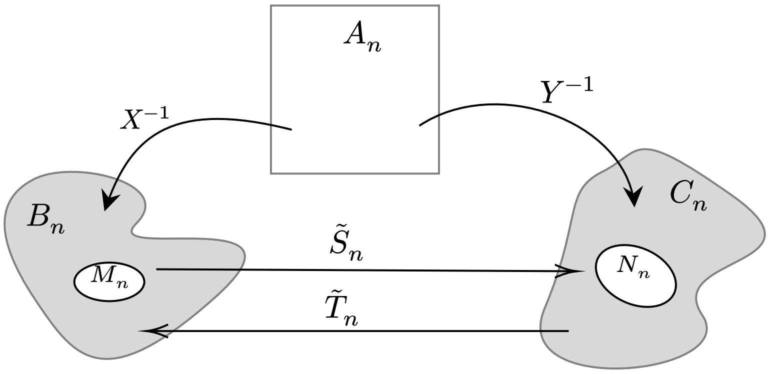
$$\tilde{S}_n : (\Omega_n^X, \mathscr{F}_n^X, \mathbb{P}_n^X) \to (\Omega_n^Y, \mathscr{F}_n^Y, \mathbb{P}_n^Y).$$

和

$$\tilde{T}_n : (\Omega_n^Y, \mathscr{F}_n^Y, \mathbb{P}_n^Y) \to (\Omega_n^X, \mathscr{F}_n^X, \mathbb{P}_n^X)$$

且 $\tilde{S}_n \circ \tilde{T}_n = \mathrm{id}_{\Omega_n^X}, \tilde{T}_n \circ \tilde{S}_n = \mathrm{id}_{\Omega_n^Y}$.

[①] 设 $(X_1, \mathscr{F}_1, \mu_1), (X_2, \mathscr{F}_2, \mu_2)$ 是两个测度空间, 称 $M : (X_1, \mathscr{F}_1, \mu_1) \to (X_2, \mathscr{F}_2, \mu_2)$ 是一个保测映射, 是指:M 可测且任意 $A \in \mathscr{F}_2, \mu_2(A) = \mu_1(M^{-1}(A))$.



下面把 $\tilde{S}_n, \tilde{T}_n$ 从 $B_n - \tilde{M}_n, C_n - \tilde{N}_n$ 延拓到 B_n, C_n 上, 记为 S_n, T_n. 只需补充在 $\tilde{M}_n, \tilde{N}_n$ 上的定义, 具体地说, 任取 $x_n \in M_n, y_n \in N_n$, 令 $\forall x \in \tilde{M}_n, S_n(x) = y_n, \forall y \in \tilde{N}_n, T_n(y) = x_n$ 即可. 容易验证, 延拓后的映射 S_n 是 $(B_n, \mathscr{F}|_{B_n})$ 到 $(C_n, \mathscr{F}|_{C_n})$ 的保测映射.T_n 类似.

当 $\mathbb{P}(B_n) = 0$ 时, 取 $M_n = B_n, N_n = C_n$.

令 $S = \sum_{n\geq 1} S_n I_{B_n}, T = \sum_{n\geq 1} T_n I_{C_n}$, 则 S, T 是 $(\Omega, \mathscr{F}, \mathbb{P})$ 到自身的保测映射. 且以概率 1 互为逆.

最后, 对几乎处处 ω, 存在 n, 使得 $\omega \in C_n$, 则 $T(\omega) \in B_n$, 所以 $X(T(\omega)), Y(\omega) \in A_n$. ■

因为 $L^2(\Omega; \mathbb{R}^d)$ 是一个 Hilbert 空间, 所以对函数 $F : L^2(\Omega; \mathbb{R}^d) \to \mathbb{R}$ 在一点 X 处的 Frechet 导数 $DF(X)$ 作为一个 Hilbert 空间的对偶空间的元素, 总可以视为 $L^2(\Omega; \mathbb{R}^d)$ 中的某个随机变量, 对任意 $Y \in L^2(\Omega; \mathbb{R}^d)$ 的作用以 $DF(X)(Y) = \langle DF(X), Y\rangle = \mathbb{E}[DF(X) \cdot Y]$ 的形式表现.

定理 1.2　设函数 $f : \mathscr{P}_2 \to \mathbb{R}$ 在 $\mu_0 \in \mathscr{P}_2$ 处 L-可微. 则对于任意分布为 μ_0 的随机变量 $X \in L^2(\Omega; \mathbb{R}^d)$,$f$ 的提升 $\tilde{f}$ 在 X 处是 Frechet 可微的, 且 $(X, D\tilde{f}(X))$ 的联合分布不依赖于 X 的选取.

证明　由定义, 存在分布为 μ 的随机变量 $X_0 \in L^2(\Omega; \mathbb{R}^d)$,$\tilde{f}$ 在 X_0 处是 Frechet 可微的. 对任意分布为 μ 的随机变量 $X \in L^2(\Omega; \mathbb{R}^d)$, 任意 $\epsilon > 0$, 存在 Ω 到自身的保测映射 S_ϵ 和 T_ϵ 满足

