An Elementary Proof for the Structure of Wasserstein Derivatives

Cong Wu * and Jianfeng Zhang †

Abstract

Let $F : \mathbb{L}^2(\Omega, \mathbb{R})^1 \to \mathbb{R}$ be a law invariant and continuously Fréchet differentiable mapping. Based on Lions [4], Cardaliaguet [1] (Theorem 6.2 and 6.5) proved that:

$$DF(\xi) = g(\xi), \tag{1}$$

where $g : \mathbb{R} \to \mathbb{R}$ is a deterministic function which depends only on the law of ξ. See also Carmona & Delarue [2] Section 5.2 and Gangbo & Tudorascu [3]. In this short note we provide an elementary proof for this well known result. This note is part of our accompanying paper [5], which deals with a more general situation.

Let $\mathcal{P}_2(\mathbb{R})$ denote the set of square integrable probability measures on $\mathbb{R}$, and consider a mapping $f : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$. As in standard literature, we lift f to a function $F : \mathbb{L}^2(\Omega, \mathbb{R}) \to \mathbb{R}$ by $F(\xi) := f(\mathcal{L}_\xi)$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is an atomless Polish probability space and $\mathcal{L}_\xi$ denotes the law of ξ. If F is Frechét differentiable, then $DF(\xi)$ can be identified as an element of $\mathbb{L}^2(\Omega, \mathbb{R})$:

$$\mathbb{E}\big[DF(\xi)\eta\big] = \lim_{\varepsilon \to 0} \frac{F(\xi + \varepsilon\eta) - F(\xi)}{\varepsilon}, \quad \text{for all } \eta \in \mathbb{L}^2(\Omega, \mathbb{R}). \tag{2}$$

We start with the simple case that ξ is discrete. Let δ_x denote the Dirac measure of x.

Proposition 1 *Assume ξ is discrete: $\mathbb{P}(\xi = x_i) = p_i$, $i \geq 1$. If F is Fréchet differentiable at ξ, then* (1) *holds with*

$$g(x_i) := \lim_{\varepsilon \to 0} \frac{f(\sum_{j \neq i} p_j \delta_{x_j} + p_i \delta_{x_i + \varepsilon}) - f(\sum_{j \geq 1} p_j \delta_{x_j})}{\varepsilon p_i}, \quad i \geq 1. \tag{3}$$

*Department of Mathematics, University of Southern California, Los Angeles, CA 90089. E-mail: congw@usc.edu.

†Department of Mathematics, University of Southern California, Los Angeles, CA 90089. E-mail: jianfenz@usc.edu. This author is supported in part by NSF grant #1413717.

[1]The space $\mathbb{R}$ can be replaced with general $\mathbb{R}^d$. We assume $d = 1$ here for simplicity.

To prove the proposition, we need the following result.

Lemma 2 *Let $X \in \mathbb{L}^2(\Omega, \mathbb{R})$. Assume $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ satisfies*

$$\mathbb{E}[X\mathbf{1}_{A_1}] = \mathbb{E}[X\mathbf{1}_{A_2}], \quad \text{for all } A_1, A_2 \subset A \text{ such that } \mathbb{P}(A_1) = \mathbb{P}(A_2). \tag{4}$$

Then X is a constant, $\mathbb{P}$-a.s. in A.

Proof This result is elementary, we nevertheless provide a proof for completeness.

Assume the result is not true. Denote $c := \frac{\mathbb{E}[X\mathbf{1}_A]}{\mathbb{P}(A)}$ and $A_1 := \{X < c\} \cap A$, $A_2 := \{X > c\} \cap A$. Then $\mathbb{P}(A_1) > 0$, $\mathbb{P}(A_2) > 0$. Assume without loss of generality that $\mathbb{P}(A_1) \le \mathbb{P}(A_2)$. Since $(\Omega, \mathcal{F}, \mathbb{P})$ is atomless, there is a random variable U with uniform distribution on $[0,1]$. Denote $A_{2,x} := A_2 \cap \{U \le x\}$, $x \in [0,1]$. Clearly there exists x_0 such that $\mathbb{P}(A_{2,x_0}) = \mathbb{P}(A_1)$. Apply (4) on A_1 and A_{2,x_0} we obtain the desired contradiction. ■

Remark 3 Lemma 2 may not hold if $(\Omega, \mathcal{F}, \mathbb{P})$ has atoms. Indeed, consider $\Omega := \{\omega_1, \omega_2\}$ with $\mathbb{P}(\omega_1) = \frac{1}{3}, \mathbb{P}(\omega_2) = \frac{2}{3}$. Set $A := \Omega$ and X is an arbitrary random variable. The (4) holds true trivially because $\mathbb{P}(A_1) \neq \mathbb{P}(A_2)$ whenever $A_1 \neq A_2$. However, X may not be a constant. ■

