Stochastic Processes and their Applications 195 (2026) 104868

Contents lists available at ScienceDirect

Stochastic Processes and their Applications

journal homepage: www.elsevier.com/locate/spa

Sensitivity of functionals of McKean-Vlasov SDEs with respect to the initial distribution

Filippo de Feo [a,b,*], Salvatore Federico [c], Fausto Gozzi [b], Nizar Touzi [d]

[a] Institut für Mathematik, Technische Universität Berlin, Berlin, Germany
[b] Department of Economics and Finance, Luiss Guido Carli University, Rome, Italy
[c] Dipartimento di Matematica, Universit'a "Alma Mater" di Bologna, Italy
[d] New York University, Tandon School of Engineering, New York, USA

ARTICLE INFO

JEL classification:
60H10
91G45

Keywords:
Distributionally robust optimization
Tangent process for McKean-Vlasov SDEs
Model risk

ABSTRACT

We examine the sensitivity at the origin of the distributional robust optimization problem in the context of a model generated by a mean field stochastic differential equation. We adapt the finite dimensional argument developed by Bartl, Drapeau, Obloj, & Wiesel to our framework involving the infinite dimensional gradient of the solution of the mean field SDE with respect to its initial data. We revisit the derivation of this gradient process as previously introduced by Buckdahn, Li, Peng, & Rainer and we complement the existing properties so as to satisfy the requirement of our main result. We use the theory developed in the context of a mean-field systemic risk model by evaluating the sensitivity with respect to the initial distribution for the variance of the log-monetary reserve of a representative bank.

1. Introduction

Distributionally robust optimization (DRO) has been very popular in the recent Operations Research literature. The main idea is to formulate the traditional worst case evaluation of some criterion $g(\mu)$ of a model defined by a probability measure μ outside the restricted framework of parametric models, and to consider instead all possible deviations of the underlying model in the set of probability measures. The discrepancy between such probability measures was evaluated through the Kullback divergence in Lam [1], the total variation distance in Farokhi [2], a criterion based a cumulative distribution functions in Bayraktar and Chen [3], and more recently through the p−Wasserstein distance in Esfahani and Kuhn [4] and Blanchet and Murthy [5], see also Neufeld, En, and Zhang [6], Fuhrmann, Kupper and Nendel [7], Blanchet and Shapiro [8], Nendel and Sgarabottolo [9], Blanchet, Chen and Zhou [10], Blanchet, Kang and Murthy [11], Blanchet, Li, Lin and Zhang [12], Lanzetti, Bolognani and Dörfler [13], Yue, Kuhn and Wiesemann [14].

By restricting to models in the p−Wasserstein ball with radius δ centered at the model of interest, the DRO value function reduces to a scalar function of δ defined as the worst evaluation of the criterion g in this neighborhood. Our interest here is on the sensitivity at the origin of the DRO as derived in Bartl, Drapeau, Obloj, and Wiesel [15], see also Bartl and Wiesel [16] for similar results in the context of the adapted Wasserstein ball, Jiang and Obloj [17] for the continuous time extension of [16], and Sauldubois and Touzi [18] for the martingale Wasserstein ball with possibly restricted marginals.

* Corresponding author.
E-mail addresses: defeo@math.tu-berlin.de (F. de Feo), s.federico@unibo.it (S. Federico), fgozzi@luiss.it (F. Gozzi), nizar.touzi@nyu.edu (N. Touzi).

https://doi.org/10.1016/j.spa.2025.104868
Received 10 April 2025; Accepted 31 December 2025
Available online 3 January 2026

In this paper, we consider the case where the criterion $g(\mu) := \phi \circ (T_{\#}\mu)$ where T is a transport map of the initial law μ through the solution of a mean field stochastic differential equation. This setting is motivated by the huge interest in mean field models resulting from interacting population of agents in several application areas. The seminal papers of Lasry and Lions [19] and Huang, Malhamé, and Caines [20] introducing mean field games initiated a large interest interacting populations control and games problems and their mean field limit, see also Carmona and Delarue [21]. We are particularly interested in applications to default contagion and systemic risk modeling in financial mathematics, e.g., see Fouque, Carmona and Sun [22], Nadtotchiy and Shkolnikov [23], Djete and Touzi [24], and De Crescenzo, de Feo, and Pham [25].