$$\mathbb{P}(\{\omega \in \Omega : S(T(\omega)) = T(S(\omega)) = \omega\}) = 1$$

和

$$\mathbb{P}(|X_0 - X \circ S| \leq \epsilon) = 1.$$

注意到 $\tilde{f}$ 作为 f 的提升, 其函数值只依赖于随机变量的分布, 所以对任意 $Y \in$

$L^2(\Omega;\mathbb{R}^d)$,

$$\begin{aligned}
\tilde{f}(X+Y) &= \tilde{f}((X+Y)\circ S_\epsilon) \\
&= \tilde{f}(X_0 + (X\circ S_\epsilon - X_0) + Y\circ S_\epsilon) \\
&= \tilde{f}(X_0) + (\tilde{f})'(X_0)((X\circ S_\epsilon - X_0) + Y\circ S_\epsilon) \\
&\quad + \|(X\circ S_\epsilon - X_0) + Y\circ S_\epsilon\|_{L^2}\,\delta((X\circ S_\epsilon - X_0) + Y\circ S_\epsilon) \\
&= \tilde{f}(X_0) + \mathbb{E}[D\tilde{f}(X_0)\cdot((X\circ S_\epsilon - X_0) + Y\circ S_\epsilon)] \\
&\quad + \|(X\circ S_\epsilon - X_0) + Y\circ S_\epsilon\|_{L^2}\,\delta((X\circ S_\epsilon - X_0) + Y\circ S_\epsilon) \\
&= \tilde{f}(X) + \mathbb{E}[(D\tilde{f}(X_0)\circ T_\epsilon)\cdot Y] + \mathbb{E}[D\tilde{f}(X_0)\cdot(X\circ S_\epsilon - X_0)] \\
&\quad + \|(X\circ S_\epsilon - X_0) + Y\circ S_\epsilon\|_{L^2}\,\delta((X\circ S_\epsilon - X_0) + Y\circ S_\epsilon). \qquad (1.1)
\end{aligned}$$

其中 $\delta(\cdot)$ 与 ϵ 无关且 $\lim_{\|Y\|_{L^2}\to 0}\delta(Y)=0$, 第 1 个等号和第 5 个等号分别使用了 S_ϵ, T_ϵ 的保测性.

取 $\epsilon=\frac{1}{n}, n\in\mathbb{N}$, 记 $Z_n = D\tilde{f}(X_0)\circ T_{\frac{1}{n}}$. 由式 (3.1) 可得

$$\mathbb{E}[(Z_n - Z - m)\cdot Y] \le C(|\frac{1}{n}-\frac{1}{m}| + o(\|Y\|_{L^2})),$$

所以 $\{Z_n\}$ 会在 $L^2(\Omega;\mathbb{R}^d))$ 中收敛到某个随机变量, 最后在式 (3.1) 中令 $\epsilon\to 0$ 即可. ■

L-可微性的定义是将定义域从一个概率测度组成的空间转换到一个随机变量空间上讨论的, 但应该注意到 $f:\mathscr{P}_2\to\mathbb{R}$ 在一点处的局部性态应该只和该点和其周围的概率测度决定, 下面的定理将说明 L-可微性确实是函数的内蕴性质, 并对 Lions 导数给出一个稍微独立于随机变量的表示. 在此之前, 首先引入连续 L-可微的概念, 这是连续 Frechet 可微的自然延伸, 称 $f:\mathscr{P}_2\to\mathbb{R}$ 是连续 L-可微的, 是指:f 在任意 $\mu\in\mathscr{P}_2$ 处是 L-可微的, 且 $D\tilde{f}(X)$ 是以 X 为自变量的从 $L^2(\Omega;\mathbb{R}_d)$ 到自身的连续映射.

定理 1.3　设 $f:\mathscr{P}_2\to\mathbb{R}$ 是连续 L-可微的. 则对任意 $\mu\in\mathscr{P}_2$, 存在可测函数 $\xi:\mathbb{R}_d\to\mathbb{R}_d$, 使得任意以 μ 为分布的随机变量 X, 都有 $D\tilde{f}(X)=\xi(X)$.

证明　对给定的 X, 首先证明 $D\tilde{f}(X)$ 关于 X 生成的 σ 代数 $\sigma(X)$ 可测. 不妨设 f 有界. 证明分为以下几步.

(1) 先考虑 μ 关于 Lebsgue 测度绝对连续, 且对 $\forall q>4$,$\mu(|\cdot|^q)<\infty$. 对 $\epsilon>0$, 定义

$$\Psi(Y)=\tilde{f}(Y)+\frac{1}{2\epsilon}\mathbb{E}[|X-Y|^2]+\mathbb{E}[|Y|^4],\quad Y\in L^4(\Omega;\mathbb{R}^d),$$

则 Ψ 在 $L^4(\Omega;\mathbb{R}^d)$ 是 Frechet 可微的, 且任意 $Y\in L^4(\Omega;\mathbb{R}^d)$,

$$\Psi'(Y)=D\tilde{(f)}(Y)+\frac{1}{\epsilon}(Y-X)+4\mathbb{E}[|Y|^2]Y.$$

由于 Ψ 有下界, 从而一定有下确界, 取 $\{Z_n\} \subset L^4(\Omega;\mathbb{R}^d)$ 为 Ψ 的极小化序列, 即 $\lim_{n\to\infty}\Psi(Z_n) = \inf_{Z\in L^4(\Omega;\mathbb{R}^d)}\Psi(Z)$ 以及 $\nu_n = \mathscr{L}(\nu_n)$. 由 Brenier 定理 (陈述及证明可见 (Figalli et al., 2021) 的定理 2.5.10), 存在一列 $\mathbb{R}^d$ 上的凸函数 $\{\psi_n\}$ 满足 $\mathscr{L}(\nabla\psi(X_n)) = \nu_n$ 且是 μ—a.e. 可微的, 此外还有 $W_2(\mu,\nu_n) = \mathbb{E}[|X-\nabla\psi(X_n)|^2]$. 记 $Y_n = \nabla\psi(X_n)$, 则

$$\Psi(Y_n) = \tilde{f}(Y_n) + \frac{1}{2\epsilon}W_2(\mu,\nu_n)^2 + \mathbb{E}[|Y|^4] \leq \Psi(Z_n), \tag{1.2}$$

从而 $\{Y_n\}$ 也是 Ψ 的一个极小化序列.

