

Derivative Formulas in Measure on Riemannian Manifolds*

Panpan Ren^{b,c)}, Feng-Yu Wang^{a,b)}

^{a)} Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

^{b)} Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, United Kingdom

^{c)} Mathematical Institute, Woodstock Road, OX2 6GG, University of Oxford

673788@swansea.ac.uk, Panpan.ren@maths.ox.ac.uk; wangfy@tju.edu.cn, F.-Y.Wang@swansea.ac.uk

July 6, 2020

Abstract

We characterise the link of derivatives in measure, which are introduced in [2, 3, 8] respectively by different means, for functions on the space \mathbb{M} of finite measures over a Riemannian manifold M . For a reasonable class of functions f , the extrinsic derivative $D^E f$ coincides with the linear functional derivative $D^F f$, the intrinsic derivative $D^I f$ equals to the L -derivative $D^L f$, and

$$D^I f(\eta)(x) = D^L f(\eta)(x) = \lim_{s \downarrow 0} \frac{1}{s} \nabla f(\eta + s\delta_x)(x) = \nabla \{D^E f(\eta)\}(x), \quad (x, \eta) \in M \times \mathbb{M},$$

where ∇ is the gradient on M , δ_x is the Dirac measure at x , and

$$D^E f(\eta)(x) := \lim_{s \downarrow 0} \frac{f(\eta + s\delta_x) - f(\eta)}{s}, \quad x \in M$$

is the extrinsic derivative of f at $\eta \in \mathbb{M}$. This gives a simple way to calculate the intrinsic or L -derivative, and is extended to functions of probability measures.

AMS subject Classification: 60B05, 60B10, 58C35.

Keywords: Intrinsic derivative, extrinsic derivative, Lions derivative, linear functional derivative.

1 Introduction

To develop analysis on the space of measures, some derivatives in measure have been introduced by different means, where the intrinsic and extrinsic derivatives defined in [2, 8] have been used

*Supported in part by NNSFC (11771326, 11831014, 11921001).

to investigate measure-valued diffusion processes over Riemannian manifolds (see [7, 10, 11, 13, 14] and references therein), and the L - and linear functional derivatives were investigated in [3, 4] on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$ (the set of all probability measures on \mathbb{R}^d with finite second-order moments). See [1] and references therein for calculus and optimal transport on the space of probability measures, and see [9, 12] for the Bismut formula and estimates on the L -derivative of distribution dependent SDEs.

In this paper, we aim to clarify the link of these derivatives, and present formulas for calculations. For a broad range of applications, we will work on the space of finite/probability measures over a Riemannian manifold, which includes $\mathcal{P}_2(\mathbb{R}^d)$ as a special example.

Let $(M, \langle \cdot, \cdot \rangle)$ be a complete Riemannian manifold, i.e. M is a differentiable manifold equipped with the Riemannian metric $\langle \cdot, \cdot \rangle$, which is a positive definite smooth bilinear form on the tangent bundle $TM := \cup_{x \in M} T_x M$ ($T_x M$ is the tangent space at point x), such that M is a Polish space under the corresponding Riemannian distance ρ . Let \mathbb{M} denote the class of all nonnegative finite measures on M equipped with the weak topology induced by bounded continuous functions.

For a fixed point $o \in M$, let $\rho_o = \rho(o, \cdot)$ be the Riemannian distance function to o . Denote $\eta(f) = \int_M f d\eta$ for a measure η and a function $f \in L^1(\eta)$. For any $p \in [0, \infty)$, consider the spaces

$$\mathbb{M}_p := \{\eta \in \mathbb{M} : \eta(\rho_o^p) < \infty\}, \quad \mathcal{P}_p := \{\eta \in \mathbb{M}_p : \eta(M) = 1\}, \quad p \in [0, \infty).$$

We will study the above mentioned derivatives on \mathbb{M}_p and \mathcal{P}_p .

For every $p \in [0, \infty)$, \mathbb{M}_p is equipped with the topology that $\eta_n \rightarrow \eta$ in \mathbb{M}_p as $n \rightarrow \infty$ if and only if the convergence holds under the weak topology and

$$\lim_{m \rightarrow \infty} \sup_{n \geq 1} \eta_n(\rho_o^p 1_{\{\rho_o \geq m\}}) = 0.$$

When $p = 0$, this is nothing but the weak topology. When $p > 0$, the topology is induced by the p -Wasserstein metric

$$\mathbb{W}_p(\gamma, \eta) := |\gamma(1 + \rho_o^p) - \eta(1 + \rho_o^p)| + \inf_{\pi \in \mathcal{C}(\gamma, \eta)} \{\pi(\rho^p)\}^{\frac{1}{p+1}},$$

where $\pi \in \mathcal{C}(\gamma, \eta)$ means that π is a finite measure on $M \times M$ such that

$$\pi(M \times \cdot) = \gamma(M)\eta, \quad \pi(\cdot \times M) = \eta(M)\gamma.$$

It is well known that $(\mathbb{M}_p, \mathbb{W}_p)$ is a Polish space for any $p \in [0, \infty)$.

We first recall the extrinsic derivative defined as partial derivative in the direction of Dirac measures, see [8, Definition 1.2].

Definition 1.1 (Extrinsic derivative). Let $p \in [0, \infty)$ and f be a real function on \mathbb{M}_p .

- (1) f is called extrinsically differentiable on \mathbb{M}_p with derivative $D^E f$, if

$$D^E f(\eta)(x) := \lim_{\varepsilon \downarrow 0} \frac{f(\eta + \varepsilon \delta_x) - f(\eta)}{\varepsilon} \in \mathbb{R}$$

exists for all $(x, \eta) \in M \times \mathbb{M}_p$.

- (2) If $D^E f(\eta)(x)$ exists and is continuous in $(x, \eta) \in M \times \mathbb{M}_p$, we denote $f \in C^{E,1}(\mathbb{M}_p)$.
- (3) We denote $f \in C_K^{E,1}(\mathbb{M}_p)$, if $f \in C^{E,1}(\mathbb{M}_p)$ and for any compact set $\mathcal{K} \subset \mathbb{M}_p$, there exists a constant $C > 0$ such that

$$\sup_{\eta \in \mathcal{K}} |D^E f(\eta)(x)| \leq C(1 + \rho_o^p(x)), \quad x \in M.$$

- (4) We denote $f \in C^{E,1,1}(\mathbb{M}_p)$, if $f \in C^{E,1}(\mathbb{M}_p)$ such that $D^E f(\eta)(x)$ is differentiable in x , $\nabla\{D^E f(\eta)(\cdot)\}(x)$ is continuous in $(x, \eta) \in M \times \mathbb{M}_p$, and $|\nabla\{D^E f(\eta)\}| \in L^2(\eta)$ for any $\eta \in \mathbb{M}_p$.
- (5) We write $f \in C_B^{E,1,1}(\mathbb{M}_p)$, if $f \in C^{E,1,1}(\mathbb{M}_p)$ and for any constant $L > 0$ there exists $C_L > 0$ such that

$$\sup_{\eta(\rho_o^p) \leq L} |\nabla\{D^E f(\eta)\}|(x) \leq C_L(1 + \rho_o^p(x)), \quad x \in M.$$

Since for a probability measure μ and $s > 0$, $\mu + s\delta_x$ is no longer a probability measure, for functions of probability measures we modify the definition of the extrinsic derivative with the convex combination $(1 - s)\mu + s\delta_x$ replacing $\mu + s\delta_x$.

Definition 1.2 (Convexity extrinsic derivative). Let $p \in [0, \infty)$ and f be a real function f on \mathcal{P}_p .

- (1) f is called extrinsically differentiable on \mathcal{P}_p , if the centered extrinsic derivative

$$\tilde{D}^E f(\mu)(x) := \lim_{s \downarrow 0} \frac{f((1 - s)\mu + s\delta_x) - f(\mu)}{s} \in \mathbb{R}$$

exists for all $(x, \mu) \in M \times \mathcal{P}_p$.

