

# 随机分析中的几种导数

姓名： 张竣淞

学号： 2024233014

时间： 二〇二五年十二月二十日

# **On Several Types of Derivatives in Stochastic Analysis**

Author: Zhang Junsong

Completion date: December 20, 2025

## 摘要

本文从泛函分析中的 Frechet 导数出发，对随机分析中常用的几种导数及其性质进行简单的介绍，包括定义、链式法则和对 Ito 公式的扩展。

**关键词：**随机分析，Frechet 导数，Lions 导数，内蕴导数，外在导数，Malliavin 导数

## 目录

第 1 章 导论 .....	1
第 2 章 赋范向量空间中的 Fréchet 导数和 Gâteaux 导数 .....	3
2.1 Fréchet 导数和 Gâteaux 导数的定义和例子 .....	3
2.1.1 Fréchet 导数的定义 .....	3
2.1.2 Gâteaux 导数的定义 .....	6
2.1.3 若干例子 .....	6
2.2 链式法则 .....	8
2.3 高阶导数和 Taylor 展开 .....	9
第 3 章 $\mathcal{P}_2$ 上的函数的 Lions 导数 .....	13
3.1 L-可微性 .....	14
3.1.1 定义 .....	14
3.1.2 若干例子 .....	18
3.2 链式法则和 Itô 公式 .....	21
3.2.1 完全 $C^2$ 正则性 .....	22
3.2.2 Ito 公式 .....	24
参考文献 .....	28

## 第1章 导论

在古典微积分中, 对于一个性质足够好的函数

$$f : \mathbb{R} \rightarrow \mathbb{R},$$

可以从两种观点去看待它在  $x \in \mathbb{R}$  处的局部性态, 一种是差商的极限, 称  $f$  在  $x$  处可导, 是指极限:

$$f'(x) := \lim_{y \rightarrow 0} \frac{f(x+y) - f(x)}{y},$$

存在. 另一种则是用线性映射局部逼近, 即称  $f$  在  $x \in \mathbb{R}$  处可微, 是指: 存在一个连续线性映射

$$L_x : \mathbb{R} \rightarrow \mathbb{R},$$

使得当  $y \rightarrow 0$  时,

$$\delta(y) = f(x+y) - f(y) - L_x(y) \rightarrow 0.$$

并且由于  $\mathbb{R}$  上连续泛函全体  $\mathcal{L}(\mathbb{R}; \mathbb{R})$  同构于  $\mathbb{R}$ , 所以这两种角度实际上是一致的.

当考虑  $f : \mathbb{R}^d \rightarrow \mathbb{R}, d \geq 2$  时,  $\mathbb{R}^d$  中元素之间不存在天然的除法, 差商的极限的观点不能完全推广, 只能考虑方向导数, 即固定  $v \in \mathbb{R}^d$ , 考察

$$\partial_v f(x) := \lim_{h \rightarrow 0 \in \mathbb{R}} \frac{f(x+hv) - f(x)}{h},$$

上面的极限如果存在, 则称为  $f$  沿  $v$  的方向导数. 而线性逼近的观点则可以完全推广, 定义完全一致, 即称  $f$  在  $x \in \mathbb{R}^d$  处可微, 是指: 存在一个连续线性映射

$$L_x : \mathbb{R}^d \rightarrow \mathbb{R},$$

使得当  $|y| \rightarrow 0$  时,

$$\delta(y) = f(x+y) - f(y) - L_x(y) \rightarrow 0.$$

由于  $\mathbb{R}^d$  上连续泛函全体  $\mathcal{L}(\mathbb{R}^d; \mathbb{R})$  同构于  $\mathbb{R}^d$ , 该线性映射在  $\mathbb{R}$  中的对应表示就是

$$\nabla f(x) := (\partial_{e_1} f(x), \dots, \partial_{e_d} f(x)),$$

其中  $e_i$  表示  $\mathbb{R}^d$  中第  $i$  个分量为 1, 其余分量为 0 的元素,  $i = 1, \dots, d$ . 虽然这两种观点依然有紧密的联系, 但由于空间几何性质较一维情形复杂很多, 所以两者并不完全等同, 具体地说: 如果  $f$  在一点可微, 则该点处各方向导数均存在; 但各方向导数存在并不能推出可微性, 需要对方向导数添加额外的连续性要求.

当考虑更一般的赋范向量空间之间的映射时,可以继承这两种观点去考察映射的局部性态:从线性逼近的角度推广得到 Fréchet 导数的概念,而从差商极限或方向导数的角度推广得到 Gâteaux 导数的概念,这是第二章的主要内容;而在此基础上,出于随机分析研究的需要,第三章将引入 L-可微和 Lions 导数的概念,并介绍其在分布依赖型随机微分方程或称 McKean-Vlasov 型随机微分方程中的初步应用.

## 第2章 赋范向量空间中的 Fréchet 导数和 Gâteaux 导数

本章内容主要参考引用自 (Ciarlet, 2025) 的 Chapter 9: Differential Calculus in Normed Vector Spaces 及其第一版的中译本 (Ciarlet, 2017b) 的第七章相关内容。

如无特殊说明, 本章涉及  $X, Y$  为 Banach 空间, 即完备赋范向量空间,  $\|\cdot\|_X, \|\cdot\|_Y$  分别为  $X, Y$  上的完备范数, 在不致混淆的前提下, 均简记为  $\|\cdot\|; 0_X, 0_Y$  分别表示  $X, Y$  中的零元, 在不致混淆的前提下均简记为  $0; X$  上拓扑一般为该完备范数诱导的拓扑,  $\Omega$  为  $X$  中在范数拓扑下的开集.  $\mathcal{L}(X; Y)$  表示  $X$  到  $Y$  的有界线性算子全体,  $A \in \mathcal{L}(X; Y)$  的算子范数为

$$\|A\|_{\mathcal{L}(X; Y)} := \sup_{x \in X: \|x\|_X > 0} \frac{\|Ax\|_Y}{\|x\|_X}.$$

$\mathcal{L}_k(X; Y)$  表示从  $\underbrace{X \times \cdots \times X}_{k\text{个}}$  到  $Y$  的连续  $k-$  线性映射全体. 对赋范向量空间  $X$  和给定的  $a, b \in X$ , 定义开区间  $(a, b) := \{ta + (1 - t)b : 0 < t < 1\}$  和闭区间  $[a, b] := \{ta + (1 - t)b : 0 \leq t \leq 1\}$

### 2.1 Fréchet 导数和 Gâteaux 导数的定义和例子

#### 2.1.1 Fréchet 导数的定义

**定义 2.1** 给定映射  $f : \Omega \rightarrow Y, x \in \Omega$ , 称  $f$  在  $x$  处 Frechet 可微, 是指: 存在  $A \in \mathcal{L}(X; Y)$ , 使得当  $h \rightarrow 0_X$ , 时

$$\delta(h) := \frac{f(x + h) - f(x) - Ah}{\|h\|_X} \rightarrow 0_Y. \quad (2.1)$$

记此连续线性算子  $A$  为  $f'(x)$  或  $df(x)$ , 称为  $f$  在  $x$  处的 Fréchet 导数.

根据定义, 可以得到 Frechet 导数以下两个基本性质.

**命题 2.1** 给定映射  $f : \Omega \rightarrow Y, x \in \Omega$ , 若  $f'(x)$  存在, 则  $f'(x)$  是唯一的.

**证明** 假设存在  $A_1, A_2 \in \mathcal{L}(X; Y)$  均满足定义2.1中的要求. 由于  $\Omega$  为开集, 存在  $r > 0$  使得  $B(x, r) \subset \Omega$ . 对任意  $h \in B(0, r)$ ,

$$\begin{aligned} f(x + h) &= f(x) + A_1 h + \|h\| \delta_1(h) \\ &= f(x) + A_2 h + \|h\| \delta_2(h), \end{aligned} \quad (2.2)$$

所以对任意  $h \in B(0, r)$ ,

$$\|(A_1 - A_2)h\| = \|h\| \|\delta_1(h) - \delta_2(h)\|, \quad (2.3)$$

对任意  $u \in X$ , 存在充分大的  $R > 0$  使得  $\frac{1}{R}u \in B(0, r)$ , 则

$$\begin{aligned}\|(A_1 - A_2)u\| &= R \left\| (A_1 - A_2)\left(\frac{u}{R}\right) \right\| \\ &\leq R \left\| \frac{u}{R} \right\| \left\| \delta_1\left(\frac{u}{R}\right) - \delta_2\left(\frac{u}{R}\right) \right\| \\ &= \|u\| \left\| \delta_1\left(\frac{u}{R}\right) - \delta_2\left(\frac{u}{R}\right) \right\|. \end{aligned}\quad (2.4)$$

令  $R \rightarrow \infty$  即可. ■

**命题 2.2** 若  $f : \Omega \rightarrow Y$  在  $x \in \Omega$  处 Frechet 可微, 则  $f$  在  $x$  处连续.

**定义 2.2** 称  $f : \Omega \rightarrow Y$  在  $\Omega$  内可微, 是指: 任意  $x \in \Omega$ ,  $f$  在  $x$  处可微. 此时定义映射

$$\begin{aligned}f' : \Omega &\rightarrow \mathcal{L}(X; Y) \\ x &\mapsto f'(x).\end{aligned}$$

若  $f'$  是连续的, 则称  $f$  在  $\Omega$  内连续可微, 简称在  $\Omega$  内是  $C^1$  的.