而 $\sup_{n\geq 1}\nu_n(|\cdot|^4)\infty$, 结合 Chebyshev 不等式可得 $\{\nu_n\}$ 是胎紧的. 不妨设 ν_n 在 W_2 距离和弱收敛意义下收敛到 ν, 所以

$$\begin{aligned} f(\nu) &= \lim_{n\to\infty} f(\nu), \\ W_2(\mu,\nu) &= \lim_{n\to\infty} W_2(\mu,\nu_n). \end{aligned} \tag{1.3}$$

由 Fatou 引理的测度收敛版本 (见 (Bogachev, 2018) 的推论 2.2.6) 可得

$$\int_{\mathbb{R}^d}|x|^4\nu(\mathrm{d}x) \leq \liminf_{n\to\infty}\mathbb{E}[|Y|^4] < \infty.$$

再次利用 Brenier 定理可得, 存在凸函数 $\psi:\mathbb{R}^d\to\mathbb{R}$ 使得 $Y := \nabla\psi(X)$ 分布为 ν 且 $W_2(\mu,\nu) = \mathbb{E}[|X-Y|^2]$. 因为 Ψ 在 Y 处取得极小值, 由 (Ciarlet, 2017) 的定理 7.1-5 知 $D\Psi(Y) = 0$. 即

$$D\tilde{f}(Y) = -\frac{1}{\epsilon}(Y-X) - 4|Y|^2Y,$$

这说明 $D\tilde{f}(Y) \in \sigma(X)$.

注意到这里的 Y 是依赖于 ϵ 的, 不妨记为 Y_ϵ, 由于 Y_ϵ 极小化 Ψ_ϵ, 所以对任意 ϵ,

$$\mathbb{E}[|X-Y_\epsilon|^2] \leq 2(\tilde{f}(X) + \mathbb{E}[|X|^4] - (\tilde{f}(Y_\epsilon) + \mathbb{E}[|Y_\epsilon|^4])),$$

所以令 $\epsilon\to 0, Y_\epsilon\to X$. 这就证得了 $D\tilde{f}(X)\in\sigma(X)$.

(2) 假设 $\mu\in\mathscr{P}_2$ 关于 Lebsgue 测度绝对连续但不一定有 (1) 中的矩条件, 取分布为 μ 的随机变量 $X\in L^2(\Omega;\mathbb{R}^d)$, 令 $X_n = \frac{nX}{\sqrt{n^2+X^2}}$, 再令 $n\to\infty$ 即可.

(3) 若 $\mu\in\mathscr{P}_2$ 不是关于 Lebsgue 测度绝对连续的, 取分布为 μ 的随机变量 $X\in L^2(\Omega;\mathbb{R}^d)$ 和 Ω 上两个独立的标准正态分布的随机变量 N_1,N_2, 且 X 独立于 N_1,N_2. 令 $X_{n,i} = X + \frac{1}{n}N_i, n\in\mathbb{N}, i=1,2$, 则 $X_{n,i}$ 的分布总是关于 Lebsgue 测度绝对连续的, 从而可得 $D\tilde{f}(X)\in\sigma(X,N_i)$, 而由 N_1,N_2 的独立性, 可得 $D\tilde{f}(X)\in\sigma(X)$.

最后我们说明 ξ 不依赖于 X 的选取. 假设对两个分布为 μ 的二阶矩有限随机变量 X_1, X_2 分别有对应的 ξ_1, ξ_2, 任取 $A \in \mathscr{B}(\mathbb{R}^d)$, 易见 $\mu(\xi_1 I_A) = \mu(\xi_2 I_A)$, 所以 $\xi_1 = \xi_2, \mu$—a.e.. ■

有了以上定理, 我们可以借助概率空间和随机变量给出 L-可微函数 $f : \mathscr{P}_2 \to \mathbb{R}$ 在 μ_0 处的局部性态:

$$f(\mu) = f(\mu_0) + \mathbb{E}[D\tilde{f}(X_0) \cdot (X - X_0)] + o(\|X - X_0\|_{L^2}),$$

其中 $\mathscr{L}(X) = \mu, \mathscr{L}(X_0) = \mu_0$.

另一方面, 为了突出 L-可微的内蕴性, 我们将定理1.3中的 ξ 记为 $\partial_\mu f(\mu_0)$, 称为 f 在 μ_0 处的 Lions 导数或 L-导数. 此时有

$$f(\mu) = f(\mu_0) + \mathbb{E}[\partial_\mu f(\mu_0)(X_0) \cdot (X - X_0)] + o(\|X - X_0\|_{L^2}),$$

其中 $\mathscr{L}(X) = \mu, \mathscr{L}(X_0) = \mu_0$.

1.1.2　若干例子

例 1.1 (线性函数)　我们首先计算积分形式给出的线性函数. 给定光滑函数 $h : \mathbb{R}^d \to \mathbb{R}$, 且 $|\nabla h(x)| \leq C(1 + |x|)$. 令 $f(\mu) = \int_{\mathbb{R}^d} h(x)\mu(\mathrm{d}x)$,$f$ 的提升为 $\tilde{f}(X) = \mathbb{E}[h(X)]$, 其中 $\mathscr{L}(X) = \mu$. 计算 $\tilde{f}$ 的 Frechet 导数, 首先由中值定理,

$$h(x+y) = h(x) + \int_0^1 (\nabla h(x+ty) \cdot y)\mathrm{d}t, \tag{1.4}$$

$$\begin{aligned}\mathbb{E}[h(X+Y)] &= \mathbb{E}[h(X) + \int_0^1 (\nabla h(X+tY) \cdot Y)\mathrm{d}t] \\ &= \mathbb{E}[h(X)] + \mathbb{E}[\nabla h(X) \cdot Y] + \mathbb{E}[\int_0^1 (\nabla h(X+tY) \cdot Y - \nabla h(X) \cdot Y)\mathrm{d}t],\end{aligned} \tag{1.5}$$