- (2) We write $f \in C^{E,1}(\mathcal{P}_p)$, if $\tilde{D}^E f(\mu)(x)$ exists and is continuous in $(x, \mu) \in M \times \mathcal{P}_p$.
- (3) We denote $f \in C_K^{E,1}(\mathcal{P}_p)$, if $f \in C^{E,1}(\mathcal{P}_p)$ and for any compact set $\mathcal{K} \subset \mathcal{P}_p$, there exists a constant $C > 0$ such that

$$\sup_{\mu \in \mathcal{K}} |D^E f(\mu)(x)| \leq C(1 + \rho_o^p(x)), \quad x \in M.$$

- (4) We write $f \in C^{E,1,1}(\mathcal{P}_p)$, if $f \in C^{E,1}(\mathcal{P}_p)$ such that $\tilde{D}^E f(\mu)(x)$ is differentiable in $x \in M$, $\nabla\{\tilde{D}^E f(\mu)\}(x)$ is continuous in $(x, \mu) \in M \times \mathcal{P}_p$, and $|\nabla\{\tilde{D}^E f(\mu)\}| \in L^2(\mu)$ for any $\mu \in \mathcal{P}_p$.
- (5) We write $f \in C_B^{E,1,1}(\mathcal{P}_p)$, if $f \in C^{E,1,1}(\mathcal{P}_p)$ and for any constant $L > 0$ there exists $C > 0$ such that

$$\sup_{\mu(\rho_o^p) \leq L} |\nabla\{\tilde{D}^E f(\mu)\}|(x) \leq C(1 + \rho_o^p(x)), \quad x \in M.$$

By Lemma 3.2 below with $\gamma = \delta_x$ and $r = 0$, we have

$$\lim_{s \downarrow 0} \frac{f((1-s)\eta + s\delta_x) - f(\eta)}{s} = D^E f(\eta)(x) - \eta(D^E f(\eta)), \quad f \in C_K^{E,1}(\mathbb{M}_p), x \in M.$$

So, the convexity extrinsic derivative is indeed the centralised extrinsic derivative.

To introduce the intrinsic derivative, for any $v \in \Gamma_0(TM)$, the class of smooth vector fields on M with compact support, consider the flow $(\phi_s^v)_{s \geq 0}$ generated by v :

$$\frac{d}{ds} \phi_s^v = v(\phi_s^v), \quad \phi_0^v = \text{Id}, s \geq 0,$$

where Id is the identity map. Let $\mathcal{B}(TM)$ be the set of all measurable vector fields on M . Then for any $\eta \in \mathbb{M}_p$,

$$L^2(\mathcal{B}(TM); \eta) := \{v \in \mathcal{B}(TM) : \eta(|v|^2) < \infty\}$$

is a Hilbert space, where $|v| := \sqrt{\langle v, v \rangle}$. Then $\Gamma_0(TM)$ is dense in $L^2(\mathcal{B}(TM); \eta)$. When $M = \mathbb{R}^d$, we have $\Gamma_0(TM) = C_0^\infty(\mathbb{R}^d \rightarrow \mathbb{R}^d)$ and $\mathcal{B}(TM) = \mathcal{B}(\mathbb{R}^d \rightarrow \mathbb{R}^d)$.

By the Riesz representation theorem, for any bounded linear functional $U : \Gamma_0(TM) \rightarrow \mathbb{R}$, there exists a unique element $U^* \in L^2(\mathcal{B}(TM); \eta)$ such that

$$U(v) = \langle v, U^* \rangle_{L^2(\eta)} := \int_M \langle v, U^* \rangle d\eta, \quad v \in \Gamma_0(TM).$$

In this case, $U(v) := \int_M \langle v, U^* \rangle d\eta$ for $v \in L^2(\mathcal{B}(TM); \eta)$ is the unique continuous extension of U on $L^2(\mathcal{B}(TM); \eta)$.

Definition 1.3 (Intrinsic derivative). Let $p \in [0, \infty)$ and f be a real function on \mathbb{M}_p .

- (1) f is called intrinsically differentiable on \mathbb{M}_p , if for any $\eta \in \mathbb{M}_p$ and $v \in \Gamma_0(TM)$,

$$D_v^I f(\eta) := \lim_{\varepsilon \downarrow 0} \frac{f(\eta \circ (\phi_\varepsilon^v)^{-1}) - f(\eta)}{\varepsilon} \in \mathbb{R}$$

exists and is a bounded linear functional of $v \in \Gamma_0(TM) \subset L^2(\mathcal{B}(TM); \eta)$. In this case, the unique element $D^I f(\eta) \in L^2(\mathcal{B}(TM); \eta)$ such that

$$D_v^I f(\eta) = \langle D^I f(\eta), v \rangle_{L^2(\eta)} := \int_M \langle D^I f(\eta), v \rangle d\eta, \quad v \in \Gamma_0(TM)$$

is called the intrinsic derivative of f at $\eta \in \mathbb{M}_p$.

- (2) We denote $f \in C^{I,1}(\mathbb{M}_p)$, if f is intrinsically differentiable on \mathbb{M}_p such that $D^I f$ has a version $D^I f(\eta)(x)$ continuous in $(x, \eta) \in M \times \mathbb{M}_p$.

We now introduce the L - and linear functional derivatives following [3, 4] where $\mathcal{P}_2(\mathbb{R}^d)$ is considered. Comparing with the definition of intrinsic derivative, to define the L -derivative one replaces the flow ϕ_s^v by the geodesic flow

$$\phi_{sv}(x) := \exp_x[sv(x)], \quad s \geq 0, x \in M,$$

where

$$\exp_x : T_x M \rightarrow M$$

is the exponential map, so that for each $u \in T_x M$,

$$[0, \infty) \ni s \mapsto \gamma(s) := \exp_x[su] \in M$$

is the unique geodesic starting from x with initial tangent vector $\frac{d}{ds}\gamma(s)|_{s=0} = u$. When $M = \mathbb{R}^d$, we have $\phi_{sv}(x) = x + sv(x)$. By the triangle inequality, we have

$$(1.1) \quad \rho_o(\phi_{sv}(x)) := \rho(o, \phi_{sv}(x)) \leq \rho(o, x) + \rho(x, \exp_x[sv(x)]) \leq \rho_o(x) + |v(x)|, \quad s \in [0, 1].$$

So, when $p \leq 2$, $\eta \in \mathbb{M}_p$ implies

$$\eta \circ \phi_v^{-1} \in \mathbb{M}_p, \quad v \in L^2(\mathcal{B}(TM); \eta).$$

Thus, in the following definition of L -derivative, we assume that $p \leq 2$. See also [5] for a different characterization on the L -derivative and applications to the Hamilton-Jacobi equations on $\mathcal{P}_2(\mathbb{R}^d)$.

Definition 1.4 (L -derivative). Let $p \in [0, 2]$ and f be a real function on \mathbb{M}_p .

- (1) f is called weakly L -differentiable on \mathbb{M}_p , if for any $\eta \in \mathbb{M}_p$ and $v \in L^2(\mathcal{B}(TM); \eta)$,

$$D_v^L f(\eta) := \lim_{\varepsilon \downarrow 0} \frac{f(\eta \circ \phi_{\varepsilon v}^{-1}) - f(\eta)}{\varepsilon} \in \mathbb{R}$$

exists and is a bounded linear functional of $v \in L^2(\mathcal{B}(TM); \eta)$. In this case, the unique element $D^L f(\eta) \in L^2(\mathcal{B}(TM); \eta)$ such that

$$(1.2) \quad D_v^L f(\eta) = \langle D^L f(\eta), v \rangle_{L^2(\eta)}, \quad v \in L^2(\mathcal{B}(TM); \eta)$$

is called the weak L -derivative of f at η .

- (2) f is called L -differentiable on \mathbb{M}_p , if f is weakly L -differentiable with

$$\lim_{\|v\|_{L^2(\eta)} \downarrow 0} \frac{|f(\eta \circ \phi_v^{-1}) - f(\eta) - D_v^L f(\eta)|}{\|v\|_{L^2(\eta)}} = 0, \quad 0 \neq \eta \in \mathbb{M}_p.$$

In this case, we call $D^L f$ the L -derivative of f .

- (3) We denote $f \in C^{L,1}(\mathbb{M}_p)$, if f is L -differentiable on \mathbb{M}_p such that $D^L f$ has a version $D^L f(\eta)(x)$ continuous in $(x, \eta) \in M \times \mathbb{M}_p$.