若  $f$  为单射,  $f(\Omega)$  为  $Y$  中开集且  $f^{-1} : f(\Omega) \rightarrow X$  在  $f(\Omega)$  内是  $C^1$  的, 则称  $f$  是一个  $C^1$  微分同胚.

**注** 记  $C^1(\Omega; Y) := \{f : \Omega \rightarrow Y : f \text{ 在 } \Omega \text{ 内是 } C^1 \text{ 的}\}$ , 特别地, 当  $Y = \mathbb{R}$  时, 记  $C^1(\Omega) = C^1(\Omega; \mathbb{R})$ . 容易验证,  $C^1(\Omega; Y)$  是向量空间.

类似于欧式空间, 考虑  $X$  或  $Y$  为有限个 Banach 空间的乘积空间的情形, 对乘积空间配备最大值范数, 即考虑  $X = X_1 \times \cdots \times X_n$ , 其中  $X_i, 1 \leq i \leq n$  为 Banach 空间, 对  $x = (x_1, \dots, x_n) \in X$ , 定义  $\|x\|_\infty = \max_{1 \leq i \leq n} \|x_i\|_{X_i}$ . 乘积空间在最大值范数下仍为 Banach 空间且最大值范数诱导了乘积拓扑 (见 (Ciarlet, 2017a) 的 2.2 节).

**定理 2.3** 设  $Y_1, \dots, Y_m$  为 Banach 空间,  $Y = Y_1 \times \cdots \times Y_m$  并配备最大值范数. 给定映射  $f_i : \Omega \rightarrow Y, 1 \leq i \leq m$  和映射

$$\begin{aligned}f : \Omega &\rightarrow Y = Y_1 \cdots Y_m \\ x &\mapsto (f_1(x), \dots, f_m(x)).\end{aligned}\quad (2.5)$$

则  $f$  在  $a \in \Omega$  处可微等价于任意  $1 \leq i \leq m, f_i$  在  $a$  处可微, 且此时  $f'(a) \in \mathcal{L}(X; Y)$  等同于  $(f'_1(a), \dots, f'_m(a)) \in \mathcal{L}(X; Y_1) \times \cdots \times \mathcal{L}(X; Y_m)$ .

**证明** 若  $f$  在  $a \in \Omega$  处可微, 则

$$f(a + h) = f(a) + f'(a)h + \|h\| \delta(h),$$

其中当  $h \rightarrow 0_X$  时,  $\delta(h) \rightarrow 0_Y$ .

写成分量形式即:

$$f_i(a + h) = f_i(a) + A_i h + \|h\| \delta_i(h), i = 1, \dots, m,$$

其中  $A_i \in \mathcal{L}(X; Y_i)$  是  $f \in \mathcal{L}(X; Y)$  的第  $i$  个分量,  $i = 1, \dots, m$ . 由此即得  $f_i$  在  $a$  处的可微性以及  $f'_i(a) = A_i$ .

反过来, 若任意  $1 \leq i \leq m$ ,  $f_i$  在  $a \in \Omega$  处可微, 即

$$f_i(a + h) = f_i(a) + f'_i(a)h + \|h\| \delta_i(h), \quad i = 1, \dots, m,$$

定义线性映射

$$A : X \rightarrow Y$$

$$x \mapsto (f'_1(a)x, \dots, f'_m(a)x),$$

则有

$$f(a + h) = f(a) + Ah + \|h\| \delta(h), \quad (2.6)$$

其中  $\delta(h) = (\delta_1(h), \dots, \delta_m(h))$ . 由于

$$\begin{aligned} \|Ah\|_Y &= \max_{1 \leq i \leq m} \|f'_i(a)h\|_{Y_i} \\ &= \max_{1 \leq i \leq m} \|f'_i(a)\|_{\mathcal{L}(X; Y_i)} \|h\|_X, \end{aligned} \quad (2.7)$$

所以  $A \in \mathcal{L}(X; Y)$ , 从而  $f$  在  $a$  处可微且  $f'(a) = A$ . ■

下面考虑  $X$  为乘积空间的情况.

**定义 2.3** 给定  $X_1, \dots, X_n$  为 Banach 空间,  $X = X_1 \times \dots \times X_n$ , 配备最大值范数,  $\Omega$  为 Banach 空间  $X$  中开集,  $a \in \Omega$ , 且任意  $1 \leq i \leq n$ , 存在  $X_i$  中开集  $\Omega_i$ , 使得  $\Omega_1 \times \dots \times \Omega_n \subset \Omega$ . 若对某个  $1 \leq i \leq n$ , 映射

$$\begin{aligned} f(a_1, \dots, a_{i-1}, \cdot, a_{i+1}, \dots, a_n) : \Omega_i &\rightarrow Y \\ x &\mapsto f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n) \end{aligned}$$

在  $a_i$  处可微, 则称其导数为  $f$  在  $a$  处的第  $i$  个偏导数, 记为  $\partial_i f(a)$ .

下面的定理给出了映射在一点的微分和其偏导数间的关系.

**定理 2.4** 给定  $X_1, \dots, X_n$  为 Banach 空间,  $X = X_1 \times \dots \times X_n$ . 若映射  $f : \Omega \subset X \rightarrow Y$  在  $a \in \Omega$  处可微, 则其各偏导数均存在, 且任意

$$f'(a)h = \sum_{i=1}^n \partial_i f(a)h_i, \quad h = (h_1, \dots, h_n) \in X.$$

**证明** 只证  $n = 2$  的情形,  $n \geq 3$  时证明没有本质区别.

定义线性映射

$$A_1 : X_1 \rightarrow Y$$

$$h_1 \mapsto f'(a)(h_1, 0),$$

和

$$\begin{aligned} A_2 : X_2 &\rightarrow Y \\ h_2 &\mapsto f'(a)(0, h_2). \end{aligned}$$

则

$$\begin{aligned} f(a_1 + h_1, a_2) &= f(a_1, a_2) + f'(a)(h_1, 0) + \|(h_1, 0)\|_\infty \delta(h_1, 0) \\ &= f(a_1, a_2) + A_1 h_1 + \|h_1\|_{X_1} \delta(h_1, 0). \end{aligned} \quad (2.8)$$

且  $\|A_1 h_1\|_Y = \|f'(a)(h_1, 0)\|_Y \leq \|f'(a)\|_{\mathcal{L}(X; Y)} \|h_1\|_{X_1}$ , 所以  $A_1 = \partial_1 f(a) \in \mathcal{L}(X_1; Y)$ . 对第二个分量的证明类似. ■

### 2.1.2 Gateaux 导数的定义

Fréchet 可微可以看作是欧氏空间之间映射局部线性逼近的自然推广, 下面考虑推广方向导数的概念.

**定义 2.4** 给定映射  $f : \Omega \rightarrow Y, a \in \Omega, h \in X$  以及一个开区间  $I_h \in \mathbb{R}$ , 其中  $I_h$  满足:

- (1)  $0 \in I_h$ ;
- (2)  $\forall \theta \in I_h, a + \theta h \in \Omega$ .

称  $f$  在  $a$  处有沿方向  $h$  的 Gateaux 导数, 是指: 极限

$$\partial_h f(a) := \lim_{\theta \rightarrow 0} \frac{f(a + \theta h) - f(a)}{\theta}$$

在  $Y$  中存在,  $\partial_h f(a)$  即为  $f$  在  $a$  处沿方向  $h$  的 Gateaux 导数.

**注** 与欧式空间之间的映射相同, 若  $f$  在一点处可微, 则沿任意方向的 Gateaux 导数均存在, 并且  $\partial_h f(a) = f'(a)h$ ; 反之则不成立. 此外, 定义 2.3 中的偏导数可以看作特殊方向上的 Gateaux 导数.

### 2.1.3 若干例子

下面介绍几个简单的例子.

**例 2.1 (仿射映射)** 给定  $A \in \mathcal{L}(X; Y), b \in Y$ , 定义映射

$$\begin{aligned} f : X &\rightarrow Y \\ x &\mapsto Ax + b. \end{aligned}$$

易见  $\forall x \in X, f'(x) = A$ .

**例 2.2 (双线性映射)** 设  $B : X \times X \rightarrow Y$  是连续双线性映射. 由于  $B$  是双线性的,

$$B(a_1 + h_1, a_2 + h_2) = B(a_1, a_2) + (B(h_1, a_2) + B(a_1, h_2)) + B(h_1, h_2),$$

且  $\|B(h_1, h_2)\| \leq \|B\| \|h_1\| \|h_2\|$ , 所以

$$B'(a_1, a_2)(h_1, h_2) = B(h_1, a_2) + B(a_1, h_2), (h_1, h_2) \in X_1 \times X_2,$$

$$\partial_1 B(a_1, a_2)(h_1) = B(h_1, a_2), h_1 \in X_1,$$

$$\partial_2 B(a_1, a_2)(h_2) = B(a_1, h_2), h_2 \in X_2.$$

**注**  $B'(a_1, a_2)$  表示一个连续线性算子, 而  $B'(a_1, a_2)(h_1, h_2)$  则表示该算子作用于  $(h_1, h_2)$ .

**例 2.3 (续例2.2)** 设  $B$  是对称的, 即  $B(x, y) = B(y, x), \forall x, y \in X$ . 定义映射  $\tilde{B}(x) := B(x, x)$ . 则  $\tilde{B}'(a)(h) = 2B(a, h)$ .