其中,

$$
\begin{aligned}
&|\mathbb{E}[\int_0^1 ((\nabla h(X+tY)-\nabla h(X))\cdot Y)\,\mathrm{d}t]| \\
\leq &\int_0^1 \mathbb{E}[(((\nabla h(X+tY)-\nabla h(X))\cdot Y))\, I_{\{|Y|\leq \|Y\|_{L^2}^{1/2}\}}]\mathrm{d}t \\
&\quad + \int_0^1 \mathbb{E}[(((\nabla h(X+tY)-\nabla h(X))\cdot Y))\, I_{\{|Y|>\|Y\|_{L^2}^{1/2}\}}]\mathrm{d}t \\
\leq &\mathbb{E}[\sup_{y\leq \|Y\|_{L^2}^{1/2}} (\nabla h(X+y)-\nabla h(X))]\, \|Y\|_{L^2} \\
&\quad + C'\mathbb{E}[(1+|X|+|Y|)|Y|] \\
\leq &\mathbb{E}[\sup_{y\leq \|Y\|_{L^2}^{1/2}} (\nabla h(X+y)-\nabla h(X))]\, \|Y\|_{L^2} \\
&\quad + C'\, \|Y\|_{L^2} (\|Y\|_{L^2} + \|Y\|_{L^2}^{1/2} + \sup_{\mathbb{P}(A)\leq \|\|_{L^2} Y} \mathbb{E}[|X|^2 I_A]^{1/2})
\end{aligned}
\tag{1.6}
$$

所以 $\partial_\mu f(\mu)=\nabla h, \forall \mu\in\mathscr{P}_2$.

这个例子虽然简单, 但也说明了 Lions 导数和古典导数以及 Frechet 导数的一些区别: 尽管对线性函数, 其 L-导数仍然是固定的, 但并非 h, 而是其梯度.

例 1.2 (续例1.1, 卷积函数)　接下来考虑一个稍微复杂的函数, 令 $h*\mu(x)=\int_{\mathbb{R}^d} h(x-y)\mu(\mathrm{d}y)$,$g(\mu)=f(h*\mu)$. 易见 $\tilde{g}(X)=\int_{\mathbb{R}^d}\mathbb{E}[h(x-X)]\mathrm{d}\mathscr{L}(X)(x)$, 由第一章相关讨论可知, 当 $\tilde{g}$ 在一点 Frechet 可微时, 为了确定该点处的 Frechet 导数, 只需计算所有的 Gateaux 导数, 所以先计算 X 处沿 Y 方向的 Gateaux 导数,

$$
\begin{aligned}
&\tilde{g}(X+\epsilon Y)-\tilde{g}(X) \\
=&\int_{\mathbb{R}^d}\mathbb{E}[h(x-(X+\epsilon Y))]\mathrm{d}\mathscr{L}(X+\epsilon Y)(x)-\int_{\mathbb{R}^d}\mathbb{E}[h(x-X)]\mathrm{d}\mathscr{L}(X)(x) \\
=&\left(\int_{\mathbb{R}^d}\mathbb{E}[h(x-X-\epsilon Y)]\mathrm{d}\mathscr{L}(X+\epsilon Y)(x)-\int_{\mathbb{R}^d}\mathbb{E}[h(x-X)]\mathrm{d}\mathscr{L}(X+\epsilon Y)(x)\right) \\
&\quad+\left(\int_{\mathbb{R}^d}\mathbb{E}[h(x-X)]\mathrm{d}\mathscr{L}(X+\epsilon Y)(x)-\int_{\mathbb{R}^d}\mathbb{E}[h(x-X)]\mathrm{d}\mathscr{L}(X)(x)\right) \\
\triangleq& I_1+I_2,
\end{aligned}
\tag{1.7}
$$

对于第二项, 结合 Fubini 定理交换积分次序可得

$$
\lim_{\epsilon\to 0}\frac{I_2}{\epsilon}=\int_{\mathbb{R}^d}\mathbb{E}[\nabla h(x-X)\cdot Y]\mathrm{d}\mathscr{L}(X)(x),
$$

而对于第一项, 取 (X,Y) 的一个独立复制 $(\hat{X},\hat{Y})$, 则

$$\begin{aligned}I_1 &= \mathbb{E}[h((\hat{X}+\epsilon\hat{Y})-(X+\epsilon Y))]-\mathbb{E}[h(\hat{X}+\epsilon\hat{Y}-X)]\\ &= \mathbb{E}[h((\hat{X}+\epsilon\hat{Y})-(X+\epsilon Y))]-\mathbb{E}[h(\hat{X}-X)]\\ &\quad+\mathbb{E}[h(\hat{X}-X)]-\mathbb{E}[h(\hat{X}+\epsilon\hat{Y}-X)],\end{aligned}\tag{1.8}$$

所以

$$\lim_{\epsilon\to0}\frac{I_2}{\epsilon}=\mathbb{E}[\nabla h(\hat{X}-X)\cdot(\hat{Y}-Y)]-\mathbb{E}[\nabla h(\hat{X}-X)\cdot(\hat{Y})]=-\mathbb{E}[\nabla h(\hat{X}-X)\cdot Y].$$

综上可得,$\partial_\mu g(\mu)=(\nabla(h+\bar{h}))*\mu$, 其中 $\bar{h}(x):=h(-x)$.

例 1.3　考虑连续函数 $v:\mathbb{R}^d\times\to\mathscr{P}_2$ 满足:

(1) $\forall\mu\in\mathscr{P}_2$,$v(x,\mu)$ 关于 x 可微, 记 $\nabla v(x,\mu)$ 表示 v 在 (x,μ) 处关于 x 的梯度. 任意有界集 $K\subset\mathscr{P}_2$, 存在常数 C_1, 使得任意 $\mu\in K,|\nabla v(x,\mu)|\le C_1(1+|x|)$.