Definition 1.5 (Linear functional derivative). Let $p \in [0, \infty)$ and f be a real function on \mathbb{M}_p . A measurable function

$$M \ni y \mapsto D^F f(\eta)(y)$$

is called the linear functional derivative of f at $\eta \in \mathbb{M}_p$, if for any constant $L > 0$ there exists a constant $C > 0$ such that

$$(1.3) \quad \sup_{\eta(\rho_o^p) \leq L} \left| D^F f(\eta)(y) \right| \leq C(1 + \rho_o^p(y)), \quad y \in M,$$

and for any $\eta, \gamma \in \mathbb{M}_p$,

$$(1.4) \quad f(\gamma) - f(\eta) = \int_0^1 dr \int_M D^F f(r\gamma + (1-r)\eta)(y)(\gamma - \eta)(dy).$$

Since $(1-s)\mu + s\nu \in \mathcal{P}_p$ for $s \in [0, 1]$ and $\mu, \nu \in \mathcal{P}_p$, the definition of D^F also applies to functions on \mathcal{P}_p .

The remainder of the paper is organized as follows. In Section 2, we state the main results of the paper. Section 3, we present some lemmas which will be used in Sections 4 to prove the main results. The main results of the paper have been reported in the survey [15].

2 Main results

Theorem 2.1. *Let $p \in [0, \infty)$.*

- (1) *If f is L -differentiable on \mathbb{M}_p , then it is intrinsic differentiable and $D^I f = D^L f$.*
- (2) *If $f \in C_K^{E,1}(\mathbb{M}_p)$, then f has linear functional derivative $D^F f = D^E f$.*
- (3) *Let $f \in C^{E,1,1}(\mathbb{M}_p)$. Then $f \in C^{I,1}(\mathbb{M}_p)$ with*

$$(2.1) \quad D^I f(\eta)(x) = \nabla \{ D^E f(\eta)(\cdot) \}(x), \quad (x, \eta) \in M \times \mathbb{M}_p.$$

When $p \in [0, 2]$ and $f \in C_B^{E,1,1}(\mathbb{M}_p)$, we have $f \in C^{L,1}(\mathbb{M}_p)$ and

$$(2.2) \quad D^L f(\eta)(x) = \nabla \{ D^E f(\eta)(\cdot) \}(x), \quad (x, \eta) \in M \times \mathbb{M}_p.$$

- (4) *If $f \in C^{L,1}(\mathbb{M}_p)$, then for any $s \geq 0$, $f(\eta + s\delta) \in C^1(M)$ with*

$$(2.3) \quad \nabla f(\eta + s\delta)(x) = s D^L f(\eta + s\delta_x)(x), \quad x \in M, s \geq 0.$$

Consequently,

$$(2.4) \quad D^L f(\eta)(x) = \lim_{s \downarrow 0} \frac{1}{s} \nabla f(\eta + s\delta)(x), \quad f \in C^{L,1}(\mathbb{M}_p), (x, \eta) \in M \times \mathbb{M}_p.$$

Remark 1.1. (a) Theorem 2.1(3) implies $C_B^{E,1,1}(\mathbb{M}_p) \subset C^{L,1}(\mathbb{M}_p)$, $p \in [0, 2]$. However, a function $f \in C^{L,1}(\mathbb{M}_p)$ is not necessarily extrinsically differentiable. For instance, let $\psi \in C([0, \infty))$ but not differentiable, and let $f(\eta) = \psi(\eta(M))$. Then $f(\eta + s\delta_x) = \psi(\eta(M) + s)$ which is not differentiable in s , so that f is not extrinsically differentiable. But it is easy to see that $f \in C^{L,1}(\mathbb{M}_p)$ with $D^L f(\eta) = 0$. Of course, this counter-example does not work for functions on the space of probability measures

(b) According to [4, Proposition 5.48], if f is a function on $\mathcal{P}_2(\mathbb{R}^d)$ having linear functional derivative $D^F f(\mu) \in C^1(M)$ for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, then f is L -differentiable and

$$(2.5) \quad D^L f(\eta) = \nabla\{D^F f(\eta)\}.$$

By Theorem 2.1(1)-(3), this formula (2.5) is extended to (2.2) for the present general framework. Since the definition of D^E is more straightforward than that of D^F , (2.2) is more explicit than (2.5). Note that in [4] the weak L -derivative is named by intrinsic derivative, where the latter was however introduced much earlier by [2] as in Definition 1.3.

(c) To illustrate the link between derivatives presented in Theorem 2.1, let us consider the class of cylindrical functions $\mathcal{F}C_b^1$, which consists of functions of type

$$f(\eta) := g(\eta(h_1), \dots, \eta(h_n)), \quad \eta(\eta_i) := \int_M h_i d\eta, \quad \eta \in \mathbb{M},$$

where $n \geq 1$, $g \in C_b^1(\mathbb{R}^n)$ and $h_i \in C_b^1(M)$, $1 \leq i \leq n$. Then f is extrinsically and L -differentiable, and has linear functional derivative:

$$D^L f(\eta) = \sum_{i=1}^n (\partial_i g)(\eta(h_1), \dots, \eta(h_n)) \nabla h_i, \quad D^E f(\eta) = D^F f(\eta) = \sum_{i=1}^n (\partial_i g)(\eta(h_1), \dots, \eta(h_n)) h_i,$$

where ∇ is the gradient operator on M . Therefore, we have

$$D^L f(\eta)(x) = \nabla\{D^E f(\eta)(\cdot)\}(x), \quad (x, \eta) \in M \times \mathbb{M}_p, \quad p \in [0, \infty)$$

as indicated in (2.2).

Next, we consider derivatives on the space $\mathcal{P}_p := \mathbb{M}_p \cap \mathcal{P}$ for $p \in [0, \infty)$. Since for any $\mu \in \mathcal{P}_p$ and any $v \in \Gamma_0(TM)$, we have $\mu \circ \phi_{\varepsilon v}^{-1}, \mu \circ (\phi_{\varepsilon}^v)^{-1} \in \mathcal{P}_p$ for $\varepsilon \geq 0$. So, the definitions of D^I and D^L work also for functions on \mathcal{P}_p , and we define the classes $C^{I,1}(\mathcal{P}_p)$ and $C^{L,1}(\mathcal{P}_p)$ as in Definitions 1.2 and 1.3 for \mathcal{P}_p replacing \mathbb{M}_p .

By extending a function on \mathcal{P}_p to \mathbb{M}_p , we may apply Theorem 2.1 to establish the corresponding link for functions on \mathcal{P}_2 . As an application, we will present derivative formula for the distribution of random variables. For $s_0 > 0$ and a family of M -valued random variables $\{\xi_s\}_{s \in [0, s_0]}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that $\dot{\xi}_0 := \frac{d}{ds} \xi_s|_{s=0}$ exists in $L^q(\Omega \rightarrow TM; \mathbb{P})$ for some $q \geq 1$, if $\dot{\xi}_0 \in T_{\xi_0} M$ with $\mathbb{E}|\dot{\xi}_0|^q < \infty$ such that

$$(2.6) \quad \lim_{s \downarrow 0} \mathbb{E} \left| \frac{1}{s} \exp_{\xi_0}^{-1}[\xi_s] - \dot{\xi}_0 \right|^q = 0.$$

Since $\xi_s \rightarrow \xi_0$ as $s \rightarrow 0$, note that the inverse of the exponential map $\exp_{\xi_0}^{-1}[\xi_s]$ is well-defined for small $s > 0$, see the proof of Theorem 2.1(1) below for details. In particular, for $M = \mathbb{R}^d$ we have $\exp_{\xi_0}^{-1}[\xi_s] = \xi_s - \xi_0$.