The current setting in this paper with criterion involving the law of a mean field SDE is in the general spirit of the static setting considered by Bartl, Drapeau, Obloj, and Wiesel [15]. However, the criterion considered in this paper $g(\mu) := \phi \circ (T_{\#}\mu)$ does not fall in the setting considered in there, and involves a nontrivial nonlinearity due to the mean field dependence. In order to adapt their derivation approach of the DRO sensitivity at the origin and to characterize it in our dynamic setting, we essentially analyze the regularity of this transport map. This requires a careful study of the gradient process of the solution of the mean field stochastic differential equation with respect to its initial condition which was introduced previously by Buckdahn, Li, Peng, and Rainer [26] (see also Chassagneux, Crisan, and Delarue [27]). We revisit the approach of [26] and complement their result as follows. Under appropriate conditions, this gradient process is the Gateaux derivative of the solution of the mean field SDE, it is a uniformly strongly continuous functional of its initial condition, and it is bounded in the operator norm. The corresponding adjoint process inherits the last properties and is the main ingredient for the expression of the DRO sensitivity at the origin. We use the theory developed in the context of a mean-field systemic risk model [22, Section 2] by evaluating the sensitivity with respect to the initial distribution for the variance of the log-monetary reserve of a representative bank.

The paper is organized as follows. Section 2 introduces the general framework. Section 3 contains the main results together with the justification of the distributionally robust sensitivity. All arguments related to the differentiability of the solution of the McKean-Vlasov SDE with respect to its initial law are reported in Section 4. Finally, we provide in Section 5 an application in the context of the systemic risk model of Carmona, Fouque and Sun [22].

2. The underlying framework

2.1. Basic notations

We denote by $\mathbb{R}^{d\times h}$ the space of $d \times h-$matrices with entries in $\mathbb{R}$ and by $|\cdot|$ the Frobenius norm on this space. The transpose of a matrix is denoted by the superscript ^{T}.

Given a Polish space E, we denote by $\mathcal{B}(E)$ the Borel $\sigma-$algebra on E and by $\mathcal{P}(E)$ the set of probability measures on $(E, \mathcal{B}(E))$. The Dirac's delta measure concentrated at $x \in E$ is denoted by δ_x. We denote the set of couplings in $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with given marginals $\mu, \mu' \in \mathcal{P}(\mathbb{R}^d)$ by $\Pi(\mu,\mu') := \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \ \pi(\cdot \times \mathbb{R}^d) = \mu, \ \pi(\mathbb{R}^d \times \cdot) = \mu'\}$.

Given a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and a random variable $X : \Omega \to \mathbb{R}^{d\times h}$, we denote by P_X the law of X. For $p > 0$ we denote by $L^p(\mathcal{F};\mathbb{R}^{d\times h})$ the Banach space of $p-$integrable random variables X with norm $\|X\|_{L^p} := (\mathsf{E}|X|^p)^{1/p}$. We denote $L^p(\mathcal{F}) := L^p(\mathcal{F};\mathbb{R})$. When $p = 2$ we denote by $\langle\cdot,\cdot\rangle_{L^2}$ the corresponding inner product on this space. A probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is said to be atomless if for every $A \in \mathcal{F}$ such that $\mathsf{P}(A) > 0$, there exists $B \subsetneq A$ such that $\mathsf{P}(B) > 0$. It is well known [21] that, in this case, for every $\mu \in \mathcal{P}(\mathbb{R}^{d\times h})$ there exists a random variable $X : \Omega \to \mathbb{R}^{d\times h}$ random variable such that $\mathsf{P}_X = \mu$.

2.2. The wasserstein space

We define $\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d) : \|\mu\|_2 < \infty\}$, where $\|\mu\|_2 := \left(\int_{\mathbb{R}^d} |x|^2 \mu(dx)\right)^{1/2}$. We endow $\mathcal{P}_2(\mathbb{R}^d)$ with the 2-Wasserstein distance

$$\mathcal{W}_2(\mu,\mu') := \inf_{\pi\in\Pi(\mu,\mu')} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \pi(dx,dy)\right)^{\frac{1}{2}}, \quad \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d).$$

The space $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$ is a Polish space. It is well-known [21] that an optimal coupling always exists, i.e. there exists $\pi^* \in \Pi(\mu,\mu')$ such that $\mathcal{W}_2(\mu,\mu') = \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \pi^*(dx,dy)$. We recall [21] that the following equality holds:

$$\|\mu\|_2 = \mathcal{W}_2(\mu,\delta_0), \quad \mu \in \mathcal{P}_2(\mathbb{R}^d). \tag{1}$$

We denote $B^2_\delta(\mu) := \{\mu' \in \mathcal{P}_2(\mathbb{R}^d) : \ \mathcal{W}_2(\mu',\mu) \le \delta\}$.