(2) 任意 $\mu\in\mathscr{P}_2,x\in\mathbb{R}^d$, 使得 $(x,y)\mapsto$ 可测, 且 $\partial_\mu v(x,\mu)(y)$ 关于 y 连续, 并且对任意有界集 $K\subset\mathscr{P}_2$, 存在常数 C_2, 使得任意 $\mu\in K$,$\partial_\mu v(x,\mu)(y)\le C_2(1+|y|)$.

令 $u(\mu)=\int_{\mathbb{R}^d}v(x,\mu)\mathrm{d}\mu(x)$, 计算 u 的 Lions 导数, 计算中由于涉及到多个随机变量, 因此在合适的时候应该回到函数的定义以免造成混淆. 首先注意到 u 的提升 $\tilde{u}(X)=\mathbb{E}[v(X,\mathscr{L}(X))]$,

$$\begin{aligned}&\tilde{u}(X+Y)-\tilde{u}(X)\\ =&\mathbb{E}[v(X+Y,\mathscr{L}(X+Y))]-\mathbb{E}[v(X,\mathscr{L}(X))]\\ =&\mathbb{E}[v(X+Y,\mathscr{L}(X+Y))]-\mathbb{E}[v(X,\mathscr{L}(X+Y))]\\ &\quad+\mathbb{E}[v(X,\mathscr{L}(X+Y))]-\mathbb{E}[v(X,\mathscr{L}(X))]\\ \triangleq&I_1+I_2,\end{aligned}\tag{1.9}$$

由中值定理可得,

$$I_1=\mathbb{E}[\int_0^1\nabla v(X+tY,\mathscr{L}(X+Y))\cdot Y\mathrm{d}t)],$$

而

$$\begin{aligned}I_2&=\int_{\mathbb{R}^d}(v(x,\mathscr{L}(X+Y))-v(x,\mathscr{L}(X)))\mathrm{d}\mathscr{L}(X)(x)\\ &=\int_{\mathbb{R}^d}\int_0^1\mathbb{E}[\partial_\mu v(x,\mathscr{L}(X+tY))(X+tY)\cdot Y]\mathrm{d}t\mathrm{d}\mathscr{L}(X)(x),\end{aligned}$$

所以可以合理推测 $\partial_\mu f(\mu)(\cdot)=\nabla v(\cdot,\mu)+\int_{\mathbb{R}^d}\partial_\mu f(y,\mu)(\cdot)\mathrm{d}\mu(y)$, 事实上根据对 v 的假设, 采用和例1.1中类似的讨论方法可知,f 在 μ 处的 L-导数确实为上式.

例 1.4 (续1.3)　设 $K:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ 为一可测映射,∇_1K,∇_2K 分别表示对第 1 个和第 2 个自变量的梯度, 且都是至多线性增长的, 令 $v(x,\mu)=$

$\int_{\mathbb{R}^d} K(x,y)\mathrm{d}\mu(y)$, 则由上例结果可得

$$\partial_\mu f(\mu)(x) = \int_{\mathbb{R}^d} (\nabla_1 K(x,y) + \nabla_1 K(y,x))\mathrm{d}\mu(y).$$

1.2　链式法则和 Itô 公式

给定一个完备的带流概率空间 $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ 和其上的一个 m 维的 $\mathscr{F}_t-$ 布朗运动 $(B_s)_{s\geq 0}$, 设 $\mathscr{F}_t$ 是右连续的, 即 $\mathscr{F}_{t+} := \cap_{s>t}\mathscr{F}_s = \mathscr{F}_t$,$(b_t)_{t\geq 0}, (\sigma_t)_{t\geq 0}$ 分别为 Ω 上 $\mathbb{R}^d$ 和 $\mathbb{R}^{d\times m}$ 值的循序可测过程,$X_0 \in L^2(\Omega, \mathscr{F}_0, \mathbb{P}; \mathbb{R}^d)$, 且对任意 $T>0$ 都有

$$\mathbb{E}[\int_0^T (|b_t|^2 + \|\sigma_t\|_{HS}^4)\mathrm{d}t] < \infty,$$

其中 $\|\sigma_s\|_{HS} := \sqrt{\mathrm{tr}(a_s)} := \sqrt{\mathrm{tr}(\sigma_s\sigma_s^*)}$. 令

$$X_t = X_0 + \int_0^t b_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}B_s, \quad t \geq 0. \tag{1.10}$$

对于式 (3.10) 给出的 Ito 过程和任一 $\mathbb{R}^d$ 上的二次连续可微函数 f, 都有 Ito 公式

$$f(X_t) = f(X_0) + \int_0^t \nabla f(X_s)\cdot b_s \mathrm{d}s + \int_0^t \nabla f(X_s)\sigma_s \mathrm{d}B_s + \frac{1}{2}\int_0^t \mathrm{tr}(a_s \nabla^2 f(X_s))\mathrm{d}s \tag{1.11}$$

我们希望对式 (3.10) 给出的过程的时间边缘分布 $\boldsymbol{\mu} := (\mu_t)_{t\geq 0} := (\mathscr{L}(X_t))_{t\geq 0}$ 和一类性质足够好的 $f: \mathscr{P}_2 \to \mathbb{R}$ 也能建立类似的公式表示 $f(\mu_t)$.