进一步地, 设  $M \in \mathcal{L}_k(X; Y)$  是一个对称的  $k-$  线性映射, 令  $\tilde{M}(x) = M(x, \dots, x)$ , 则  $(\tilde{M})'(a)(h) = kM(a, \dots, a, h)$ , 简记  $(\tilde{M})'(a) = Ma^{k-1}$ .

**例 2.4 (方阵的函数)** 记  $M_n(\mathbb{R})$  为  $n$  阶实方阵的全体,  $U_n(\mathbb{R})$  为  $n$  阶可逆实方阵的全体,  $M_n(\mathbb{R}) \xrightarrow{\text{tr}} \mathbb{R}$  和  $M_n(\mathbb{R}) \xrightarrow{\det} \mathbb{R}$  分别表示取迹和行列式运算.

$\text{tr}(\cdot)$  作为连续线性泛函,  $\forall W \in M_n(\mathbb{R}), \text{tr}(\cdot)$  在  $W$  处的 Frechet 导数由  $(\text{tr})'(W)(H) = \text{tr}(H) = \langle I_n, H \rangle$  决定, 其中  $\langle A, B \rangle := \text{tr}(A^*B)$  表示  $n$  阶实方阵上的内积.

另一方面, 若  $W \in U_n(\mathbb{R})$ , 则

$$\begin{aligned} \det(W + H) &= \det(W) \det(I_n + W^{-1}H) \\ &= \det(W)(1 + \text{tr}(W^{-1}H) + o(\|W^{-1}H\|)) \\ &= \det(W) + \text{tr}((\det(W)W^{-1})H) + \det(W)o(\|W^{-1}H\|). \end{aligned}$$

所以  $(\det)'(W)(H) = \text{tr}((\text{Adj}(W))H) = \langle \text{Adj}(W)^T, H \rangle$ , 其中  $\text{Adj}(W)$  表示  $W$  的伴随矩阵.

**例 2.5 (续例2.4)** 考虑取逆映射

$$f : U_n(\mathbb{R}) \rightarrow U_n(\mathbb{R})$$

$$A \mapsto A^{-1}.$$

由 (Ciarlet, 2017a) 的定理 3.6-3, 当  $H$  的矩阵范数充分小时

$$\begin{aligned} f(A + H) &= (A + H)^{-1} \\ &= (I_n + A^{-1}H)^{-1}A^{-1} \\ &= (I_n - A^{-1}H + o(H))A^{-1} \\ &= f(A) - A^{-1}HA^{-1} + o(H)A^{-1}. \end{aligned}$$

所以  $f'(A)(H) = -A^{-1}HA^{-1}$ .

## 2.2 链式法则

在古典微积分中, 链式法则是计算导数的重要工具, 在一般的 Banach 空间中对 Fréchet 同样有链式法则.

**定理 2.5 (链式法则)** 给定 Banach 空间  $X, Y, Z, U, V$  分别为  $X, Y$  中开集. 映射  $f : U \rightarrow Y$  在  $a \in U$  处可微且  $f(U) \subset V$ , 映射  $g : V \rightarrow Z$  在  $f(a) \in V$  处可微. 则映射  $g \circ f : U \rightarrow Z$  在  $a$  处可微, 且

$$(g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

此外, 若  $f \in C^1(U; Y), g \in C^1(V; Z)$ , 则  $g \circ f \in C^1(U; Z)$ .

**证明** 对任意  $a + h \in U$ , 定义

$$b := f(a), k(h) = f(a + h) - f(a).$$

由  $f, g$  的可微性可得,

$$\begin{aligned} f(a + h) &= f(a) + f'(a)h + \|h\| \delta(h), \lim_{h \rightarrow 0} \delta(h) = 0 \\ g(b + k) &= g(b) + g'(b)k + \|k\| \eta(k), \lim_{k \rightarrow 0} \eta(k) = 0. \end{aligned}$$

所以

$$\begin{aligned} &(g \circ f)(a + h) - (g \circ f)(a) \\ &= g(f(a + h)) - g(f(a)) \\ &= g'(f(a))(k(h)) + \|k(h)\| \eta(k(h)) \\ &= g'(f(a))(f'(a)h + \|h\| \delta(h)) + \|k(h)\| \eta(k(h)) \\ &= g'(f(a))(f'(a)h) + \|h\| g'(f(a))\delta(h) + \|k(h)\| \eta(k(h)) \end{aligned} \tag{2.9}$$

并且当  $h \rightarrow 0$  时,

$$\frac{\|h\| g'(f(a))\delta(h) + \|k(h)\| \eta(k(h))}{\|h\|} \rightarrow 0,$$

另外连续线性映射的复合仍然是连续线性映射, 所以  $g \circ f$  在  $a$  处可微且  $(g \circ f)'(a) = g'(f(a)) \circ f'(a)$ .

若  $f \in C^1(U; Y), g \in C^1(V; Z)$ , 为证  $g \circ f \in C^1(U; Z)$ , 只需证映射

$$x \mapsto (f'(x), g'(f(x))) = (f'(x), (g' \circ f)(x))$$

和

$$\mathcal{L}(X; Y) \times \mathcal{L}(Y; Z) \ni (A, B) \mapsto B \circ A \in \mathcal{L}(X; Z)$$

的连续性. 而这两个的映射的连续性则由  $f', g'$  以及映射复合保持连续性得到.

■

### 2.3 高阶导数和 Taylor 展开

在古典微积分中, 如果一个函数  $f : \mathbb{R} \rightarrow \mathbb{R}$  处处可导, 则自然地有一个导函数  $f' : \mathbb{R} \rightarrow \mathbb{R}$ , 则可以考察导函数的连续性和可微性. 而对于一般的向量空间, 若  $X, Y$  是 Banach 空间, 则  $\mathcal{L}(X; Y)$  也是 Banach 空间, 所以对于  $f \in C^1(\Omega; Y)$ , 同样可以考察  $f' : \Omega \rightarrow \mathcal{L}(X; Y)$  的 Frechet 可微性, 也就是二阶导数.

**定义 2.5** 设  $f : \Omega \rightarrow Y$  在  $\Omega$  内可微. 称  $f$  在  $a \in \Omega$  处二次可微, 是指: 映射

$$\begin{aligned} f' : \Omega &\rightarrow \mathcal{L}(X; Y) \\ x &\mapsto f'(x) \end{aligned} \tag{2.10}$$

在  $a$  处 Frehcet 可微. 记  $f''(a) := (f')'(a)$  为  $f$  处的二次导数.

若  $f$  在  $\Omega$  内处处二次可微, 且映射

$$\begin{aligned} f'' : \Omega &\rightarrow \mathcal{L}(X; \mathcal{L}(X; Y)) \\ x &\mapsto f''(x) \end{aligned} \tag{2.11}$$

是连续的, 则称  $f$  在  $\Omega$  内是  $C^2$  的,  $\Omega$  内到  $Y$  的  $C^2$  映射全体记为  $C^2(\Omega; Y)$ , 特别地,  $Y = \mathbb{R}$  时, 记  $C^2(\Omega; Y) = C^2(\Omega)$

**注** 由 (Ciarlet, 2017a) 的定理 2.11-5,  $\mathcal{L}(X; \mathcal{L}(X; Y))$  同构于  $\mathcal{L}_2(X; Y)$ , 所以对  $h, k \in X$ ,  $(f''(a)h)k = f''(a)(h, k)$ , 这里  $f''(a)$  作为  $\mathcal{L}(X; \mathcal{L}(X; Y))$  中的元素先作用于  $h$  再作用于  $k$ , 但下面的定理说明这两步操作的顺序可以交换: 即先作用于  $h$  再作用于  $k$  和先作用于  $k$  再作用于  $h$  的结果是相同的, 也即  $f''(a)$  作为一个双线性映射是对称的.

**定理 2.6** 设  $f : \Omega \rightarrow Y$  在  $a \in \Omega$  处二次可微, 则  $f''(a) \in \mathcal{L}_2(X; Y)$  是对称的.

定理的证明见 (Ciarlet, 2017b) 的定理 7.8-1, 该证明依赖于如下的中值定理.

**定理 2.7** 给定映射  $f : \Omega \rightarrow Y$ , 闭区间  $[a, b] \subset \Omega, \forall x \in [a, b], f$  在  $x$  处连续,  $\forall x \in (a, b), f$  在  $x$  处可微. 则

$$\|f(b) - f(a)\|_Y \leq (\sup_{x \in (a, b)} \|f'(x)\|_{\mathcal{L}(X; Y)}) \|b - a\|_X.$$

**证明** 不妨设  $M = \sup_{x \in (a, b)} \|f'(x)\|_{\mathcal{L}(X; Y)} < \infty$ .

定义映射  $\phi(t) := f(ta + (1-t)b), t \in [0, 1]$ , 则  $\phi : [0, 1] \rightarrow Y$  连续, 在  $(0, 1)$  可微, 且

$$\phi'(t) = f'(ta + (1-t)b)(b - a), t \in (0, 1) \implies \sup_{t \in (0, 1)} \|\phi'(t)\| \leq M \|b - a\|$$

以下讨论被称为”连续性方法”.

对任意  $\epsilon > 0$ , 定义映射

$$F : [0, 1] \rightarrow \mathbb{R}$$

$$t \mapsto \|\phi(t) - \phi(0)\| - (M \|b - a\| + \epsilon)t - \epsilon$$

以及  $I(\epsilon) = F^{-1}((\infty, 0])$ , 由于  $0 \in I(\epsilon)$ , 所以  $I(\epsilon)$  非空.