If $(\Omega, \mathcal{F}, \mathsf{P})$ is atomless, we have $\mathcal{P}_2(\mathbb{R}^d) = \{\mathsf{P}_X : X \in L^2(\mathcal{F};\mathbb{R}^d)\}$, that is, given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we may construct $X \in L^2(\mathcal{F};\mathbb{R}^d)$ such that $\mathsf{P}_X = \mu$. Hence, given $\mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d)$, we have $\Pi(\mu,\mu') \subset \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ and $\Pi(\mu,\mu') = \left\{\mathsf{P}_{(X,X')} : \ X, X' \in L^2(\mathcal{F};\mathbb{R}^d), \ \mathbb{P}_X = \mu, \mathbb{P}_{X'} = \mu'\right\}$. Thus, for every $\mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d)$, we have $\mathcal{W}_2(\mu,\mu') := \inf\{\|X - X'\|_{L^2} : \ X, X' \in L^2(\mathcal{F};\mathbb{R}^d), \ X \sim \mu, \ X' \sim \mu'\}$. Finally, there exist $X, X' \in L^2(\mathcal{F};\mathbb{R}^d)$ such that $\mathsf{P}_{(X,X')} = \pi^*$ and $\mathcal{W}_2(\mu,\mu') = \|X - X'\|_{L^2}$.

2.3. Derivatives in the wasserstein space

We recall here the definition of linear functional derivative and of $L-$derivative [21].

Definition 1. (i) A map $\phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is said to have a *linear functional derivative* if there exists a map $\delta_\mu \phi : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ continuous for the product topology, such that the function $x \mapsto \delta_\mu \phi(\mu, x)$ has at most quadratic growth in x locally uniformly in μ and

$$\phi(\mu') - \phi(\mu) = \int_0^1 \int_{\mathbb{R}^d} \delta_\mu \phi(\mu^\lambda, x)\big(\mu' - \mu\big)(dx) d\lambda, \quad \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d),$$

where $\mu^\lambda := \lambda(\mu' - \mu) + \mu$. We denote by $C^1_{LF}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ the class of such maps.

(ii) A map $\phi : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is said to be *L-differentiable* if: it belongs to $C^1_{LF}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$; for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ the real-valued map $x \mapsto \delta_\mu \phi(\mu, x)$ is differentiable; the map $(\mu, x) \mapsto \partial_x \delta_\mu \phi(\mu, x) \in \mathbb{R}^d$ is continuous and has at most of quadratic growth in x, locally uniformly in μ. In this case, the map $\partial_x \delta_\mu \phi$ is called the *L-derivative* of ϕ. We denote the set of such maps by $C^1(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$. Analogously, we define the space $C^1(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d\times m})$.

We recall that the $L-$derivative $\partial_x \delta_\mu \phi$ coincides with the Lions derivative [21] and the Wasserstein gradient [28]. We extend the previous definition as follows.

Definition 2.

1. $C^1(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d\times m})$ is the space of maps $\phi : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \longrightarrow \mathbb{R}^{d\times m}$ such that
   - $\phi(\cdot, \mu) \in C^1(\mathbb{R}^d; \mathbb{R}^{d\times m})$ for all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with partial Jacobian denoted by $\partial_x \phi(x, \mu)$,
   - $\phi(x, \cdot) \in C^1(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d\times m})$ for all $x \in \mathbb{R}^d$, with partial $L-$derivative denoted by $\partial_{\tilde{x}} \delta_\mu \phi(x, \mu, \tilde{x})$.
2. We denote by
   - $C^1_b(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d\times m})$ the subspace of functions ϕ with bounded $\partial_x \phi$ and $\partial_{\tilde{x}} \delta_\mu \phi$.
   - $C^{1,1}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d\times m})$ (resp., $C^{1,1}_b(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d\times m})$) the corresponding subspace of functions ϕ with Lipschitz (resp., Lipschitz and bounded) partial gradients $\partial_x \phi$ and $\partial_{\tilde{x}} \delta_\mu \phi$.