在此之前,, 我们先考虑一类较简单的复合函数及求导. 首先注意到任意概率测度的凸组合仍为概率测度, 所以对函数 $f: \mathscr{P}_2 \to \mathbb{R}^d$ 可以定义它的经验投影

$$\begin{aligned} f^N: \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{N\text{个}} = (\mathbb{R}^d)^N &\quad\longrightarrow\quad \mathbb{R} \\ (x^1, \cdots, x^N) &\quad\longmapsto\quad f(\frac{1}{N}\sum_{l=1}^N \delta_{x^i}), \end{aligned} \tag{1.12}$$

其中 $N \in \mathbb{N}$,$\delta_x(A) = I_A(x), x \in \mathbb{R}^d, A \in \mathscr{B}(\mathbb{R}^d)$ 表示 x 处的 Dirac 质量.

对于经验投影的可微性和求导公式有以下命题.

定理 1.4　设 $f: \mathscr{P}_2 \to \mathbb{R}$ 是连续 L-可微的, 则经验投影 f^N 是 $(\mathbb{R}^d))^N$ 上的可微函数, 且

$$\partial_{x^i} f^N(x^1, \cdots, x^N) = \frac{1}{N}\partial_\mu f(\frac{1}{N}\sum_{l=1}^N \delta_{x^i})(x^i), i = 1, \cdots, N.$$

注　此处的 ∂_{x^i} 表示沿 x^i 的方向导数算子.

证明　设 $\theta: \Omega \to \{1, \cdots, N\}$ 为均匀分布. 则给定 $\boldsymbol{x} = (x^1, \cdots, x^N) \in (\mathbb{R}^d)^N$, 和常值随机变量 $X_i^{\boldsymbol{x}} = x^i, i = 1, \cdots, N$,$X_\theta^{\boldsymbol{x}}$ 是一个取值于 $\{x^1, \cdots, x^N\}$ 的随机变

量, 且对任意 $A \in \mathscr{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\mathbb{P}(X_\theta^{\boldsymbol{x}} \in A) &= \sum_{l=1}^N \mathbb{P}((X_\theta^{\boldsymbol{x}} \in A, \theta = l) \\
&= \sum_{l=1}^N \mathbb{P}(X_\theta^{\boldsymbol{x}} \in A | \theta = l)\mathbb{P}(\theta = l) \\
&= \frac{1}{N}\sum_{l=1}^N \mathbb{P}(X_l^{\boldsymbol{x}} \in A) \\
&= \frac{1}{N}\sum_{l=1}^N \delta_{x^i}(A),
\end{aligned} \tag{1.13}$$

也即 $\mathscr{L}(X_\theta^{\boldsymbol{x}}) = \frac{1}{N}\sum_{l=1}^N \delta_{x^i}$. 所以有定义,

$$\begin{aligned}
f^N(\boldsymbol{x} + \boldsymbol{h}) &= f(\frac{1}{N}\sum_{l=1}^N \delta_{x^l + h^l}) \\
&= \tilde{f}(X_\theta^{\boldsymbol{x}} + X_\theta^{\boldsymbol{h}}) \\
&= \tilde{f}(X_\theta^{\boldsymbol{x}}) + \mathbb{E}[D\tilde{f}(X_\theta^{\boldsymbol{x}}) \cdot X_\theta^{\boldsymbol{h}}] + o(\|X_\theta^{\boldsymbol{h}}\|_{L^2}) \\
&= f^N(\boldsymbol{x}) + \mathbb{E}[\partial_\mu f(\frac{1}{N}\sum_{l=1}^N \delta_{x^i})(X_\theta^{\boldsymbol{x}}) \cdot X_\theta^{\boldsymbol{h}}] + o(\|X_\theta^{\boldsymbol{h}}\|_{L^2}),
\end{aligned} \tag{1.14}$$

注意到 $\partial_\mu f(\frac{1}{N}\sum_{l=1}^N \delta_{x^i})(X_\theta^{\boldsymbol{x}}) \cdot X_\theta^{\boldsymbol{h}}$ 是只取有限个值的随机变量, 所以其期望是容易计算的, 计算后即可得到定理中的导数公式. ■

1.2.1　完全 C^2 正则性

由定理1.3, 当 $f : \mathscr{P}_2 \to \mathbb{R}$ 是连续 L-可微时, 对任一 $\mu \in \mathscr{P}_2$ 存在一个可测映射 $\partial_\mu f(\mu)(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$, 使得当 $\mathscr{L}(X) = \mu, D\tilde{f}(X) = \partial_\mu f(\mu)(X)$, 且该可测映射在 μ—a.e. 意义下是唯一的. 我们希望 $(\mu, x) \mapsto \partial_\mu f(\mu)(x)$ 是一个从 $\mathscr{P}_2 \times \mathbb{R}^d$ 上的性质充分好的函数, 为此我们引入以下假设.

完全 C^2 正则性 (Full C^2 Regularity) 假设

(A1) $f : \mathscr{P}_2 \to \mathbb{R}$ 是连续 L-可微的, 即 $\tilde{f}$ 是连续 Frechet 可微的, 且存在合适版本的一阶 L-导数使得映射 $(\mu, x) \mapsto \partial_\mu f(\mu)(x)$ 是 $\mathscr{P}_2 \times \mathbb{R}^d$ 到 $\mathbb{R}^d$ 的连续映射;

(A2) 对于 (A1) 中的一阶 L-导数版本, 给定任一 $\mu \in \mathscr{P}_2$, 映射 $x \mapsto \partial_\mu f(\mu)(x)$ 是可微的, 记 $\partial_\mu f(\mu)(x)$ 关于空间变量 x 的微分为 $\partial_x \partial_\mu f(\mu)(x)$, 映射 $(\mu, x) \mapsto \partial_x \partial_\mu f(\mu)(x)$ 是 $\mathscr{P}_2 \times \mathbb{R}^d$ 到 $\mathbb{R}^{d\times d}$ 的连续映射;

(A3) 对于 (A1) 中的一阶 L-导数版本, 给定任意 $x \in \mathbb{R}^d$, 映射 $\mu \mapsto \partial_\mu f(\mu)(x)$ 的各分量映射是 L-可微的, 也称映射 $\mu \mapsto \partial_\mu f(\mu)(x)$ 是 L-可微的, 记为 $\partial_\mu^2 f(\mu)(x, \cdot) :== \partial_\mu^2 f(\mu)(x)(\cdot)$, 且映射 $(\mu, x, y) \mapsto \partial_\mu^2 f(\mu)(x, y)$ 是 $\mathscr{P}_2 \times \mathbb{R}^d \times \mathbb{R}^d$ 上的连续映射.