Corollary 2.2. *Let $p \in [0, \infty)$.*

- (1) *If f is L -differentiable on \mathcal{P}_p , then it is intrinsic differentiable and $D^I f = D^L f$.*
- (2) *If $f \in C_p^{E,1}(\mathcal{P}_p)$, then f has linear functional derivative on \mathcal{P}_p and $D^F f = \tilde{D}^E f$.*
- (3) *Let $f \in C^{E,1,1}(\mathcal{P}_p)$. Then $f \in C^{I,1}(\mathcal{P}_p)$ and*

$$(2.7) \quad D^I f(\mu)(x) = \nabla \{ \tilde{D}^E f(\mu)(\cdot) \}(x), \quad (x, \mu) \in M \times \mathcal{P}_p.$$

When $p \leq 2$ and $f \in C_B^{I,1}(\mathcal{P}_p)$, we have $f \in C^{L,1}(\mathcal{P}_p)$ with

$$D^L f = \nabla \{ \tilde{D}^E f(\mu)(\cdot) \}, \quad \mu \in \mathcal{P}_p, f \in C_B^{E,1,1}(\mathcal{P}_p).$$

- (4) *If $f \in C^{L,1}(\mathcal{P}_p)$, then $f((1-s)\mu + s\delta_x) \in C^1(M)$ with*

$$(2.8) \quad \nabla f((1-s)\mu + s\delta_x)(x) = s D^L f((1-s)\mu + s\delta_x)(x), \quad x \in M.$$

Consequently,

$$(2.9) \quad D^L f(\mu)(x) = \lim_{s \downarrow 0} \frac{1}{s} \nabla f((1-s)\mu + s\delta_x)(x), \quad f \in C^{L,1}(\mathcal{P}_p), (x, \eta) \in M \times \mathbb{M}.$$

- (5) *Let $\{\xi_s\}_{s \in [0, s_0]}$ be random variables on M with $\mathcal{L}_{\xi_s} \in \mathcal{P}_p$ continuous in s , such that $\dot{\xi}_0 := \frac{d}{ds} \xi_s \Big|_{s=0}$ exists in $L^q(\Omega \rightarrow TM; \mathbb{P})$ for some $q \geq 1$. Then*

$$(2.10) \quad \lim_{s \downarrow 0} \frac{f(\mathcal{L}_{\xi_s}) - f(\mathcal{L}_{\xi_0})}{s} = \mathbb{E} \langle D^L f(\mathcal{L}_{\xi_0}) \rangle (\xi_0, \dot{\xi}_0)$$

holds for any $f \in C^{E,1,1}(\mathcal{P}_p)$ such that for any compact set $\mathcal{K} \subset \mathcal{P}_p$,

$$(2.11) \quad \sup_{\mu \in \mathcal{K}} |\nabla \{ \tilde{D}^E f(\mu) \}|(x) \leq C(1 + \rho_o)^{\frac{p(q-1)}{q}}, \quad x \in M$$

holds for some constant $C > 0$.

Let us compare (2.10) with the corresponding formula presented in [3] for $M = \mathbb{R}^d$, $\rho(x, y) = |x - y|$ and $p = 2$. In this case, the formula (2.10) is established for the probability space being Polish and $f \in C^{L,1}(\mathcal{P}_2(\mathbb{R}^d))$ with bounded $D^L f$, see also [6, Proposition A.2] and [14, Lemma 2.3] for this formula with more general functions f on $\mathcal{P}_2(\mathbb{R}^d)$. Theorem 2.2 establishes (2.10) to \mathbb{M}_p on Riemannian manifolds and $p \geq 0$.

3 Some lemmas

We first consider the variation of $f(h\eta)$ in the density function $0 \leq h \in L^1(\eta)$. Recall that for a nonnegative measurable function h on M , the measure $h\eta$ is defined by

$$(h\eta)(A) := \int_A h d\eta, \quad A \in \mathcal{B}(M),$$

where $\mathcal{B}(M)$ is the Borel σ -field of M . In the proof of Theorem 2.1, we will formulate $f(\eta \circ \phi_{\varepsilon v}^{-1})$ by $f((1 + h_\varepsilon)\eta)$ for some $h \in \mathcal{H}_{\varepsilon_0}$, where $h \in \mathcal{H}_{\varepsilon_0}$ means that $h \in C([0, \varepsilon_0] \times M; [0, \infty))$ and

(1) $h_0 = 0$, $\sup_{\varepsilon \in [0, \varepsilon_0]} \|h_\varepsilon\|_\infty < \infty$, $\text{supp} h_\varepsilon \subset K$ for some compact set $K \subset M$ and all $\varepsilon \in [0, \varepsilon_0]$;

(2) $\dot{h}_\varepsilon := \lim_{s \downarrow 0} \frac{h_{\varepsilon+s} - h_\varepsilon}{s} \in C_b(M)$ exists and is uniformly bounded for $\varepsilon \in [0, \varepsilon_0]$.

So, to calculate $D^L f(\mu)$, we first present the following lemma which links $f((1 + h_\varepsilon)\eta) - f(\eta)$ to the extrinsic derivative.

Lemma 3.1. *Let $p \in [0, \infty)$. For any $h \in \mathcal{H}_{\varepsilon_0}$ and any $f \in C^{E,1,1}(\mathbb{M}_p)$,*

$$(3.1) \quad f((1 + h_\varepsilon)\eta) - f(\eta) = \int_0^\varepsilon dr \int_M D^E f((1 + h_r)\eta)(x) \dot{h}_r(x) \eta(dx), \quad \eta \in \mathbb{M}_p, \varepsilon \in [0, \varepsilon_0].$$

Proof. (1) We first consider

$$\eta \in \mathbb{M}_{disc} := \left\{ \sum_{i=1}^n a_i \delta_{x_i} : n \geq 1, a_i > 0, x_i \in M, 1 \leq i \leq n \right\}.$$

In this case, for any $\varepsilon \in [0, \varepsilon_0)$ and $s \in (0, \varepsilon_0 - \varepsilon)$, by the definition of D^E we have

$$\begin{aligned} f((1 + h_{\varepsilon+s})\eta) - f((1 + h_\varepsilon)\eta) &= f\left((1 + h_\varepsilon)\eta + \sum_{i=1}^n \{h_{\varepsilon+s} - h_\varepsilon\}(x_i) a_i \delta_{x_i}\right) - f((1 + h_\varepsilon)\eta) \\ &= \sum_{k=1}^n \left\{ f\left((1 + h_\varepsilon)\eta + \sum_{i=1}^k \{h_{\varepsilon+s} - h_\varepsilon\}(x_i) a_i \delta_{x_i}\right) - f\left((1 + h_\varepsilon)\eta + \sum_{i=1}^{k-1} \{h_{\varepsilon+s} - h_\varepsilon\}(x_i) a_i \delta_{x_i}\right) \right\} \\ &= \sum_{k=1}^n a_k \int_{-a_k \{h_{\varepsilon+s} - h_\varepsilon\}^-(x_k)}^{a_k \{h_{\varepsilon+s} - h_\varepsilon\}^+(x_k)} \left\{ D^E f\left((1 + h_\varepsilon)\eta + \sum_{i=1}^{k-1} \{h_{\varepsilon+s} - h_\varepsilon\}(x_i) a_i \delta_{x_i} + r \delta_{x_k}\right) \right\}(x_k) dr, \end{aligned}$$

where $\sum_{i=1}^0 := 0$, $a^+ := \max\{a, 0\}$ and $a^- := (-a)^+$ for $a \in \mathbb{R}$. Multiplying by s^{-1} and letting $s \downarrow 0$, we deduce from this and the continuity of $D^E f$ that

$$(3.2) \quad \begin{aligned} \lim_{s \downarrow 0} \frac{f((1 + h_{\varepsilon+s})\eta) - f((1 + h_\varepsilon)\eta)}{s} &= \sum_{k=1}^n a_k \{ \dot{h}_\varepsilon(x_k)^+ - \dot{h}_\varepsilon(x_k)^- \} D^E f((1 + h_\varepsilon)\eta)(x_k) \\ &= \int_M D^E f((1 + h_\varepsilon)\eta)(x) \dot{h}_\varepsilon(x) \eta(dx), \quad \varepsilon \in [0, \varepsilon_0), \eta \in \mathbb{M}_{disc}. \end{aligned}$$