$F$  是由连续函数的复合和四则运算得到, 因此仍是连续的, 从而  $I(\epsilon)$  是闭集, 所以  $t_0 := \sup(I(\epsilon) \cap [0, 1]) \in I(\epsilon)$ . 往证  $t_0 = 1$ .

假设  $t_0 < 1$ , 则对于充分小的  $\delta > 0$  有  $t_0 + \delta < 1$ ,

$$\phi(t_0 + \delta) - \phi(t_0) = \phi(t_0) + \phi'(t_0)\delta + \delta\eta(\delta), \lim_{\delta \rightarrow 0} \eta(\delta) = 0.$$

取充分小  $\delta_0$  使得  $t_0 + \delta_0 < 1$  且  $\eta(\delta) < \epsilon$ , 则

$$\begin{aligned} \|\phi(t_0 + \delta_0) - \phi(0)\| &\leq \|\phi(t_0 + \delta_0) - \phi(t_0)\| + \|\phi(t_0) - \phi(0)\| \\ &\leq (M \|b - a\| + \epsilon)(t_0 + \delta_0) + \epsilon. \end{aligned} \quad (2.12)$$

矛盾, 所以  $t_0 = 1$ . 这也就证明了定理. ■

上述中值定理只对单点处可微性有要求, 如果  $f \in C^1(\Omega; Y)$ , 则还有更精确的表达.

**定理 2.8** 设  $f \in C^1(\Omega; Y)$ ,  $[a, b] \subset \Omega$ . 则

$$f(b) - f(a) = \int_0^1 f'((1-t)a + tb)(b - a)dt.$$

**注** 关于向量值函数的积分或称 Bochner 积分的相关定义和性质见 (Yosida, 2022), 上述定理的证明见 (Ciarlet, 2017a).

类似于二阶导数, 可以归纳定义高阶导数, 为统一记号, 记  $f^{(0)} = f$ ,  $f^{(1)} = f'$ ,  $f^{(2)} = f''$ .

**定义 2.6** 设  $m \in \mathbb{N}$ , 设映射  $f : \Omega \rightarrow Y$  在  $\Omega$  内  $(m - 1)$  次可微, 称  $f$  在  $a \in \Omega$  处  $m$  次可微, 是指: 映射

$$\begin{aligned} f^{(m-1)} : \Omega &\rightarrow \mathcal{L}(X; \mathcal{L}_{m-1}(X; Y)) \\ x &\mapsto f^{(m-1)}(x) \end{aligned}$$

在  $a$  处可微,  $a$  处的  $m$  阶导数记为  $f^{(m)}(a) = (f^{(m-1)})'(a)$ . 若  $m$  阶导数映射连续, 则称  $f$  在  $\Omega$  内是  $C^m$  的; 若  $f$  是任意次可微的, 则称  $f$  是光滑的.

**注** 类似  $C^1(\Omega; Y)$  和  $C^1$  微分同胚, 可以定义  $C^m(\Omega; Y)$  和  $C^m$  微分同胚,  $m \in \mathbb{N} \cup \{\infty\}$ ; 另外, 简记  $f^{(m)}(a)h^m = f^{(m)}(a)(\underbrace{h, \dots, h}_{m \text{ 个}})$

有了以上准备, 可以叙述并证明 Banach 空间中的 Taylor 公式.

**定理2.9** (Taylor公式) 给定映射  $f : \Omega \rightarrow Y, [a, a+h] \subset \Omega$ , 以及整数  $m \geq 1$ .

(1)  $f$  在  $\Omega$  中  $(m-1)$  次可微且在  $a$  处  $m$  次可微, 则

$$f(a+h) = f(a) + \sum_{i=1}^m \frac{1}{i!} f^{(i)}(a) h^i + \|h\|^m \delta(h).$$

(2) 若  $Y = \mathbb{R}$ ,  $f$  在  $\Omega$  中  $(m-1)$  次连续可微, 在  $(a, a+h)$  中  $m$  次可微, 则存在  $\theta \in (0, 1)$  使得

$$f(a+h) = f(a) + f'(a)h + \dots + \frac{1}{(m-1)!} f^{(m-1)}(a)h^{m-1} + \frac{1}{m!} f^{(m)}(a + \theta h)h^m.$$

(3) 设  $f$  在  $\Omega$  内  $m$  次连续可微, 则

$$f(a+h) = f(a) + f'(a)h + \dots + \frac{1}{(m-1)!} f^{(m-1)}(a)h^{m-1} + \frac{1}{(m-1)!} \int_0^1 (1-t)^{m-1} (f^m(a+th)h^m) dt.$$

(4) 设  $f$  在  $\Omega$  中  $(m-1)$  次连续可微, 在  $(a, a+h)$  中  $m$  次可微, 则

$$\left\| f(a+h) - f(a) + \sum_{i=1}^{m-1} \frac{1}{i!} f^{(i)}(a) h^i \right\| \leq \frac{1}{m!} \left( \sup_{x \in (a, a+h)} \|f^{(m)}(x)\| \right) \|h\|^m.$$

**注** 上述定理的(1),(2),(3) 分别对应单变量微积分中带 Peano 余项, Lagrange 余项和积分余项的 Taylor 展开公式,(4) 则是中值定理2.3对高阶导数的推广.

### 证明

先证明(1). 当  $m = 1$  时根据导数定义自动成立. 假设对  $m = 1, \dots, k-1$  都成立. 存在  $r > 0$  使得  $B(a, r) \subset \Omega$  且映射

$$g : B(a, r) \rightarrow Y$$

$$x \mapsto f(a+x) - f(a) - \left( \sum_{i=1}^k \frac{1}{i!} f^{(i)}(a) x^i \right)$$

在  $B(a, r)$  内可微, 且

$$g'(x) = f'(a+x) - \left( \sum_{i=1}^k \frac{1}{(i-1)!} f^{(i)}(a) x^{i-1} \right).$$

根据归纳假设,

$$f'(a+x) = f'(a) + \dots + \frac{1}{(k-1)!} f^{(k)}(a) x^{k-1} + \|x\|^{k-1} \delta(x)$$

且  $\lim_{x \rightarrow 0} \delta(x) = 0$ . 由中值定理2.3,

$$\begin{aligned} & \left\| f(a+h) - f(a) - \left( \sum_{i=1}^m \frac{1}{i!} f^{(i)}(a) h^i \right) \right\| \\ &= \|g(h) - g(0)\| \\ &\leq \left( \sup_{x \in (a, a+h)} \|g'(x)\| \right) \|h\| \\ &\leq \left( \sup_{x \in (a, a+h)} \|x\|^{k-1} \|\delta(x)\| \right) \|h\| \\ &\leq \|h\|^k \|\eta(h)\|. \end{aligned} \tag{2.13}$$

且  $\lim_{h \rightarrow 0} \eta(h) = 0_Y$ .

定义辅助函数  $\phi(t) := f(a + th), t \in (0, 1)$ , 则在 (2) 的条件下分别使用链式法则和单变量实函数的带 Lagrang 余项的 Taylor 展开式即可证得 (2).

而在 (3) 的条件下可知  $\phi$  在一个包含  $[0, 1]$  的开区间上是  $m$  次连续可微的. 定义函数  $\psi(t) := \phi(t) + \sum_{i=1}^{m-1} (1-t)^i \phi^{(i)}(t)$ , 再将定理 2.8 应用于

$$\psi(1) - \psi(0) = \int_0^1 \psi'(t) dt$$

即可证得 (3).

最后证明 (4). 当  $m = 1$  时由中值定理 2.3 成立, 假设对  $m = 1, \dots, k-1$  均成立.

令  $u(t) = \phi(t) - (f(a) + \sum_{i=1}^{k-1} \frac{1}{i!} f^{(i)}(a)(th)^{i-1})$ . 由  $f$  的可微性可得  $u$  在包含  $[0, 1]$  的一个开区间上可微.