2.4. McKean-Vlasov stochastic differential equation

Let $T > 0$ and let $\{B_t\}_{t\in[0,T]}$ be an m-dimensional Brownian motion defined on a complete atomless probability space $(\Omega, \mathcal{F}, \mathsf{P})$. We denote by $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ the completion of the filtration generated by B, which is also right-continuous, so that it satisfies the usual conditions. We denote by $\mathcal{S}^2([0,T]; \mathbb{R}^{d\times h})$ the space of continuous $\mathbb{F}-$adapted processes Y with values in $\mathbb{R}^{d\times h}$ such that $\|Y\|^2_{\mathcal{S}^2} := \mathsf{E}\big[\sup_{t\in[0,T]} |Y_t|^2\big] < \infty$.

Given $\xi \in L^2(\mathcal{F}_0; \mathbb{R}^d)$ and $b = (b^0, b^1) : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d \times \mathbb{R}^{d\times m}$, we consider a SDE of McKean-Vlasov (MKV) type:

$$X^\xi_t = \xi + \int_0^t b(X^\xi_s, \mathsf{P}_{X^\xi_s}) dZ_s, \quad t \in [0,T], \text{ where } Z_t = \begin{bmatrix} Z^0_t \\ Z^1_t \end{bmatrix} := \begin{bmatrix} t \\ B_t \end{bmatrix}. \tag{2}$$

Under the standard Lipschitz condition

$$\Big|b(x, \mu) - b(x', \mu')\Big| \le C\big(|x - x'| + \mathcal{W}_2(\mu, \mu')\big), \ (x, \mu), (x', \mu') \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d), \tag{3}$$

for some constant $C > 0$, there exists a unique solution X^ξ to (2) in the class of processes $\mathcal{S}^2([0,T]; \mathbb{R}^d)$, see e.g. [21, Th. 4.21 and Lemma 4.34]. Moreover, there exists $C > 0$ such that

$$\|X^\xi\|_{\mathcal{S}^2} \le C\|\xi\|_{L^2} \text{ and } \|X^\xi - X^{\xi'}\|_{\mathcal{S}^2} \le C\|\xi - \xi'\|_{L^2}, \text{ for all } \xi, \xi' \in L^2(\mathcal{F}_0; \mathbb{R}^d). \tag{4}$$

By the uniqueness in law of the solution to the MKV SDE, it follows that the law of the process X^ξ is independent of the choice of the initial r.v. ξ in the set

$$\mathcal{R}_{\mu_0} := \big\{\xi \in L^2(\mathcal{F}_0; \mathbb{R}^d) : \ \mathsf{P}_\xi = \mu_0\big\}. \tag{5}$$

In the following, we shall often confuse μ_0 with the reference initial conditions $\xi \in \mathcal{R}_{\mu_0}$, and abuse notation writing $\mathsf{P}_{X^{\mu_0}_t}$ instead of $\mathsf{P}_{X^\xi_t}$ for an arbitrary $\xi \in \mathcal{R}_{\mu_0}$.

3. Sensitivity of functionals for McKean-Vlasov SDEs

Let $\phi : \mathcal{P}_2(\mathbb{R}^d) \longrightarrow \mathbb{R}$, our goal is to analyze the differentiability at the origin $r = 0$ of the map

$$\Phi(\mu_0, r) := \sup_{\mu'_0 \in B^2_r(\mu_0)} \phi\big(\mu'_T\big), \qquad \mu_0 \in \mathcal{P}_2(\mathbb{R}^d), \ r \ge 0,$$

where $\mu'_t := \mathsf{P}_{X^{\mu'_0}_t} = \mathsf{P}_{X^{\xi'}_t}$ for $t \in [0,T]$, and $X^{\xi'}$ is the solution to (2) with an arbitrary initial condition $\xi' \in \mathcal{R}_{\mu'_0}$. When $\mu'_0 = \mu_0$, we denote the corresponding reference starting condition by ξ and set $\mu_t := \mathsf{P}_{X^\xi_t}$ for $t \in [0,T]$.