Proof of Proposition 1. Fix an $i \ge 1$. For an arbitrary $A_1 \subset A := \{\xi = x_i\}$, set $\eta := \mathbf{1}_{A_1}$. Note that, for any $\varepsilon > 0$, we have

$$\mathcal{L}_{\xi+\varepsilon\eta} = \sum_{j \neq i} p_j \delta_{x_j} + \mathbb{P}(A_1)\delta_{x_i+\varepsilon} + [p_i - \mathbb{P}(A_1)]\delta_{x_i},$$

which depends only on $\mathcal{L}_\xi$ and $\mathbb{P}(A_1)$. By (2),

$$\mathbb{E}\big[DF(\xi)\mathbf{1}_{A_1}\big] = \lim_{\varepsilon \to 0} \frac{f\big(\sum_{j\neq i} p_j\delta_{x_j} + \mathbb{P}(A_1)\delta_{x_i+\varepsilon} + [p_i - \mathbb{P}(A_1)]\delta_{x_i}\big) - f\big(\sum_{j \ge 1} p_j \delta_{x_j}\big)}{\varepsilon}. \tag{5}$$

In particular, $\mathbb{E}\big[DF(\xi)\mathbf{1}_{A_1}\big]$ depends only on $\mathbb{P}(A_1)$ for $A_1 \subset \{\xi = x_i\}$. Applying Lemma 2, we see that $DF(\xi)$ is a constant, $\mathbb{P}$-a.s. on $\{\xi = x_i\}$. Now set $A_1 := \{\xi = x_i\}$ in (5), we obtain (3) immediately. ■

We now consider the general case.

Theorem 4 *If F is continuously Fréchet differentiable, then* (1) *holds with g depending only on $\mathcal{L}_\xi$ but not on the particular choice of ξ.*

Proof For each $n \geq 1$, denote $x_i^n := i2^{-n}, i \in \mathbb{Z}$, and $\xi_n := \sum_{i=-\infty}^{\infty} x_i^n \mathbf{1}_{\{x_i^n \leq \xi < x_{i+1}^n\}}$. Since ξ_n is discrete, by Proposition 1 we have $DF(\xi_n) = g_n(\xi_n) = \tilde{g}_n(\xi)$, where g_n is defined on $\{x_i^n, i \in \mathbb{Z}\}$ by (3) (with $g_n(x_i^n) := 0$ when $\mathbb{P}(\xi_n = x_i^n) = 0$) and $\tilde{g}_n(x) := g_n(x_i^n)$ for $x \in [x_i^n, x_{i+1}^n)$. Clearly $\lim_{n\to\infty} \mathbb{E}[|\xi_n - \xi|^2] = 0$. Then by the continuous differentiability of F we see that $\lim_{n\to\infty} \mathbb{E}[|\tilde{g}_n(\xi) - DF(\xi)|^2] = 0$. Thus, there exists a subsequence $\{n_k\}_{k\geq 1}$ such that $\tilde{g}_{n_k}(\xi) \to DF(\xi)$, $\mathbb{P}$-a.s. Denote $K := \{x : \overline{\lim}_{k\to\infty} \tilde{g}_{n_k}(x) = \underline{\lim}_{k\to\infty} \tilde{g}_{n_k}(x)\}$, and $g(x) := \lim_{k\to\infty} \tilde{g}_{n_k}(x) \mathbf{1}_K(x)$. Then $\mathbb{P}(\xi \in K) = 1$ and $DF(\xi) = g(\xi)$, $\mathbb{P}$-a.s.

Moreover, let ξ' be another random variable such that $\mathcal{L}_{\xi'} = \mathcal{L}_\xi$. Define ξ_n' similarly. Then $DF(\xi_n') = \tilde{g}_n(\xi')$ for the same function $\tilde{g}_n$. Note that $\mathbb{P}(\xi' \in K) = \mathbb{P}(\xi \in K) = 1$, then $\lim_{k\to\infty} \tilde{g}_{n_k}(\xi') = g(\xi')$, $\mathbb{P}$-a.s. On the other hand, $DF(\xi_{n_k}') \to DF(\xi')$ in $\mathbb{L}^2$. So $DF(\xi') = g(\xi')$, and thus g does not depend on the choice of ξ. ■

Remark 5 One may also write $DF(\xi) = g(\mathcal{L}_\xi, \xi)$, where $g : \mathcal{P}_2(\mathbb{R}) \times \mathbb{R} \to \mathbb{R}$. When DF is uniformly continuous, one may easily construct g jointly measurable in $(\mu, x) \in \mathcal{P}_2(\mathbb{R}) \times \mathbb{R}$. One may also extend the result to the case that F is a function of processes. We leave the details to [5]. ■

References

[1] Cardaliaguet P. (2013), *Notes on Mean Field Games (from P.-L. Lions lectures at College de France)*, preprint, www.ceremade.dauphine.fr/∼cardalia/MFG100629.pdf.

[2] Carmona, R. and Delarue, F. (2017) *Probabilistic Theory of Mean Field Games I Mean Field FBSDEs, Control, and Games*, Springer Verlag, 2017.

[3] Gangbo, W. and Tudorascu, A. *On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations*. Technical report, 2017.

[4] Lions, P.-L. *Cours au Collége de France*. www.college-de-france.fr.

[5] Wu, C. and Zhang, J. *Viscosity Solutions to Parabolic Master Equations and McKean-Vlasov SDEs with Closed-loop Controls*, preprint, arXiv:1805.02639.