对  $f'$  用归纳假设可得,

$$\begin{aligned} \|u'(t)\| &= \left\| (f'(a+th) - (f'(a) + \dots + \frac{1}{(k-2)!} f^{(k-1)}(th)^{k-2})h) \right\| \\ &\leq \frac{1}{(k-1)!} \left( \sup_{x \in (a, a+h)} \|f^{(k)}(x)\| \right) t^{k-1} \|h\|^k, \quad 0 \leq t \leq 1. \end{aligned} \quad (2.14)$$

令  $\chi(t) := \frac{1}{(k)!} \left( \sup_{x \in (a, a+h)} \|f^{(k)}(x)\| \right) t^k \|h\|^k$ . 则

$$\begin{aligned} \|u(1) - u(0)\| &\leq \sum_{i=0}^{l-1} \left\| u\left(\frac{j+1}{l}\right) - u\left(\frac{j}{l}\right) \right\| \\ &\leq \frac{1}{l} \sum_{i=0}^{l-1} \sup_{t \in (\frac{j}{l}, \frac{j+1}{l})} \|u'(t)\| \\ &\leq \frac{1}{l} \sum_{j=0}^{l-1} \sup_{t \in (\frac{j}{l}, \frac{j+1}{l})} \chi'(t). \end{aligned} \quad (2.15)$$

右边是一个 Riemann 和的形式, 令  $l \rightarrow \infty$ , 即得

$$\|u(1) - u(0)\| \leq \chi(1) - \chi(0).$$

也即证得了结论. ■

## 第3章 $\mathcal{P}_2$ 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  $\mathcal{P}$  表示  $\mathbb{R}^d$  上全体概率测度,  $\mathcal{P}_p$  表示全体  $p$  阶矩有限概率测度, 即

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

$\mathcal{P}_p$  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

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

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

对一个函数  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$ , 我们希望考察  $\mu \in \mathcal{P}_2$  处的局部性态, 但由于  $\mathcal{P}_2$  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  $\mu$  附近函数的性态, 而是考虑将  $f$  先“提升”为定义在  $L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$  上的函数  $\tilde{f}$ , 考察分布为  $\mu$  的随机变量  $X \in L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$  附近  $\tilde{f}$  的性态, 而  $L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$  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  $\tilde{f} = f \circ \mathcal{L}$ , 其中  $\mathcal{L}(X)$  表示  $\mathbb{R}^d$  上由随机变量  $X$  诱导的概率分布, 也即  $\mathcal{L}(X)(A) = \mathbb{P}(X^{-1}(A)), \forall A \in \mathcal{B}(\mathbb{R}^d)$ .

$$\begin{array}{ccccc}
 X & \xrightarrow{\hspace{3cm}} & \mu = \mathcal{L}(X) & & \\
 & \searrow \tilde{f} = f \circ \mathcal{L} & \downarrow f & \uparrow & \\
 L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) & \xrightarrow{\mathcal{L}} & B & & \\
 & \swarrow & \downarrow & & \\
 & & \mathbb{R} & & f(\mu)
 \end{array}$$

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  $\mu \in \mathcal{P}$ , 总存在一个随机变量  $X$  使得  $\mathcal{L}(X) = \mu$ , 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  $(\Omega, \mathcal{F}, \mathbb{P})$  总是固定的, 简记  $L^2(\Omega; \mathbb{R}^d) = L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ . 对任意  $A \in \mathcal{F}$ , 记  $\mathcal{F}|_A := \{E \cap A : E \in \mathcal{F}\}$ . 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.

### 3.1 L-可微性

#### 3.1.1 定义

**定义 3.1** 给定一个函数  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$ , 称  $f$  在  $\mu_0 \in \mathcal{P}_2$  是 L-可微的, 是指: 存在一个分布为  $\mu_0$  的随机变量  $X_0$ , 使得  $f$  的提升  $\tilde{f} = f \circ \mathcal{L} : L^2(\Omega; \mathbb{R}^d) \rightarrow \mathbb{R}$  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 3.1** 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<sup>①</sup>  $S, T : \Omega \rightarrow \Omega$  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  $\mathcal{F}^*, \mathbb{P}^*$  分别为  $\mathcal{F}, \mathbb{P}$  的完备化扩张.

对给定的  $\epsilon$ , 设  $\{A_n\}_{n=1}^\infty \subset \mathcal{B}(\mathbb{R}^d)$  是  $\mathbb{R}^d$  的一个剖分, 且任意  $A_n$  直径不超过  $\epsilon$ (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  $(\Omega, \mathcal{F}, \mathbb{P})$  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  $\Omega_n^X = B_n - M_n, \mathcal{F}_n^X = \{A \cap \Omega_n^X : A \in \mathcal{F}^*\}, \mathbb{P}_n^X(E) = \frac{\mathbb{P}^*(E)}{\mathbb{P}(B_n)}, E \in \mathcal{F}_n^X$  以及  $\Omega_n^Y = C_n - N_n, \mathcal{F}_n^Y = \{A \cap \Omega_n^Y : A \in \mathcal{F}^*\}, \mathbb{P}_n^Y(E) = \frac{\mathbb{P}^*(E)}{\mathbb{P}(C_n)}, E \in \mathcal{F}_n^Y$ .

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

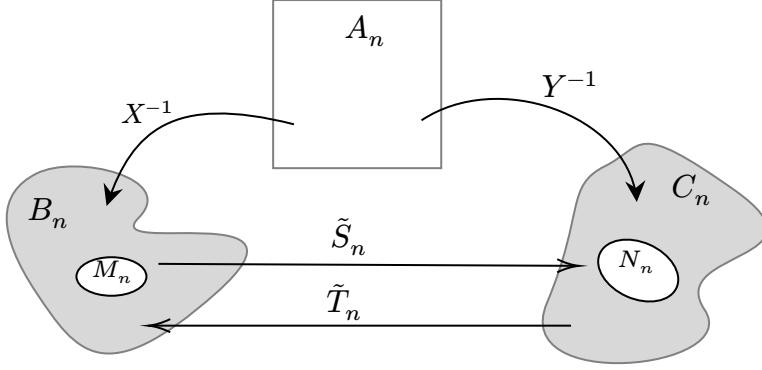
$$\tilde{S}_n : (\Omega_n^X, \mathcal{F}_n^X, \mathbb{P}_n^X) \rightarrow (\Omega_n^Y, \mathcal{F}_n^Y, \mathbb{P}_n^Y).$$

和

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

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

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



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  $(B_n, \mathcal{F}|_{B_n})$  到  $(C_n, \mathcal{F}|_{C_n})$  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  $(\Omega, \mathcal{F}, \mathbb{P})$  到自身的保测映射. 且以概率 1 互为逆.

最后, 对几乎处处  $\omega$ , 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  $F : L^2(\Omega; \mathbb{R}^d) \rightarrow \mathbb{R}$  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 3.2** 设函数  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  在  $\mu_0 \in \mathcal{P}_2$  处 L-可微. 则对于任意分布为  $\mu_0$  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  $\mu$  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  $\mu$  的随机变量  $X \in L^2(\Omega; \mathbb{R}^d)$ , 任意  $\epsilon > 0$ , 存在  $\Omega$  到自身的保测映射  $S_\epsilon$  和  $T_\epsilon$  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  $Y \in$

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

$$\begin{aligned}
 \tilde{f}(X + Y) &= \tilde{f}((X + Y) \circ S_\epsilon) \\
 &= \tilde{f}(X_0 + (X \circ S_\epsilon - X_0) + Y \circ S_\epsilon) \\
 &= \tilde{f}(X_0) + (\tilde{f})'(X_0)((X \circ S_\epsilon - X_0) + Y \circ S_\epsilon) \\
 &\quad + \|(X \circ S_\epsilon - X_0) + Y \circ S_\epsilon\|_{L^2} \delta((X \circ S_\epsilon - X_0) + Y \circ S_\epsilon) \\
 &= \tilde{f}(X_0) + \mathbb{E}[D\tilde{f}(X_0) \cdot ((X \circ S_\epsilon - X_0) + Y \circ S_\epsilon)] \\
 &\quad + \|(X \circ S_\epsilon - X_0) + Y \circ S_\epsilon\|_{L^2} \delta((X \circ S_\epsilon - X_0) + Y \circ S_\epsilon) \\
 &= \tilde{f}(X) + \mathbb{E}[(D\tilde{f}(X_0) \circ T_\epsilon) \cdot Y] + \mathbb{E}[D\tilde{f}(X_0) \cdot (X \circ S_\epsilon - X_0)] \\
 &\quad + \|(X \circ S_\epsilon - X_0) + Y \circ S_\epsilon\|_{L^2} \delta((X \circ S_\epsilon - X_0) + Y \circ S_\epsilon). \quad (3.1)
 \end{aligned}$$

其中  $\delta(\cdot)$  与  $\epsilon$  无关且  $\lim_{\|Y\|_{L^2} \rightarrow 0} \delta(Y) = 0$ , 第 1 个等号和第 5 个等号分别使用了  $S_\epsilon, T_\epsilon$  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

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

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  $\epsilon \rightarrow 0$  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  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  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  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  是连续 L-可微的, 是指:  $f$  在任意  $\mu \in \mathcal{P}_2$  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 3.3** 设  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  是连续 L-可微的. 则对任意  $\mu \in \mathcal{P}_2$ , 存在可测函数  $\xi : \mathbb{R}_d \rightarrow \mathbb{R}_d$ , 使得任意以  $\mu$  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  $\sigma$  代数  $\sigma(X)$  可测. 不妨设  $f$  有界. 证明分为以下几步.

(1) 先考虑  $\mu$  关于 Lebesgue 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  $\Psi$  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  $\Psi$  有下界, 从而一定有下确界, 取  $\{Z_n\} \subset L^4(\Omega; \mathbb{R}^d)$  为  $\Psi$  的极小化序列, 即  $\lim_{n \rightarrow \infty} \Psi(Z_n) = \inf_{Z \in L^4(\Omega; \mathbb{R}^d)} \Psi(Z)$  以及  $\nu_n = \mathcal{L}(\nu_n)$ . 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  $\mathcal{L}(\nabla \psi(X_n)) = \nu_n$  且是  $\mu$ -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

$$\Psi(Y_n) = \tilde{f}(Y_n) + \frac{1}{2\epsilon} W_2(\mu, \nu_n)^2 + \mathbb{E}[|Y|^4] \leq \Psi(Z_n), \quad (3.2)$$

从而  $\{Y_n\}$  也是  $\Psi$  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  $\nu_n$  在  $W_2$  距离和弱收敛意义下收敛到  $\nu$ , 所以

$$\begin{aligned} f(\nu) &= \lim_{n \rightarrow \infty} f(\nu_n), \\ W_2(\mu, \nu) &= \lim_{n \rightarrow \infty} W_2(\mu, \nu_n). \end{aligned} \quad (3.3)$$

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

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

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

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

这说明  $D\tilde{f}(Y) \in \sigma(X)$ .