(2) In general, for any $\eta \in \mathbb{M}_p$, let $\{\eta_n\}_{n \geq 1} \subset \mathbb{M}_{disc}$ such that $\eta_n \rightarrow \eta$ in \mathbb{M}_p . By (3.2), for any $\varepsilon \in (0, \varepsilon_0)$ and $s \in (0, \varepsilon_0 - \varepsilon)$, we have

$$(3.3) \quad f((1 + h_\varepsilon)\eta_n) - f(\eta_n) = \int_0^\varepsilon dr \int_M D^E f((1 + h_r)\eta_n)(x) \dot{h}_r(x) \eta_n(dx), \quad n \geq 1.$$

Next, since $D^E f \in C(M \times \mathbb{M}_p)$ and $h_r, \dot{h}_r \in C_b(M)$ for $r \in [0, \varepsilon_0]$ with compact support $\subset K$, and $\eta_n \rightarrow \eta$ in \mathbb{M}_p , we obtain

$$(3.4) \quad \lim_{n \rightarrow \infty} \int_M D^E f((1 + h_r)\eta)(x) \dot{h}_r(x) \eta_n(dx) = \int_M D^E f((1 + h_r)\eta)(x) \dot{h}_r(x) \eta(dx).$$

Moreover, $\eta_n \rightarrow \eta$ in \mathbb{M}_p and $h \in \mathcal{H}_{\varepsilon_0}$ imply that the set

$$\mathcal{K}_r := \{(1 + h_r)\eta, (1 + h_r)\eta_n : n \geq 1\}$$

is compact in \mathbb{M}_p for any $r \in [0, \varepsilon_0]$. Combining this with $D^E f \in C(M \times \mathbb{M}_p)$, we see that the function

$$\mathcal{K}_r \times M \ni (\gamma, x) \mapsto D^E f(\gamma)(x) \dot{h}_r(x)$$

is uniformly continuous and has compact support $\subset \mathcal{K}_r \times K$, so that (3.4) implies

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left| \int_M D^E f((1 + h_r)\eta_n)(x) \dot{h}_r(x) \eta_n(dx) - \int_M D^E f((1 + h_r)\eta)(x) \dot{h}_r(x) \eta(dx) \right| \\ &= \limsup_{n \rightarrow \infty} \left| \int_M D^E f((1 + h_r)\eta_n)(x) \dot{h}_r(x) \eta_n(dx) - \int_M D^E f((1 + h_r)\eta)(x) \dot{h}_r(x) \eta_n(dx) \right| \\ &\leq \limsup_{n \rightarrow \infty} \left\{ \eta_n(K) \sup_{x \in K} |D^E f((1 + h_r)\eta_n)(x) \dot{h}_r(x) - D^E f((1 + h_r)\eta)(x) \dot{h}_r(x)| \right\} \\ &= 0. \end{aligned}$$

Combining this with

$$\sup_{(\gamma, x) \in \mathcal{K}_r \times K, r \in [0, \varepsilon_0]} |D^E f(\gamma)(x) \dot{h}_r(x)| < \infty,$$

we deduce from the dominated convergence theorem that

$$\begin{aligned} (3.5) \quad & \lim_{n \rightarrow \infty} \int_0^\varepsilon dr \int_M \{D^E f\}((1 + h_\varepsilon)\eta_n)(x) \dot{h}_r(x) \eta_n(dx) \\ &= \int_0^\varepsilon dr \int_M \{D^E f\}((1 + h_r)\eta)(x) \dot{h}_r(x) \eta(dx). \end{aligned}$$

Therefore, by letting $n \rightarrow \infty$ in (3.3) and using the continuity of f , we prove (3.1). \square

To calculate the convexity extrinsic derivative, we present the following result.

Lemma 3.2. *Let $p \in [0, \infty)$. Then for any $f \in C_K^{E,1}(\mathbb{M}_p)$ and $\eta, \gamma \in \mathbb{M}_p$,*

$$\begin{aligned} \frac{d}{dr} f((1 - r)\eta + r\gamma) &:= \lim_{\varepsilon \downarrow 0} \frac{f((1 - r - \varepsilon)\eta + (r + \varepsilon)\gamma) - f((1 - r)\eta + r\gamma)}{\varepsilon} \\ &= \int_M \{D^E f((1 - r)\eta + r\gamma)(x)\} (\gamma - \eta)(dx), \quad r \in [0, 1]. \end{aligned}$$

Consequently, for any $f \in C_K^{E,1}(\mathbb{M}_p)$,

$$\begin{aligned} \tilde{D}^E f(\eta)(x) &:= \lim_{s \downarrow 0} \frac{f((1 - s)\eta + s\delta_x) - f(\eta)}{s} \\ &= D^E f(\eta)(x) - \eta(D^E f(\eta)), \quad (x, \eta) \in M \times \mathbb{M}_p. \end{aligned}$$

The assertions also hold for \mathcal{P}_p replacing \mathbb{M}_p .

Proof. As in the proof of Lemma 3.1, we take

$$\eta_n = \sum_{i=1}^n \alpha_{n,i} \delta_{x_{n,i}}, \quad \gamma_n = \sum_{i=1}^n \beta_{n,i} \delta_{x_{n,i}}$$

for some $x_{n,i} \in M$ and $\alpha_{n,i}, \beta_{n,i} \geq 0$, such that

$$\eta_n \rightarrow \eta, \quad \gamma_n \rightarrow \gamma \quad \text{in } \mathbb{M}_p \text{ as } n \rightarrow \infty.$$

For any $r \in [0, 1)$ and $\varepsilon \in (0, 1 - r)$, let

$$\Lambda_{n,i}^\varepsilon := (1 - r)\eta_n + r\gamma_n + \sum_{k=1}^{i-1} \varepsilon(\beta_k - \alpha_k) \delta_{x_{n,k}} \in \mathbb{M}_p, \quad 1 \leq i \leq n,$$

where by convention $\sum_{i=1}^0 := 0$. Then by the definition of $D^E f$, we have

$$\begin{aligned} & f((1 - r - \varepsilon)\eta_n + (r + \varepsilon)\gamma_n) - f((1 - r)\eta_n + r\gamma_n) \\ &= \sum_{i=1}^n \{f(\Lambda_{n,i}^\varepsilon + \varepsilon(\beta_{n,i} - \alpha_{n,i})\delta_{x_{n,i}}) - f(\Lambda_{n,i}^\varepsilon)\} \\ &= \sum_{i=1}^n \int_{-\varepsilon(\beta_{n,i} - \alpha_{n,i})^-}^{\varepsilon(\beta_{n,i} - \alpha_{n,i})^+} D^E f(\Lambda_{n,i}^\varepsilon + s\delta_{x_{n,i}})(x_{n,i}) ds, \quad \varepsilon \in (0, 1 - r). \end{aligned}$$

Multiplying by ε^{-1} and letting $\varepsilon \downarrow 0$, due to the continuity of $D^E f$ we derive

$$\begin{aligned} \frac{d}{dr} f((1 - r)\eta_n + r\gamma_n) &= \sum_{i=1}^n (\beta_{n,i} - \alpha_{n,i}) D^E f((1 - r)\eta_n + r\gamma_n)(x_{n,i}) \\ &= \int_M \{D^E f((1 - r)\eta_n + r\gamma_n)(x)\} (\gamma_n - \eta_n)(dx), \quad r \in [0, 1), \quad n \geq 1. \end{aligned}$$

Consequently, for any $r \in [0, 1)$,

$$\begin{aligned} & f((1 - r - \varepsilon)\eta_n + (r + \varepsilon)\gamma_n) - f((1 - r)\eta_n + r\gamma_n) \\ &= \int_r^{r+\varepsilon} ds \int_M \{D^E f((1 - s)\eta_n + s\gamma_n)(x)\} (\gamma_n - \eta_n)(dx), \quad \varepsilon \in (0, 1 - r), \quad n \geq 1. \end{aligned}$$

Noting that the set $\{\eta_n, \gamma_n : n \geq 1\}$ is relatively compact in \mathbb{M}_p , by this and the condition on f , we may let $n \rightarrow \infty$ to derive

$$\begin{aligned} & f((1 - r - \varepsilon)\eta + (r + \varepsilon)\gamma) - f((1 - r)\eta + r\gamma) \\ &= \int_r^{r+\varepsilon} ds \int_M \{D^E f((1 - s)\eta + s\gamma)(x)\} (\gamma - \eta)(dx), \quad \varepsilon \in (0, 1 - r). \end{aligned}$$

Multiplying by ε^{-1} and letting $\varepsilon \downarrow 0$, we finish the proof. \square

The following is a consequence of Lemma 3.2 for functions on \mathcal{P}_p .