■

注

(1) 粗略地讲, 完全 C^2 正则性对应于多元微积分中的 C^2 可微概念, 即要求二阶导数存在并连续;

(2) 完全 C^2 正则性假设中满足条件的映射 $\partial_\mu f(\mu)(x), \partial_x\partial_\mu f(\mu)(x), \partial_\mu^2 f(\mu)(x,y)$ 是存在且唯一的, 具体证明参见 (Carmona et al., 2018) 的注 5.82;

(3) 与多元微积分中的梯度,Jacobi 矩阵等保持一致, 对 $\mu \in \mathscr{P}_2(\mathbb{R}^d)$,$x, y \in \mathbb{R}^d$, 将 $\partial_x\partial_\mu f(\mu)(x)$ 表示为 d 维方阵, 第 i 行, 第 j 列元素为

$$(\partial_x\partial_\mu f(\mu)(x))_{i,j} = \partial_{x_j}(\partial_\mu f(\mu))^{(i)}(x),$$

以及将 $\partial_\mu^2 f(\mu)(x,y)$ 表示为 d 维方阵, 第 i 行, 第 j 列元素为

$$(\partial_\mu^2 f(\mu)(x,y))_{i,j} = (\partial_\mu((\partial_\mu f(\mu))(x))^{(i)}(y))^{(j)}.$$

在将 $\mathbb{R}^d$ 中的元素 z 看作列向量的情况下,$\partial_x\partial_\mu f(\mu)(x)z, \partial_\mu^2 f(\mu)(x,y)z$ 均为由通常矩阵乘法的 $d\times 1$ 列向量; 而对于 d 维方阵 A,$\langle \partial_x\partial_\mu f(\mu)(x), A\rangle = \partial_x\partial_\mu f(\mu)(x)\cdot A = \mathrm{tr}(\partial_x\partial_\mu f(\mu)(x)A^T)$ 表示矩阵内积,$\partial_\mu^2 f(\mu)(x,y)\cdot A$ 类似.

(4) 在无特殊说明下, 无前缀的 C^2 正则性是指完全 C^2 正则性.

(5) 若函数 $f : \mathscr{P}_2 \to \mathbb{R}^d$ 满足完全 C^2 正则性假设中的 (A1),(A2) 和 (A3), 则称 f 是 (完全)C^2 正则的或 f 具有 (完全)C^2 正则性.

对于 C^2 正则的函数 f, 在此陈述一些相关性质.

首先是定理1.4中经验投影的可微性的加强, 有如下定理.

定理 1.5　设 $f : \mathscr{P}_2 \to \mathbb{R}$ 是完全 C^2 正则的. 则对任意 $N \in \mathbb{N}$, 经验投影 $f^N : (\mathbb{R}^d)^N \to \mathbb{R}$ 是经典意义下 C^2 的, 且对任意 $\boldsymbol{x} = (x^1, \cdots, x^N) \in (\mathbb{R}^d)^N$, 和 $1 \le i, j \le N$,

$$\partial^2_{x^i x^j} f^N(\boldsymbol{x}) = \frac{1}{N}\partial_x\partial_\mu f(\frac{1}{N}\sum_{l=1}^N \delta_{x^l})(x_i)I_{i=j} + \frac{1}{N^2}\partial_\mu^2 f(\frac{1}{N}\sum_{l=1}^N \delta_{x^l})(x^i, x^j).$$

1.2.2　Ito 公式

对于1.10定义的 Ito 过程和性质足够好的函数 $f : \mathscr{P}_2 \to \mathbb{R}$, 下面的定理给出了其时间边缘分布满足的 Ito 公式.

定理 1.6　设 $f : \mathscr{P}_2 \to \mathbb{R}$ 是 C^2 正则的, 且对任意 $\mathscr{P}$ 的紧子集 K, 存在常数 $C = C(K)$, 使得任意 $\mu \in K$,

$$\int_{\mathbb{R}^d} |\partial_x\partial_\mu f(\mu)(x)|^2 \mathrm{d}\mu(x) < C.$$

则对任意 $t \geq 0$,

$$f(\mu_t) = f(\mu_0) + \int_0^t \mathbb{E}[\partial_\mu f(\mu_s)(X_s) \cdot b_s]\mathrm{d}s + \frac{1}{2}\int_0^t \mathbb{E}[\partial_x \partial_\mu f(\mu_s)(X_s) \cdot a_s]\mathrm{d}s. \tag{1.15}$$

证明　证明主要分为三步.

第一步:

设 f, 其一阶导数和其二阶导数都是有界且一致连续的. 且过程 $(b_s), (\sigma_s)$ 是有界的.

设 $((b_s^l, \sigma_s^l, B_s^l), X_0^l)_(l \in \mathbb{N})$ 是 $((b_s, \sigma_s, B_s), X_0)$ 的一列独立复制, 则

$$X_t^l = X_0^l + \int_0^t b_s^l \mathrm{d}s + \int_0^t \sigma_s^l \mathrm{d}B_s^l, \quad t \geq 0, l \geq 1,$$

则 $(X_t), (X_t^1), (X_t^2), \cdots$ 独立同分布.