注意到这里的  $Y$  是依赖于  $\epsilon$  的, 不妨记为  $Y_\epsilon$ , 由于  $Y_\epsilon$  极小化  $\Psi_\epsilon$ , 所以对任意  $\epsilon$ ,

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

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

(2) 假设  $\mu \in \mathcal{P}_2$  关于 Lebesgue 测度绝对连续但不一定有 (1) 中的矩条件, 取分布为  $\mu$  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  $n \rightarrow \infty$  即可.

(3) 若  $\mu \in \mathcal{P}_2$  不是关于 Lebesgue 测度绝对连续的, 取分布为  $\mu$  的随机变量  $X \in L^2(\Omega; \mathbb{R}^d)$  和  $\Omega$  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 Lebesgue 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  $\xi$  不依赖于  $X$  的选取. 假设对两个分布为  $\mu$  的二阶矩有限随机变量  $X_1, X_2$  分别有对应的  $\xi_1, \xi_2$ , 任取  $A \in \mathcal{B}(\mathbb{R}^d)$ , 易见  $\mu(\xi_1 I_A) = \mu(\xi_2 I_A)$ , 所以  $\xi_1 = \xi_2, \mu\text{-a.e.}$ . ■

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  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  在  $\mu_0$  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  $\mathcal{L}(X) = \mu, \mathcal{L}(X_0) = \mu_0$ .

另一方面, 为了突出 L-可微的内蕴性, 我们将定理3.3中的  $\xi$  记为  $\partial_\mu f(\mu_0)$ , 称为  $f$  在  $\mu_0$  处的 Lions 导数或 L-导数. 此时有

$$f(\mu) = f(\mu_0) + \mathbb{E}[\partial_\mu f(\mu_0)(X_0) \cdot (X - X_0)] + o(\|X - X_0\|_{L^2}), \quad (3.4)$$

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

### 3.1.2 若干例子

**例 3.1** (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  $h : \mathbb{R}^d \rightarrow \mathbb{R}$ , 且  $|\nabla h(x)| \leq C(1 + |x|)$ . 令  $f(\mu) = \int_{\mathbb{R}^d} h(x)\mu(dx), f$  的提升为  $\tilde{f}(X) = \mathbb{E}[h(X)]$ , 其中  $\mathcal{L}(X) = \mu$ . 计算  $\tilde{f}$  的 Frechet 导数, 首先由中值定理,

$$h(x + y) = h(x) + \int_0^1 (\nabla h(x + ty) \cdot y) dt, \quad (3.5)$$

$$\begin{aligned} \mathbb{E}[h(X + Y)] &= \mathbb{E}[h(X) + \int_0^1 (\nabla h(X + tY) \cdot Y) dt] \\ &= \mathbb{E}[h(X)] + \mathbb{E}[\nabla h(X) \cdot Y] + \mathbb{E}\left[\int_0^1 (\nabla h(X + tY) \cdot Y - \nabla h(X) \cdot Y) dt\right], \end{aligned} \quad (3.6)$$

其中,

$$\begin{aligned}
 & |\mathbb{E}[\int_0^1 ((\nabla h(X + tY) - \nabla h(X)) \cdot Y) dt]| \\
 & \leq \int_0^1 \mathbb{E}[(((\nabla h(X + tY) - \nabla h(X)) \cdot Y)) I_{\{|Y| \leq \|Y\|_{L^2}^{1/2}\}}] dt \\
 & \quad + \int_0^1 \mathbb{E}[(((\nabla h(X + tY) - \nabla h(X)) \cdot Y)) I_{\{|Y| > \|Y\|_{L^2}^{1/2}\}}] dt \\
 & \leq \mathbb{E}[\sup_{y \leq \|Y\|_{L^2}^{1/2}} (\nabla h(X + y) - \nabla h(X))] \|Y\|_{L^2} \\
 & \quad + C' \mathbb{E}[(1 + |X| + |Y|)|Y|] \\
 & \leq \mathbb{E}[\sup_{y \leq \|Y\|_{L^2}^{1/2}} (\nabla h(X + y) - \nabla h(X))] \|Y\|_{L^2} \\
 & \quad + C' \|Y\|_{L^2} (\|Y\|_{L^2} + \|Y\|_{L^2}^{1/2} + \sup_{\mathbb{P}(A) \leq \|Y\|_{L^2}} \mathbb{E}[|X|^2 I_A]^{1/2})
 \end{aligned} \tag{3.7}$$

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

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 3.2** (续例 3.1, 卷积函数) 接下来考虑一个稍微复杂的函数,令  $h * \mu(x) = \int_{\mathbb{R}^d} h(x - y) \mu(dy), g(\mu) = f(h * \mu)$ . 易见  $\tilde{g}(X) = \int_{\mathbb{R}^d} \mathbb{E}[h(x - X)] d\mathcal{L}(X)(x)$ ,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,

$$\begin{aligned}
 & \tilde{g}(X + \epsilon Y) - \tilde{g}(X) \\
 & = \int_{\mathbb{R}^d} \mathbb{E}[h(x - (X + \epsilon Y))] d\mathcal{L}(X + \epsilon Y)(x) - \int_{\mathbb{R}^d} \mathbb{E}[h(x - X)] d\mathcal{L}(X)(x) \\
 & = \left( \int_{\mathbb{R}^d} \mathbb{E}[h(x - X - \epsilon Y)] d\mathcal{L}(X + \epsilon Y)(x) - \int_{\mathbb{R}^d} \mathbb{E}[h(x - X)] d\mathcal{L}(X + \epsilon Y)(x) \right) \\
 & \quad + \left( \int_{\mathbb{R}^d} \mathbb{E}[h(x - X)] d\mathcal{L}(X + \epsilon Y)(x) - \int_{\mathbb{R}^d} \mathbb{E}[h(x - X)] d\mathcal{L}(X)(x) \right) \\
 & \triangleq I_1 + I_2,
 \end{aligned} \tag{3.8}$$

对于第二项,结合 Fubini 定理交换积分次序可得

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

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

$$\begin{aligned} I_1 &= \mathbb{E}[h((\hat{X} + \epsilon\hat{Y}) - (X + \epsilon Y))] - \mathbb{E}[h(\hat{X} + \epsilon\hat{Y} - X)] \\ &= \mathbb{E}[h((\hat{X} + \epsilon\hat{Y}) - (X + \epsilon Y))] - \mathbb{E}[h(\hat{X} - X)] \\ &\quad + \mathbb{E}[h(\hat{X} - X)] - \mathbb{E}[h(\hat{X} + \epsilon\hat{Y} - X)], \end{aligned} \tag{3.9}$$

所以

$$\lim_{\epsilon \rightarrow 0} \frac{I_2}{\epsilon} = \mathbb{E}[\nabla h(\hat{X} - X) \cdot (\hat{Y} - Y)] - \mathbb{E}[\nabla h(\hat{X} - X) \cdot (\hat{Y})] = -\mathbb{E}[\nabla h(\hat{X} - X) \cdot Y].$$

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

**例 3.3** 考虑连续函数  $v : \mathbb{R}^d \times \mathcal{P}_2 \rightarrow \mathcal{P}_2$  满足:

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

(2) 任意  $\mu \in \mathcal{P}_2, x \in \mathbb{R}^d$ , 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  $K \subset \mathcal{P}_2$ , 存在常数  $C_2$ , 使得任意  $\mu \in K, \partial_\mu v(x, \mu)(y) \leq C_2(1 + |y|)$ .

令  $u(\mu) = \int_{\mathbb{R}^d} v(x, \mu) d\mu(x)$ , 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  $\tilde{u}(X) = \mathbb{E}[v(X, \mathcal{L}(X))]$ ,

$$\begin{aligned} \tilde{u}(X + Y) - \tilde{u}(X) &= \mathbb{E}[v(X + Y, \mathcal{L}(X + Y))] - \mathbb{E}[v(X, \mathcal{L}(X))] \\ &= \mathbb{E}[v(X + Y, \mathcal{L}(X + Y))] - \mathbb{E}[v(X, \mathcal{L}(X + Y))] \\ &\quad + \mathbb{E}[v(X, \mathcal{L}(X + Y))] - \mathbb{E}[v(X, \mathcal{L}(X))] \\ &\stackrel{\triangle}{=} I_1 + I_2, \end{aligned} \tag{3.10}$$

由中值定理可得,

$$I_1 = \mathbb{E}\left[\int_0^1 \nabla v(X + tY, \mathcal{L}(X + Y)) \cdot Y dt\right],$$

而

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

所以可以合理推测  $\partial_\mu f(\mu)(\cdot) = \nabla v(\cdot, \mu) + \int_{\mathbb{R}^d} \partial_\mu f(y, \mu)(\cdot) d\mu(y)$ , 事实上根据对  $v$  的假设, 采用和例3.1中类似的讨论方法可知,  $f$  在  $\mu$  处的 L-导数确实为上式.