Assumption 1. The functional ϕ satisfies the following.

(i) There exists $C > 0$ such that $|\phi(\mu)| \le C(1 + \|\mu\|^2_2)$, for all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

(ii) $\phi \in C^1(\mathcal{P}_2(\mathbb{R}^d))$, there exists a modulus of continuity ϖ such that

$$|\partial_x\delta_\mu\phi(\mu',x) - \partial_x\delta_\mu\phi(\mu,x)| \le \varpi(\mathcal{W}_2(\mu',\mu)), \qquad \mu,\mu' \in \mathcal{P}_2(\mathbb{R}^d), x \in \mathbb{R}^d,$$

and for all $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ a modulus of continuity ϖ_μ such that

$$|\partial_x\delta_\mu\phi(\mu,x') - \partial_x\delta_\mu\phi(\mu,x)| \le \varpi_\mu(|x'-x|), \qquad x,x' \in \mathbb{R}^d.$$

We provide some examples of functionals satisfying the previous assumptiontion and which justify that the variance criterion considered in the example of Section 5 is included in our setting.

Example 1. Let $f \in C^1(\mathbb{R}^d)$ be a function with sub-quadratic growth and uniformly continuous gradient ∇f. Then $\phi(\mu) := \int_{\mathbb{R}^d} f(x)\mu(dx)$ satisfies Assumptiontion 1. Indeed, (i) follows from the subquadratic growth of f; (ii) follows from the fact that $\partial_x\delta_\mu\phi(\mu,x) = \nabla f(x)$.

Example 2. Let $\psi \in C^1(\mathbb{R})$ have uniformly continuous derivative ψ' and let $f \in C_b^1(\mathbb{R}^d)$ have uniformly continuous gradient ∇f. Then $\phi(\mu) := \psi\left(\int_{\mathbb{R}^d} f(x)\mu(dx)\right)$ satisfies Assumption 1. Indeed, (i) follows from sublinear growth of ψ, f. As for (ii), we have $\partial_x\delta_\mu\phi(\mu,x) = \psi'\left(\int_{\mathbb{R}^d} f(y)\mu(dy)\right)\nabla f(x)$, so that for some modulus ϖ (which may change from line to line), denoting by $\mathcal{W}_1$ the 1-Wasserstein distance, we have

$$\begin{aligned}|\partial_x\delta_\mu\phi(\mu',x) - \partial_x\delta_\mu\phi(\mu,x)| &\le \left|\psi'\left(\int f\,d\mu'\right) - \psi'\left(\int f\,d\mu\right)\right||\nabla f(x)| \\ &\le \varpi\left(\left|\int f\,d\mu' - \int f\,d\mu\right|\right) \le \varpi\left(\mathcal{W}_1(\mu',\mu)\right) \le \varpi\left(\mathcal{W}_2(\mu',\mu)\right),\end{aligned}$$

where in the third inequality we have used the Kantorovich-Rubinstein duality [21, Corollary 5.4] up to rescaling by the Lipschitz constant of f. Finally, we have

$$|\partial_x\delta_\mu\phi(\mu,x') - \partial_x\delta_\mu\phi(\mu,x)| \le \left|\psi'\left(\int f\,d\mu\right)\right||\nabla f(x') - \nabla f(x)| \le \left|\psi'\left(\int f\,d\mu\right)\right|\varpi(|x'-x|).$$

For later use, we start by providing some estimates.

Lemma 1. *Let Condition (3) and Assumption 1 (i) hold. Then:*

(i) There exists $C > 0$ such that $|\Phi(\mu_0, r)| \le C(1 + r^2 + \|\mu_0\|_2^2)$, for every $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $r > 0$.
(ii) Under the additional Assumption 1 (ii), we have:

$$K_R := \sup_{t\in[0,T]} \sup_{\mu_0 \in B_R^2(\mu_0)} \left\|\partial_x\delta_\mu\phi(\mu_t, X_t^\xi)\right\|_{L^2} < \infty \text{ for all } R > 0.$$