Lemma 3.3. *Let $p \in [0, \infty)$. Then for any $f \in C_K^{E,1}(\mathcal{P}_p)$ and $\mu, \nu \in \mathcal{P}_p$,*

$$\lim_{s \downarrow 0} \frac{f((1-s)\mu + s\nu) - f(\mu)}{s} = \int_M \{\tilde{D}^E f((\mu)(x))\}(\nu - \mu)(dx).$$

Proof. To apply Lemma 3.2, we extend a function f on \mathcal{P}_p to \tilde{f} on \mathbb{M}_p by letting

$$\tilde{f}(\eta) = h(\eta(M))f(\eta/\eta(M)), \quad \eta \in \mathbb{M}_p,$$

where $h \in C_0^\infty(\mathbb{R})$ with support contained by $[\frac{1}{4}, 2]$ and $h(r) = 1$ for $r \in [\frac{1}{2}, \frac{3}{2}]$. It is easy to see that

$$f((1-s)\mu + s\nu) = \tilde{f}((1-s)\mu + s\nu), \quad s \in [0, 1], \mu, \nu \in \mathcal{P}_p,$$

and $f \in C_K^{E,1}(\mathcal{P}_p)$ implies that $\tilde{f} \in C_K^{E,1}(\mathbb{M}_p)$ and

$$D^E \tilde{f}(\mu) = \tilde{D}^E f(\mu), \quad \mu \in \mathcal{P}.$$

Then the desired formula is implied by Lemma 3.2 with $r = 0$. □

Finally, we prove a derivative formula for the distribution of random variables.

Lemma 3.4. *Let $\{\xi_s\}_{s \in [0, s_0]}$ be M -valued random variables such that $\lim_{s \rightarrow 0} \mathcal{L}_{\xi_s} = \mathcal{L}_{\xi_0}$ in \mathcal{P}_p , and $\dot{\xi}_0 := \frac{d}{ds} \xi_s \Big|_{s=0}$ exists in $L^q(\Omega \rightarrow TM; \mathbb{P})$ for some $q \geq 1$. Then*

$$(3.6) \quad \lim_{s \downarrow 0} \frac{f(\mathcal{L}_{\xi_s}) - f(\mathcal{L}_{\xi_0})}{s} = \mathbb{E} \langle \nabla \{ \tilde{D}^E f(\mathcal{L}_{\xi_0}) \}(\xi_0), \dot{\xi}_0 \rangle$$

holds for functions $f \in C^{E,1,1}(\mathcal{P}_p)$ satisfying (2.11) for any compact $\mathcal{K} \subset \mathcal{P}_p$ and some constant $C = C(\mathcal{K}) > 0$.

Proof. By Lemma 3.3, we have

$$(3.7) \quad \begin{aligned} f(\mathcal{L}_{\xi_s}) - f(\mathcal{L}_{\xi_0}) &= \int_0^1 \left\{ \frac{d}{dr} f(r\mathcal{L}_{\xi_s} + (1-r)\mathcal{L}_{\xi_0}) \right\} dr \\ &= \int_0^1 dr \int_M \{ \tilde{D}^E f(r\mathcal{L}_{\xi_s} + (1-r)\mathcal{L}_{\xi_0}) \}(x) (\mathcal{L}_{\xi_s} - \mathcal{L}_{\xi_0})(dx) \\ &= \int_0^1 \mathbb{E} \left[\{ \tilde{D}^E f(r\mathcal{L}_{\xi_s} + (1-r)\mathcal{L}_{\xi_0}) \}(\xi_s) - \{ \tilde{D}^E f(r\mathcal{L}_{\xi_s} + (1-r)\mathcal{L}_{\xi_0}) \}(\xi_0) \right] dr. \end{aligned}$$

For each $s \geq 0$, let $\gamma_{s,\cdot} : [0, 1] \rightarrow M$ be the minimal geodesic such that $\gamma_{s,0} = \xi_0$ and $\gamma_{s,1} = \xi_s$. Then $\lim_{s \downarrow 0} \gamma_{s,\theta} = \xi_0$, and by (2.6),

$$\lim_{s \downarrow 0} \mathbb{E} \left| \frac{1}{s} //_{\theta \rightarrow 0} \frac{d}{d\theta} \gamma_{s,\theta} - \dot{\xi}_0 \right|^q = 0,$$

where $//_{\theta \rightarrow 0} : T_{\gamma_{s,\theta}}M \rightarrow T_{\xi_0}M$ is the parallel displacement along the geodesic $\gamma_{s,\cdot} : [0, \theta] \rightarrow M$. Combining these with (3.7) and (2.11) with $\mathcal{K} := \{\mathcal{L}_{\xi_0}, \mathcal{L}_{\xi_{s_n}} : n \geq 1\}$ for a sequence $s_0 > s_n \downarrow 0$, we may apply the dominated convergence theorem to derive

$$\begin{aligned}
& \lim_{s_n \downarrow 0} \frac{f(\mathcal{L}_{\xi_{s_n}}) - f(\mathcal{L}_{\xi_0})}{s_n} \\
&= \lim_{s_n \downarrow 0} \frac{1}{s_n} \int_0^1 \mathbb{E} \left[\{D^E f(r\mathcal{L}_{\xi_{s_n}} + (1-r)\mathcal{L}_{\xi_0})\}(\xi_{s_n}) - \{\tilde{D}^E f(r\mathcal{L}_{\xi_{s_n}} + (1-r)\mathcal{L}_{\xi_0})\}(\xi_0) \right] dr \\
&= \lim_{s_n \downarrow 0} \int_0^1 dr \int_0^1 \mathbb{E} \left[\langle \nabla \{ \tilde{D}^E f(r\mathcal{L}_{\xi_{s_n}} + (1-r)\mathcal{L}_{\xi_0}) \}(\gamma_{s_n, \theta}), \frac{1}{s} \frac{d}{d\theta} \gamma_{s_n, \theta} \rangle \right] d\theta \\
&= \mathbb{E} \langle \nabla \{ \tilde{D}^E f(\mathcal{L}_{\xi_0}) \}(\xi_0), \dot{\xi}_0 \rangle.
\end{aligned}$$

□

4 Proofs of Theorem 2.1 and Corollary 2.2

Obviously, assertion (2) follows from Lemma 3.2. Below we prove assertions (1), (3), (4) in Theorem 2.1 as well as Corollary 2.2 respectively.

Proof of Theorem 2.1(1). Although the flows ϕ_s^v and ϕ_{sv} are different, their derivative at $s = 0$ are all equal to v , so that both D_v^I and D_v^L are directional derivatives along v . Thus, it is reasonable that for a large class of functions we have $D^I f = D^L f$. To see this, we need the inverse exponential map \exp_x^{-1} . For any $z \in M$, let $u \in T_x M$ such that

$$[0, 1] \ni s \mapsto \exp_x[su] \in M$$

is the minimal geodesic from x to z , and we denote $u = \exp_x^{-1}[z]$. If z is not in the cut-locus of x , the minimal geodesic from x to z is unique, and $\exp_x^{-1}[z]$ is smooth in z . In case that z belongs to the cut-locus of x , such a vector $u \in T_x M$ may be not unique. For any compact set $\mathcal{K} \subset M$, there exists a constant $R > 0$ such that for any $x \in \mathcal{K}$, the distance between x and its cut-locus is larger than R . So, for any $x \in \mathcal{K}$,

$$\exp_x : \{u \in T_x M : |u| \leq R\} \rightarrow B_x(R) := \{y \in M : \rho(x, y) \leq R\}$$

is a diffeomorphism, such that

$$\exp_x^{-1} : B_x(R) \rightarrow T_x M$$

is smooth. Thus, for any $v \in \Gamma_0(TM)$ and small enough $\varepsilon > 0$, we have $v_\varepsilon := \varepsilon^{-1} \exp^{-1}[\phi_\varepsilon^v] \in \Gamma_0(TM)$. Moreover,

$$v_\varepsilon = \varepsilon v + o(\varepsilon),$$

where $\varepsilon^{-1} \|o(\varepsilon)\|_\infty \rightarrow 0$ as $\varepsilon \downarrow 0$. Hence, for any L -differentiable function f and $\eta \in \mathbb{M}$, when ε is small enough we have

$$\limsup_{\varepsilon \downarrow 0} \left| \frac{f(\eta \circ (\phi_\varepsilon^v)^{-1}) - f(\eta)}{\varepsilon} - D_v^L f(\eta) \right| = \limsup_{\varepsilon \downarrow 0} \left| \frac{f(\eta \circ \phi_{v_\varepsilon}^{-1}) - f(\eta)}{\varepsilon} - D_v^L f(\eta) \right|$$

$$\leq \limsup_{\varepsilon \downarrow 0} \left\{ \left| \frac{f(\eta \circ \phi_{v_\varepsilon}^{-1}) - f(\eta) - D_{v_\varepsilon}^L f(\eta)}{\varepsilon} \right| + |D_{v-\varepsilon^{-1}v_\varepsilon}^L f(\eta)| \right\} = 0.$$