**例 3.4 (续3.3)** 设  $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  为一可测映射,  $\nabla_1 K, \nabla_2 K$  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  $v(x, \mu) =$

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

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

### 3.2 链式法则和 Itô 公式

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

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

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

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s, \quad t \geq 0. \quad (3.11)$$

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

$$f(X_t) = f(X_0) + \int_0^t \nabla f(X_s) \cdot b_s ds + \int_0^t \nabla f(X_s) \sigma_s dB_s + \frac{1}{2} \int_0^t \text{tr}(a_s \nabla^2 f(X_s)) ds \quad (3.12)$$

我们希望对式 (3.10) 给出的过程的时间边缘分布  $\mu := (\mu_t)_{t \geq 0} := (\mathcal{L}(X_t))_{t \geq 0}$  和一类性质足够好的  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  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  $f : \mathcal{P}_2 \rightarrow \mathbb{R}^d$  可以定义它的经验投影

$$\begin{aligned} f^N : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{N \text{ 个}} &= (\mathbb{R}^d)^N \longrightarrow \mathbb{R} \\ (x^1, \dots, x^N) &\mapsto f\left(\frac{1}{N} \sum_{l=1}^N \delta_{x^l}\right), \end{aligned} \quad (3.13)$$

其中  $N \in \mathbb{N}, \delta_x(A) = I_A(x), x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d)$  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 3.4** 设  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  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

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

注 此处的  $\partial_{x^i}$  表示沿  $x^i$  的方向导数算子.

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

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

$$\begin{aligned}
 \mathbb{P}(X_\theta^\mathbf{x} \in A) &= \sum_{l=1}^N \mathbb{P}((X_\theta^\mathbf{x} \in A, \theta = l)) \\
 &= \sum_{l=1}^N \mathbb{P}(X_\theta^\mathbf{x} \in A | \theta = l) \mathbb{P}(\theta = l) \\
 &= \frac{1}{N} \sum_{l=1}^N \mathbb{P}(X_l^\mathbf{x} \in A) \\
 &= \frac{1}{N} \sum_{l=1}^N \delta_{x^i}(A),
 \end{aligned} \tag{3.14}$$

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

$$\begin{aligned}
 f^N(\mathbf{x} + \mathbf{h}) &= f\left(\frac{1}{N} \sum_{l=1}^N \delta_{x^l + h^l}\right) \\
 &= \tilde{f}(X_\theta^\mathbf{x} + X_\theta^\mathbf{h}) \\
 &= \tilde{f}(X_\theta^\mathbf{x}) + \mathbb{E}[D\tilde{f}(X_\theta^\mathbf{x}) \cdot X_\theta^\mathbf{h}] + o(\|X_\theta^\mathbf{h}\|_{L^2}) \\
 &= f^N(\mathbf{x}) + \mathbb{E}[\partial_\mu f\left(\frac{1}{N} \sum_{l=1}^N \delta_{x^i}\right)(X_\theta^\mathbf{x}) \cdot X_\theta^\mathbf{h}] + o(\|X_\theta^\mathbf{h}\|_{L^2}),
 \end{aligned} \tag{3.15}$$

注意到  $\partial_\mu f\left(\frac{1}{N} \sum_{l=1}^N \delta_{x^i}\right)(X_\theta^\mathbf{x}) \cdot X_\theta^\mathbf{h}$  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. ■

### 3.2.1 完全 $C^2$ 正则性

由定理3.3, 当  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  是连续 L-可微时, 对任一  $\mu \in \mathcal{P}_2$  存在一个可测映射  $\partial_\mu f(\mu)(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , 使得当  $\mathcal{L}(X) = \mu$ ,  $D\tilde{f}(X) = \partial_\mu f(\mu)(X)$ , 且该可测映射在  $\mu$ -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  $\mathcal{P}_2 \times \mathbb{R}^d$  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)  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  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  $\mathcal{P}_2 \times \mathbb{R}^d$  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  $\mu \in \mathcal{P}_2$ , 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  $\mathcal{P}_2 \times \mathbb{R}^d$  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  $\mathcal{P}_2 \times \mathbb{R}^d \times \mathbb{R}^d$  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  $\mu \in \mathcal{P}_2(\mathbb{R}^d), x, y \in \mathbb{R}^d$ , 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  $A, \langle \partial_x \partial_\mu f(\mu)(x), A \rangle = \partial_x \partial_\mu f(\mu)(x) \cdot A = \text{tr}(\partial_x \partial_\mu f(\mu)(x)A^T)$  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  $f : \mathcal{P}_2 \rightarrow \mathbb{R}^d$  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 3.4 中经验投影的可微性的加强, 有如下定理.

**定理 3.5** 设  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  是完全  $C^2$  正则的. 则对任意  $N \in \mathbb{N}$ , 经验投影  $f^N : (\mathbb{R}^d)^N \rightarrow \mathbb{R}$  是经典意义下  $C^2$  的, 且对任意  $x = (x^1, \dots, x^N) \in (\mathbb{R}^d)^N$ , 和  $1 \leq i, j \leq N$ ,

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

在实变函数的情形下, 一个函数与一个性质充分好的函数做卷积后会继承部分好的性质, 对于测度变量的函数也可以有类似操作, 但这里不是通过卷积来平滑, 而是通过拉回映射来实现, 具体陈述如下.

**引理 3.6** 设  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  是  $C^2$  正则的,  $\rho : \mathbb{R}^d \rightarrow \mathbb{R}^d$  为光滑紧支映射. 则函数

$$\begin{aligned} f^* \rho : \mathcal{P}_2 &\rightarrow \mathbb{R} \\ \mu &\mapsto f(\mu \circ \rho) \end{aligned} \tag{3.16}$$

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.

**证明梗概** 首先  $f^*\rho$  的提升为  $\widetilde{f^*\rho}(X) = \tilde{f}(\rho \circ X)$ ,  $X \in L^2(\Omega; \mathbb{R}^d)$ , 这是因为若  $\mathcal{L}(X)$ , 则  $\mathcal{L}(\rho \circ X) = \mu \circ \rho^{-1}$ . 考虑  $\mathcal{L}(X) = \mu \in \mathcal{P}_2$ ,

$$\begin{aligned}
& \widetilde{f^*\rho}(X+Y) - \widetilde{f^*\rho}(X) \\
&= \tilde{f}(\rho \circ (X+Y)) - \tilde{f}(\rho \circ X) \\
&= \mathbb{E}[D\tilde{f}(\rho \circ X) \cdot (\rho \circ (X+Y) - \rho \circ X)] + o(\|\rho \circ (X+Y) - \rho \circ X\|_{L^2}) \quad (3.17) \\
&= \mathbb{E}[D\tilde{f}(\rho \circ X) \cdot ((D\rho \circ X)Y + o(|Y|))] + o(\|\rho \circ (X+Y) - \rho \circ X\|_{L^2}) \\
&= \mathbb{E}[D\tilde{f}(\rho \circ X) \cdot ((D\rho \circ X)Y)] + o(\|Y\|_{L^2}) \\
&= \mathbb{E}[(D\rho \circ X)D\tilde{f}(\rho \circ X) \cdot Y] + o(\|Y\|_{L^2}),
\end{aligned}$$

其中第四个等号是因为紧支光滑映射的任意阶微分仍是紧支光滑的, 特别还是有界的, 所以可以把小量统一吸收到  $o(\|Y\|_{L^2})$ . 从而

$$\partial_\mu(f^*\rho)(\mu)(x) = D\rho(x)\partial_\mu f(\mu \circ \rho^{-1})(\rho(x)) = \left( \sum_{l=1}^d (\partial_\mu f(\mu \circ \rho^{-1})(\rho(x)))^{(l)} \frac{\partial \rho_i}{\partial x_l}(x) \right)_{1 \leq i \leq d}$$

并将其视为一个  $d \times 1$  的列向量.

二阶可微性的验证类似, 需要注意的是对测度变量的二阶导数计算中使用公式3.4会更方便. 最后给出  $f^*\rho$  的导数公式,

$$\begin{aligned}
\partial_\mu^2(f^*\rho)(\mu)(x, y) &= D\rho(y)D\rho(x)\partial_\mu^2 f(\mu \circ \rho^{-1})(\rho(x), \rho(y)), \\
\partial_x \partial_\mu(f^*\rho)(\mu)(x) &= D^2\rho(x)\partial_\mu f(\mu \circ \rho^{-1})(\rho(x)) + D\rho(x)\partial_x \partial_\mu f(\mu \circ \rho^{-1})(\rho(x)). \quad (3.18)
\end{aligned}$$

■

### 3.2.2 Ito 公式

对于3.11定义的 Ito 过程和性质足够好的函数  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$ , 下面的定理给出了其时间边缘分布满足的 Ito 公式.