Proof. (i) For $r > 0$, let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu_0' \in B_r^2(\mu_0)$. By Assumption 1 we have for some constant C that may vary from line to line in the inequalities, $\left|\phi(\mu_T')\right| \le C\left(1 + \left\|\mu_T'\right\|_2^2\right) = C\left(1 + \left\|X_T^{\xi'}\right\|_{L^2}^2\right)$. Then, it follows from (4) together with (1) and the triangle inequality for $\mathcal{W}_2(\cdot,\cdot)$ that:

$$\left|\phi(\mu_T')\right| \le C\left(1 + \left\|\xi'\right\|_{L^2}^2\right) = C\left(1 + \left\|\mu_0'\right\|_2^2\right) \le C\left(1 + \mathcal{W}_2(\mu_0',\mu_0)^2 + \left\|\mu_0\right\|_2^2\right) \le C(1 + r^2 + \|\mu_0\|_2^2),$$

which induces the required claim.

(ii) For arbitrary $t \in [0,T]$ and $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\|\mu_0\|_2 \le R$, by (4) we have $\|\mu_t\|_2^2 = \|X_t^\xi\|_{L^2}^2 \le C(1 + \|\xi\|_{L^2}^2) = C(1 + \|\mu_0\|_2^2) \le C(1 + R^2)$ for some $C > 0$. We can then use (4) and the growth bound for $|\partial_x\delta_\mu\phi|$ in Definition 1 to get, for some $C_R > 0$,

$$\left\|\partial_x\delta_\mu\phi(\mu_t, X_t^\xi)\right\|_{L^2}^2 \le C_R(1 + \|X_t^\xi\|_{L^2}^2) \le C_R(1 + \|\xi\|_{L^2}^2) = C_R(1 + \|\mu_0\|_2^2) \le C_R(1 + R^2).$$

To go towards our main result concerning the study of the sensitivity of the functional Φ with respect to μ, we first investigate the dependence of the solution of the MKV equation with respect to the initial datum. Questions of this type have been addressed in [27] and in [26] (see also [29] for extensions to McKean Vlasov SDEs with jumps and [30] for mean-field backward doubly stochastic differential equations). We notice that these papers address the smoothness of the map $\xi \mapsto X_t^\xi$ with different approaches. While [27] show the Gateaux differentiability by analyzing the dependence of the Picard iterations on the initial law, we shall follow the approach of [26] who introduced an auxiliary process with initial law concentrated at a point thus reducing the initial value of the process to a deterministic object.

Given two Banach spaces $(E, \|\cdot\|_E), (F, \|\cdot\|_F)$, we denote by $\mathcal{L}(E,F)$ the Banach space of linear bounded operators L from E to F, endowed with the operator norm $|L|_{\mathcal{L}} := \sup_{\|x\|_E=1} \|Lx\|_F$. The notion of derivative of the map $\xi \mapsto X_t^\xi$ is provided by Theorem 1 below. Such a derivative is seen as an operator $D_\xi X_t^\xi \in \mathcal{L}\left(L^2(\mathcal{F}_0;\mathbb{R}^d); L^2(\mathcal{F}_t;\mathbb{R}^d)\right)$. Relying on the results of [26], we are able to prove Gateaux differentiability of this map and some estimates concerning $D_\xi X_t^\xi$ and its adjoint $(D_\xi X_t^\xi)^*$. In Subsection 4.2, we will also derive their explicit expressions (28) and (36). We refer to Remark 3 for a discussion. Our analysis requires to strengthen condition (3) as follows.

Assumption 2. $b \in C_b^{1,1}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d \times \mathbb{R}^{d\times m})$.

Example 3. Assume $d = m = 1$ for simplicity. Let $\hat{b} \in C_b^{1,1}(\mathbb{R}\times\mathbb{R};\mathbb{R}\times\mathbb{R})$ and $g \in C_b^{1,1}(\mathbb{R})$. Then $b(x,\mu) := \hat{b}\big(x, \int_{\mathbb{R}^d} g(y)\mu(dy)\big)$ satisfies Assumption 2. Indeed, denoting by $\partial_x\hat{b}, \partial_y\hat{b}$, respectively, the partial derivatives with respect to the first and second variable of $\hat{b}$, we have

$$\partial_x b(x,\mu) = \partial_x\hat{b}\left(x, \int_{\mathbb{R}^d} g(z)\mu(dz)\right),\quad \partial_{\tilde{x}}\delta_\mu b(x,\mu,\tilde{x}) = \partial_y\hat{b}\left(x, \int_{\mathbb{R}^d} g(z)\mu(dz)\right)g'(\tilde{x}).$$