Therefore, $D^I f = D^L f$ holds for L -differentiable f . \square

Proof of Theorem 2.1(3). It suffices to prove the formulas (2.1) and (2.2) for $f \in C^{E,1,1}(\mathbb{M}_p)$ and $f \in C^{E,1,1}(\mathbb{M}_p)$ respectively.

(a) For (2.1). Since any $\eta \in \mathbb{M}_p$ can be approximated by those having smooth and strictly positive density functions with respect to the volume measure dx , by the argument leading to (3.5), it suffices to show that for any $\eta \in \mathbb{M}_p$ satisfying

$$(4.1) \quad \eta(dx) = \rho(x)dx \text{ for some } \rho \in C_b^\infty(M), \inf \rho > 0,$$

there exists a constant $\varepsilon_0 > 0$ such that

$$(4.2) \quad f(\eta \circ (\phi_\varepsilon^v)^{-1}) - f(\eta) = \int_0^\varepsilon dr \int_M \langle \nabla \{D^E f(\eta \circ \phi_{rv}^{-1})\}, v \rangle d(\eta \circ \phi_{rv}^{-1}), \quad \varepsilon \in (0, \varepsilon_0).$$

Firstly, there exists a constant $\varepsilon_0 > 0$ such that

$$\rho_\varepsilon^v := \frac{d(\eta \circ (\phi_\varepsilon^v)^{-1})}{d\eta}, \quad \dot{\rho}_\varepsilon^v := \lim_{s \downarrow 0} \frac{\rho_{\varepsilon+s}^v - \rho_\varepsilon^v}{s}$$

exist in $C_b(M)$ and are uniformly bounded and continuous in $\varepsilon \in [0, \varepsilon_0]$. Next, by Lemma 3.1, we have

$$(4.3) \quad f(\eta \circ (\phi_\varepsilon^v)^{-1}) - f(\eta) = \int_0^\varepsilon dr \int_M \{D^E f(\eta \circ (\phi_r^v)^{-1})\} \dot{\rho}_r^v d\eta, \quad \varepsilon \in [0, \varepsilon_0].$$

To calculate $\dot{\rho}_r^v$, by $\frac{d}{ds} \phi_s^v = v(\phi_s^v)$, for any $g \in C_0^\infty(M)$ we have

$$\frac{d}{dr} \{g \circ \phi_r^v\} = \langle \nabla g(\phi_r^v), v(\phi_r^v) \rangle = \langle \nabla g, v \rangle(\phi_r^v), \quad r \geq 0,$$

which is smooth and bounded in $(r, x) \in [0, \varepsilon_0] \times M$. So,

$$\begin{aligned} \int_M g \dot{\rho}_r^v d\eta &= \int_M g \lim_{s \downarrow 0} \frac{\rho_{r+s}^v - \rho_r^v}{s} d\eta = \lim_{s \downarrow 0} \frac{1}{s} \int_M g d\{\eta \circ (\phi_{r+s}^v)^{-1} - \eta \circ (\phi_r^v)^{-1}\} \\ &= \lim_{s \downarrow 0} \frac{1}{s} \int_M \{g \circ \phi_{r+s}^v - g \circ \phi_r^v\} d\eta = \int_M \frac{d}{dr} (g \circ \phi_r^v) d\eta \\ &= \int_M \langle \nabla g, v \rangle \circ \phi_r^v d\eta = \int_M \langle \nabla g, v \rangle d(\eta \circ (\phi_r^v)^{-1}) \\ &= - \int_M \{g \operatorname{div}_{\eta \circ (\phi_r^v)^{-1}}(v)\} d(\eta \circ (\phi_r^v)^{-1}) = - \int_M g \{ \operatorname{div}_{\eta \circ (\phi_r^v)^{-1}}(v) \rho_r^v \} d\eta, \quad g \in C_0^\infty(M), \end{aligned}$$

where $\operatorname{div}_{\eta \circ (\phi_r^v)^{-1}}(v) = \operatorname{div}(v) + \langle v, \nabla \log(\rho_r^v \rho) \rangle$. This implies $\dot{\rho}_r^v = -\operatorname{div}_{\eta \circ (\phi_r^v)^{-1}}(v) \rho_r^v$, so that the integration by parts formula and $\rho_r^v \eta = \eta \circ (\phi_r^v)^{-1}$ lead to

$$\int_M \{D^E f(\eta \circ (\phi_r^v)^{-1})\} \dot{\rho}_r^v d\eta = - \int_M \{D^E f(\eta \circ (\phi_r^v)^{-1})\} \operatorname{div}_{\eta \circ (\phi_r^v)^{-1}}(v) d(\eta \circ (\phi_r^v)^{-1})$$

$$= \int_M \langle \nabla \{ D^E f(\eta \circ (\phi_r^v)^{-1}) \}, v \rangle d(\eta \circ (\phi_r^v)^{-1}).$$

Combining this with (4.3) we prove (4.2).

(b) For (2.2). Let $p \in [0, 2]$. For any $\eta \in \mathbb{M}_p$ and $v \in L^2(\mathcal{B}(TM); \eta)$ with $\eta(|v|^2) \leq 1$, by (1.1) we have

$$\sup_{s \in [0, 1]} (\eta \circ \phi_{sv}^{-1})(\rho_o^p) = \eta(\rho_o(\phi_{sv})^p) \leq 2\eta(\rho_o^p + |v|^p) < \infty.$$

Then there exists a constant $K > 0$ such that

$$(4.4) \quad \sup_{s \in [0, 1], \eta(|v|^2) \leq 1} (\eta \circ \phi_{sv}^{-1} + \eta)(\rho_o^p) \leq K.$$

So, by Lemma 3.2, we obtain

$$\begin{aligned} f(\eta \circ \phi_v^{-1}) - f(\eta) &= \int_0^1 \left\{ \frac{d}{dr} f(r\eta \circ \phi_{rv}^{-1} + (1-r)\eta) \right\} dr \\ &= \int_0^1 dr \int_M (D^E f)(r\eta \circ \phi_{rv}^{-1} + (1-r)\eta) d(\eta \circ \phi_v^{-1} - \eta) \\ &= \int_0^1 dr \int_M \left\{ (D^E f)(r\eta \circ \phi_{rv}^{-1} + (1-r)\eta)(\phi_v(x)) - (D^E f)(r\eta \circ \phi_{rv}^{-1} + (1-r)\eta)(x) \right\} \eta(dx) \\ &= \int_0^1 dr \int_M \eta(dx) \int_0^1 \langle //_{\phi_{sv}(x) \rightarrow x} \nabla \{ (D^E f)(r\eta \circ \phi_{rv}^{-1} + (1-r)\eta) \}(\phi_{sv}(x)), v(x) \rangle ds, \end{aligned}$$

where $//_{\phi_{sv}(x) \rightarrow x} : T_{\phi_{sv}(x)}M \rightarrow T_xM$ is the parallel displacement along the geodesic $[0, 1] \ni \theta \mapsto \phi_{(s-\theta)v}(x)$. Thus,

$$\begin{aligned} I_v &:= \frac{|f(\eta \circ \phi_v^{-1}) - f(\eta) - \int_M \langle \nabla \{ D^E f(\eta) \}, v \rangle d\eta|^2}{\eta(|v|^2)} \\ &\leq \int_{[0, 1]^2 \times M} |//_{\phi_{sv}(x) \rightarrow x} \nabla \{ (D^E f)(r\eta \circ \phi_{rv}^{-1} + (1-r)\eta) \}(\phi_{sv}(x)) - \nabla \{ D^E f(\eta) \}(x)|^2 dr ds \eta(dx). \end{aligned}$$