**定理 3.7** 设  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  是  $C^2$  正则的, 且对任意  $\mathcal{P}$  的紧子集  $K$ , 存在常数  $C = C(K)$ , 使得任意  $\mu \in K$ ,

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

则对任意  $t \geq 0$ ,

$$f(\mu_t) = f(\mu_0) + \int_0^t \mathbb{E}[\partial_\mu f(\mu_s)(X_s) \cdot b_s] ds + \frac{1}{2} \int_0^t \mathbb{E}[\partial_x \partial_\mu f(\mu_s)(X_s) \cdot a_s] ds. \quad (3.19)$$

**证明** 证明主要分为两步.

**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

$$X_t^l = X_0^l + \int_0^t b_s^l ds + \int_0^t \sigma_s^l dB_s^l, \quad t \geq 0, l \geq 1,$$

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

考虑  $f$  到  $(\mathbb{R}^d)^N$  的经验投影, 由定理3.5可知  $f^N$  是  $C^2$  的, 所以可以应用 Ito 公式3.12, 简记  $\mathbf{X}_t = (X_t^1, \dots, X_t^N)$  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

$$\begin{aligned} & f^N(\mathbf{X}_t) - f^N(\mathbf{X}_0) \\ &= \sum_{l=1}^N \int_0^t \partial_{x^i} f^N(\mathbf{X}_s) \cdot b_s^l ds + \sum_{l=1}^N \int_0^t \partial_{x^i} f^N(\mathbf{X}_s) \cdot (\sigma_s^l dB_s^l) \\ & \quad + \sum_{l=1}^N \partial_{x^l x^l}^2 f^N(\mathbf{X}_s) \cdot a_s^l ds \\ &= \frac{1}{N} \sum_{l=1}^N \int_0^t \partial_\mu f(\bar{\mu}_s^N)(X_s^l) \cdot b_s^l ds + \sum_{l=1}^N \int_0^t \partial_\mu f(\bar{\mu}_s^N)(X_s^l) \cdot (\sigma_s^l dB_s^l) \\ & \quad + \frac{1}{2N} \sum_{l=1}^N \int_0^t \partial_x \partial_\mu f(\bar{\mu}_s^N)(X_s^l) \cdot a_s^l ds \\ & \quad + \frac{1}{2N^2} \int_0^t \partial_\mu^2 f(\bar{\mu}_s^N)(X_s^l, X_s^l) \cdot a_s^l ds, \end{aligned} \tag{3.20}$$

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,

$$\begin{aligned} & \mathbb{E}[f(\bar{\mu}_t^N)] - \mathbb{E}[f(\bar{\mu}_0^N)] \\ &= \int_0^t \mathbb{E}[\partial_\mu f(\bar{\mu}_s^N)(X_s) \cdot b_s] ds \\ & \quad + \frac{1}{2} \int_0^t \mathbb{E}[\partial_x \partial_\mu f(\bar{\mu}_s^N)(X_s) \cdot a_s] ds \\ & \quad + \frac{1}{2N} \int_0^t \mathbb{E}[\partial_\mu^2 f(\bar{\mu}_s^N)(X_s, X_s) \cdot a_s] ds \\ &\stackrel{\triangle}{=} I_1 + I_2 + I_3, \end{aligned} \tag{3.21}$$

根据有界的假设, 当  $N \rightarrow \infty$  时,  $I_3 \rightarrow 0$ ; 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, 具体地说是  $f$  和一阶, 二阶导数的连续性, 结合有界收敛定理, 可得当  $N \rightarrow \infty$ ,  $\mathbb{E}[f(\bar{\mu}_s^N)] \rightarrow f(\mu_s)$ .  $I_1, I_2$  的收敛性类似.

**第二步:**

去掉第一步中对  $f$  及其导数的有界性和一致连续性假设. 取一列紧支光滑映射  $\{\rho_n : \mathbb{R}^d \rightarrow \mathbb{R}^d\}$  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  $n \geq 1, x \in \mathbb{R}^d$ ,

$$|\rho_n(x)| \leq C|x|,$$

$$|D\rho_n(x)| \leq C,$$

$$|D^2\rho_n(x)| \leq C.$$

以及当  $\rho_n(x)I_{B(0,n)}(x) = x$ . 从而对任意  $\mu \in \mathcal{P}_2$  以及以  $\mu$  为分布的随机变量  $X$ , 都有

$$\begin{aligned} W_2(\mu \circ \rho_n^{-1}, \mu)^2 &\leq \mathbb{E}[|\rho_n(X) - X|^2] \\ &= \mathbb{E}[|\rho_n(X) - X|^2 I_{\{|X| \geq n\}}] \\ &\leq (C+1)^2 \mathbb{E}[|X|^2 I_{\{|X| \geq n\}}], \end{aligned} \quad (3.22)$$

由于  $X$  二阶矩有限, 所以当  $n \rightarrow \infty$ ,  $W_2(\mu \circ \rho_n^{-1}, \mu) \rightarrow 0$ . 结合连续性可得以下收敛在  $\mathbb{P}$ -a.s. 意义下成立:

$$\begin{aligned} f^*\rho_n(\mu) &\rightarrow mu, \\ \partial_\mu(f^*\rho_n)(\mu)(X) &\rightarrow \partial_\mu f(\mu)(X), \\ \partial_x \partial_\mu(f^*\rho_n)(\mu)(X) &\rightarrow \partial_x \partial_\mu f(\mu)(X). \end{aligned} \quad (3.23)$$

由于

$$\sup_{n \geq 1} \mathbb{E}[|\partial_\mu(f^*\rho_n)(\mu)(X)|^2 + |\partial_x \partial_\mu(f^*\rho_n)(\mu)(X)|^2] < \infty,$$

结合 (Bogachev, 2007) 的定理 4.5.9, 可得  $\{\partial_\mu(f^*\rho_n)(\mu)(X)\}, \{\partial_x \partial_\mu(f^*\rho_n)(\mu)(X)\}$  都是一致可积的, 所以可以应用 Vitali 收敛定理得到对任意  $t \geq 0$  和任意  $s \in [0, t]$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[\partial_\mu(f^*\rho_n)(\mu)(X) \cdot b_s] &= \mathbb{E}[\partial_\mu f(\mu)(X) \cdot b_s], \\ \lim_{n \rightarrow \infty} \mathbb{E}[\partial_x \partial_\mu(f^*\rho_n)(\mu)(X) \cdot a_s] &= \mathbb{E}[\partial_x \partial_\mu f(\mu)(X) \cdot a_s]. \end{aligned} \quad (3.24)$$

对每个  $n$  应用 Ito 公式, 可得

$$\begin{aligned} (f^*\rho_n)(\mu_t) - (f^*\rho_n)(\mu_0) &= \int_0^t \mathbb{E}[\partial_\mu(f^*\rho_n)(\mu_s)(X_s) \cdot b_s] ds + \frac{1}{2} \int_0^t \mathbb{E}[\partial_x \partial_\mu(f^*\rho_n)(\mu)(X_S) \cdot a_s] ds \\ &\triangleq \int_0^t g_n(s) ds + \frac{1}{2} \int_0^t h_n(s) ds, \end{aligned} \quad (3.25)$$

其中左边收敛到  $f(\mu_t) - f(\mu_0)$ , 右边则有被积函数  $(g_n(s), h_n(s))$  逐点收敛到  $(g(s), h(s)) = (\mathbb{E}[\partial_\mu f(\mu_s)(X_s) \cdot b_s], \mathbb{E}[\partial_x \partial_\mu f(\mu_s)(X_s) \cdot a_s])$ , 下面只需说明  $\{g_n\}, \{h_n\}$  是一致可积的. 事实上, 由  $\{b_s\}, \{\sigma_s\}$  的矩条件可知映射  $s \mapsto \mu_s$  是连续的, 所以  $\{\mu_s\}_{0 \leq s \leq t}$  作为紧集  $[0, t]$  的连续映射下的像仍是紧的, 所以

$$\sup_{n \geq 1} \sup_{s \in [0, t]} (|h_n(s)|^2 + |g_n(s)|^2) < \infty,$$

从而一致可积性成立,也就完成了证明.

■

在这一节的最后,要说明的是虽然完全  $C^2$  正则性假设中要求  $\partial_\mu^2 f(\mu)(x, y)$  的存在性和连续性,但在 Ito 公式中并未出现这一项,实际上对于它的要求确实可以减弱,从而将完全  $C^2$  正则性假设的要求降低到部分  $C^2$  正则性 ((Partial  $C^2$  Regularity) 假设,同时其他设定不变,保持 Ito 公式仍然成立;此外,还可以将  $f : \mathcal{P}_2 \rightarrow \mathbb{R}$  扩展到  $f : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2 \rightarrow \mathbb{R}$ ,给出

$$f(t, X_t, \mathcal{L}(X_t) - f(0, X_0, \mathcal{L}(X_0)))$$

的联合 Ito 公式.具体可见 (Carmona et al., 2018) 的 5.4.5 节和 5.5.5 节相关内容

## 参考文献

- Ciarlet P G, 2017a. 线性与非线性泛函分析及其应用（上册）[M]. 秦铁虎, 童裕孙, 译. 北京: 高等教育出版社. 4, 7, 9, 10
- Ciarlet P G, 2017b. 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. 3, 9, 17
- Yosida K, 2022. 泛函分析[M]. 吴元恺, 孙顺华, 唐志远, 等译. 6 版. 北京: 高等教育出版社. 10
- Bogachev, 2007. Measure theory[M]. Springer Berlin, Heidelberg. 13, 14, 26
- Bogachev V I, 2018. Weak convergence of measures[M]. Providence, Rhode Island : American Mathematical Society. 17
- Carmona R, Delarue F, 2018. Probabilistic theory of mean field games with applications I[M]. Springer Cham. 13, 23, 27
- Ciarlet P G, 2025. Linear and nonlinear functional analysis with applications, second edition[M]. 2nd ed. Philadelphia, PA: Society for Industrial and Applied Mathematics. 3
- Figalli A, Glaudo F, 2021. An invitation to optimal transport, wasserstein distances, and gradient flows[M]. EMS Press. 17
- Fournier N, Guillin A, 2015. On the rate of convergence in wasserstein distance of the empirical measure[J]. Probability Theory and Related Fields. 25