考虑 f 到 $(\mathbb{R}^d)^N$ 的经验投影, 由定理1.5可知 f^N 是 C^2 的, 所以可以应用 Ito 公式1.11, 简记 $\boldsymbol{X}_t = (X_t^1, \cdots, X_t^N)$ 以及 $\bar{\mu}_t^N = \frac{1}{N}\sum_{l=1}^N \delta_{X_t^l}$; 此外, 对两个独立的布朗运动 $B^i = (B_s^i), B^j = (B_s^j)$ 是正交的, 即 $\langle B^i, B^j \rangle = 0$, 所以

$$\begin{aligned}
& f^N(\boldsymbol{X}_t) - f^N(\boldsymbol{X}_0) \\
=& \sum_{l=1}^N \int_0^t \partial_{x^i} f^N(\boldsymbol{X}_s) \cdot b_s^l \mathrm{d}s + \sum_{l=1}^N \int_0^t \partial_{x^i} f^N(\boldsymbol{X}_s) \cdot (\sigma_s^l \mathrm{d}B_s^l) \\
& + \sum_{l=1}^N \partial_{x^l x^l}^2 f^N(\boldsymbol{X}_s) \cdot a_s^l \mathrm{d}s \\
=& \frac{1}{N}\sum_{l=1}^N \int_0^t \partial_\mu f(\bar{\mu}_t^N)(X_s^l) \cdot b_s^l \mathrm{d}s + \sum_{l=1}^N \int_0^t \partial_\mu f(\bar{\mu}_t^N)(X_s^l) \cdot (\sigma_s^l \mathrm{d}B_s^l) \\
& + \frac{1}{2N}\sum_{l=1}^N \int_0^t \partial_x \partial_\mu f(\bar{\mu}_s^N)(X_s^l) \cdot a_s^l \mathrm{d}s \\
& + \frac{1}{2N^2}\int_0^t \partial_\mu^2 f(\bar{\mu}_s^N)(X_s^l, X_s^l) \cdot a_s^l \mathrm{d}s,
\end{aligned} \tag{1.16}$$

注意到 $\bar{\mu}_t^N$ 实际上是一个随机概率测度, 所以上式应理解为在一个 $\mathbb{P}-$ 零测集外逐点意义下成立.

对上式左右两边同时取期望, 并结合独立同分布的条件,

$$\begin{aligned}&\mathbb{E}[f(\bar{\mu}_t^N)]-\mathbb{E}[f(\bar{\mu}_0^N)]\\&=\int_0^t\mathbb{E}[\partial_\mu f(\bar{\mu}_t^N)(X_s)\cdot b_s]\mathrm{d}s\\&\quad+\frac{1}{2}\int_0^t\mathbb{E}[\partial_x\partial_\mu f(\bar{\mu}_s^N)(X_s)\cdot a_s]\mathrm{d}s\\&\quad+\frac{1}{2N}\int_0^t\mathbb{E}[\partial_\mu^2 f(\bar{\mu}_s^N)(X_s,X_s)\cdot a_s]\mathrm{d}s\\&\triangleq I_1+I_2+I_3,\end{aligned}\tag{1.17}$$

根据有界性的假设, 当 $N\to\infty$ 时,$I_3\to 0$; 同样是由于有界性, 应用 (Fournier et al., 2015) 的定理 1 以及强大数定律, 可得任意 $s\in[0.t]$,$W_2(\mu_s^N,\mu_s)$ 以概率 1 收敛到 0. 由 f 的 C^2 正则性, 可得当 $N\to\infty$,$\mathbb{E}[f(\bar{\mu}_s^N)]\to f(\mu)$.

第二步:

这一步证明去掉对 f 及其导数的有界性和一致连续性后, 公式 (3.15) 仍然成立. 首先我们承认以下引理:

引理 1.7　设 $f:\mathscr{P}_2\to\mathbb{R}$ 是 C^2 正则的,$\rho:\mathbb{R}^d\to\mathbb{R}^d$ 为光滑紧支映射. 则函数

$$\begin{aligned}f^*\rho:\mathscr{P}_2&\to\mathbb{R}\\\mu&\mapsto f(\mu\circ\rho)\end{aligned}\tag{1.18}$$

是完全 C^2 正则的, 且 $f^*\rho$ 及其一阶和二阶导数都是有界和一致连续的.

取一列紧支光滑映射 $\{\rho_n:\mathbb{R}^d\to\mathbb{R}^d\}$ 使得在紧集上,$(\rho_n(x),D\rho_n(x),D^2\rho_n(x))$ 一致收敛到 $(x,I_d,0)$. 不妨设存在常数 C 使得任意 $n\geq 1,x\in\mathbb{R}^d$,

$$\begin{aligned}|\rho_n(x)|&\leq C|x|,\\|D\rho_n(x)|&\leq C,\\|D^2\rho_n(x)|&\leq C.\end{aligned}$$

■

参考文献

Ciarlet P G, 2017. 线性与非线性泛函分析及其应用（下册）[M]. 秦铁虎, 译. 北京: 高等教育出版社. 5

Bogachev, 2007. Measure theory[M]. Springer Berlin, Heidelberg. 1, 2

Bogachev V I, 2018. Weak convergence of measures[M]. Providence, Rhode Island : American Mathematical Society. 5

Carmona R, Delarue F, 2018. Probabilistic theory of mean field games with applications I[M]. Springer Cham. 1, 11

Figalli A, Glaudo F, 2021. An invitation to optimal transport, wasserstein distances, and gradient flows[M]. EMS Press. 5

Fournier N, Guillin A, 2015. On the rate of convergence in wasserstein distance of the empirical measure[J]. Probability Theory and Related Fields. 13