By (4.4), as $\|v\|_{L^2(\eta)} \rightarrow 0$ we have $\phi_{sv}(x) \rightarrow x$ η -a.e. and $\eta \circ \phi_{sv}^{-1} \rightarrow \eta$ in \mathbb{M}_p for any $s \geq 0$. Combining these with (4.4) we may apply the dominated convergence theorem to derive $I_v \rightarrow 0$ as $\|v\|_{L^2(\eta)} \rightarrow 0$. Therefore, f is L -differentiable such that (2.2) holds. \square

Proof of Theorem 2.1(4). It suffices to prove (2.3). Let $f \in C^{L,1}(\mathbb{M})$. We first prove the formula for $\eta \in \mathbb{M}_p$ and $x \in M$ with $\eta(\{x\}) = 0$, then extend to the general situation.

(a) Let $\eta(\{x\}) = 0$. In this case, for any $v_0 \in T_xM$, let $v = 1_{\{x\}}v_0$. Then

$$\phi_{rv}(z) = \begin{cases} z, & \text{if } z \neq x, \\ \exp_x[rv_0], & \text{if } z = x. \end{cases}$$

By $\eta(\{x\}) = 0$, we have

$$(4.5) \quad (\eta + s\delta_x) \circ \phi_{rv}^{-1} = \eta + s\delta_{\exp_x[rv_0]}.$$

Since v can be approximated in $L^2(\eta + s\delta_x)$ by elements in $\Gamma_0(TM)$, the L -differentiability of f and $\eta(\{x\}) = 0$ imply

$$\begin{aligned} & \lim_{r \downarrow 0} \frac{f((\eta + s\delta_x) \circ \phi_{rv}^{-1}) - f(\eta + s\delta_x)}{r} \\ &= \int_M \langle D^L f(\eta + s\delta_x), v \rangle d(\eta + s\delta_x) = s \langle D^L f(\eta + s\delta_x)(x), v_0 \rangle. \end{aligned}$$

Combining this with (4.5), we obtain

$$\lim_{r \downarrow 0} \frac{f(\eta + s\delta_{\exp_x[rv_0]}) - f(\eta + s\delta_x)}{r} = s \langle D^L f(\eta + s\delta_x)(x), v_0 \rangle.$$

This implies that $f(\eta + s\delta_x)$ is differentiable at point x and (2.3) holds.

(b) In general, for any $v_0 \in T_x M$, there exists $r_0 > 0$ such that v_0 extends to a smooth vector field v on $B(x, r_0)$ by parallel displacement; i.e. $v(x)$ is the parallel displacement along the minimal geodesic from x to z . Since $\eta(\{\exp_x[\theta v_0]\}) = 0$ for a.e. $\theta \geq 0$, by the continuity of f and the formula (2.3) for $\eta(\{x\}) = 0$ proved above, we obtain

$$\begin{aligned} & \frac{f(\eta + s\delta_{\exp_x[rv_0]}) - f(\eta + s\delta_x)}{r} = \frac{1}{r} \int_0^r \frac{d}{d\theta} f(\eta + s\delta_{\exp_x[\theta v_0]}) d\theta \\ &= \frac{1}{r} \int_0^r \langle \nabla f(\eta + s\delta_x)(\exp_x[\theta v_0]), v(\exp_x[\theta v_0]) \rangle d\theta \\ &= \frac{s}{r} \int_0^r \langle D^L f(\eta + s\delta_x)(\exp_x[\theta v_0]), v(\exp_x[\theta v_0]) \rangle d\theta, \quad r \in (0, r_0). \end{aligned}$$

By the continuity of $D^L f$, with $r \downarrow 0$ this implies (2.3). \square

Proof of Corollary 2.2. To apply Theorem 2.1, we extend a function f on \mathcal{P}_p to \tilde{f} on \mathbb{M}_p as in the proof of Lemma 3.3, i.e. by letting

$$\tilde{f}(\eta) = h(\eta(M))f(\eta/\eta(M)), \quad \eta \in \mathbb{M}_p,$$

where $h \in C_0^\infty(\mathbb{R})$ with support contained in $[\frac{1}{4}, 2]$ and $h(r) = 1$ for $r \in [\frac{1}{2}, \frac{3}{2}]$. It is easy to see that

$$f((1-s)\mu + s\nu) = \tilde{f}((1-s)\mu + s\nu), \quad s \in [0, 1], \mu, \nu \in \mathcal{P}_p,$$

and $f \in C^{E,1,1}(\mathcal{P}_p)$ implies that $\tilde{f} \in C^{E,1,1}(\mathbb{M}_p)$ and

$$D^E \tilde{f}(\mu) = \tilde{D}^E f(\mu), \quad D^L f(\mu) = D^L \tilde{f}(\mu), \quad D^I f(\mu) = D^{int} \tilde{f}(\mu), \quad \mu \in \mathcal{P}.$$

Then Corollary 2.2(1)-(4) follow from the corresponding assertions in Theorem 2.1 with \tilde{f} replacing f .

Finally, since $\nabla\{\tilde{D}^E f(\mu)\} = \nabla\{D^E \tilde{f}\}(\mu) = D^L f(\mu)$ for $\mu \in \mathcal{P}_p$ and $f \in C^{E,1,1}(\mathcal{P}_p)$, (2.10) follows from Lemma 3.4. \square

References

- [1] Ambrosio L, Gigli N, Savaré G. Gradient Flows in Metric Spaces and in the Spaces of Probability Measures. Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.
- [2] Alberverio S, Kondratiev Y G, Röckner M. Differential geometry of Poisson spaces. C R Acad Sci Paris Sér I Math, 1996, 323: 1129–1134.
- [3] Cardaliaguet P. Notes on mean field games. P.-L. Lions lectures at College de France. Online at <https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf>.
- [4] Carmona R, Delarue F. Probabilistic Theory of Mean Field Games with Applications I. Springer 2019.
- [5] Gangbo W, Tudorascu A. On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. J Math Pures Appl, 2019, 125: 119–174.
- [6] Hammersley W, Šiška D, Szpruch L. McKean-Vlasov SDE under measure dependent Lyapunov conditions. arXiv:1802.03974v1.
- [7] Kondratiev Y, Lytvynov E, Vershik A. Laplace operators on the cone of Radon measures. J Funct Anal, 2015, 269: 2947–2976.
- [8] Overbeck L, Röckner M, Schmuland B. Analytic approach to Fleming-Viot processes with interactive selection. Ann Probab, 1995, 23: 1–36.
- [9] Ren P, Wang F-Y. Bismut formula for Lions derivative of distribution dependent SDEs and applications. J Diff Equat, 2019, 267: 4745–4777.
- [10] Ren R, Wang F-Y. Spectral gap for measure-valued diffusion processes. J Math Anal Appl, 2020, 483: 123624, 16pp.
- [11] von Renesse M-K, Sturm K-T. Entropic measure and Wasserstein diffusion. Ann Probab, 2009, 37: 1114–1191.
- [12] Song Y. Gradient estimates and exponential ergodicity for mean-field SDEs with jumps. J Theor Probab, 2020, 33: 201–238.
- [13] Wang F-Y. Functional inequalities for weighted Gamma distributions on the space of finite (signed) measures. Elect J Probab, to appear; arXiv:1810.10387.
- [14] Wang F-Y. Diffusions and PDEs on Wasserstein spaces. arXiv:1903.02148.
- [15] Wang F Y, Ren P P. Stochastic analysis for measure-valued processes (in Chinese). Sci Sin Math, 2020, 50: 1–22.