Proceeding as in Example 2, we have the claim.

abc

Theorem 1. *Let Assumption 2 hold.*
(i) For all $t \in [0,T]$, the map $\xi \in L^2(\mathcal{F}_0;\mathbb{R}^d) \mapsto X_t^\xi \in L^2(\mathcal{F}_t;\mathbb{R}^d)$ is Gateaux differentiable with bounded differential; i.e., for all $\xi \in L^2(\mathcal{F}_0;\mathbb{R}^d)$, there exists $D_\xi X_t^\xi \in \mathcal{L}\big(L^2(\mathcal{F}_0;\mathbb{R}^d); L^2(\mathcal{F}_t;\mathbb{R}^d)\big)$ such that

$$\sup_{\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d)}\sup_{t\in[0,T]}\left\|D_\xi X_t^\xi\right\|_{\mathcal{L}} < \infty \quad\textit{and}\quad \lim_{r\to 0}\frac{1}{|r|}\left\|X_t^{\xi+r\eta} - X_t^\xi - r\,D_\xi X_t^\xi\eta\right\|_{L^2} = 0,\quad \eta\in L^2(\mathcal{F}_0;\mathbb{R}^d).$$

Moreover, $\xi \mapsto D_\xi X_t^\xi$ is uniformly strongly continuous; that is, for every $\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$ there exists ϖ_η modulus of continuity such that

$$\left\|\left(D_\xi X_t^{\xi'} - D_\xi X_t^\xi\right)\eta\right\|_{L^2} \le \varpi_\eta(\|\xi'-\xi\|_{L^2}),\quad \xi,\xi'\in L^2(\mathcal{F}_0;\mathbb{R}^d). \tag{6}$$

(ii) The adjoint $\left(D_\xi X_t^\xi\right)^* \in \mathcal{L}(L^2(\mathcal{F}_t;\mathbb{R}^d); L^2(\mathcal{F}_0;\mathbb{R}^d))$ inherits the boundedness and the uniform strong continuity, i.e. $\left\|\left(D_\xi X_t^\xi\right)^*\right\|_{\mathcal{L}}$ bounded uniformly in (t,ξ) and for all $t\in[0,T]$:

$$\left\|\left[\left(D_\xi X_t^{\xi'}\right)^* - \left(D_\xi X_t^\xi\right)^*\right]\eta\right\|_{L^2} \le \varpi_\eta(\|\xi'-\xi\|_{L^2}),\ \xi,\xi'\in L^2(\mathcal{F}_0;\mathbb{R}^d),\ \eta\in L^2(\mathcal{F}_t;\mathbb{R}^d).$$

for some modulus of continuity ϖ_η.

Proof. See Subsection 4.2.
Our main result is the following.

Theorem 2. *Let Assumptions 1 and 2 hold. Then, for each $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, the function $r \longmapsto \Phi(\mu_0, r)$ is differentiable at 0 and*

$$\frac{\partial\Phi}{\partial r}(\mu_0,0) = \left\|\left(D_\xi X_T^\xi\right)^*\partial_x\delta_\mu\phi(\mu_T, X_T^\xi)\right\|_{L^2}.$$

Proof. Denote $\Delta_r := \frac{\Phi(\mu_0,r)-\Phi(\mu_0,0)}{r}$, and $\zeta := \left(D_\xi X_T^\xi\right)^*\partial_x\delta_\mu\phi(\mu_T,X_T^\xi)$ for an arbitrary $\xi\in\mathcal{R}_{\mu_0}$. We organize the proof in two steps.
Step 1. We first prove that $\limsup_{r\searrow 0}\Delta_r \le \|\zeta\|_{L^2}$. By definition of linear functional derivative and L−derivative, we may write

$$\begin{aligned} r\Delta_r = \sup_{\mu_0'\in B_r^2(\mu_0)}\left\{\phi(\mu_T') - \phi(\mu_T)\right\} &= \sup_{\mu_0'\in B_r^2(\mu_0)}\int_0^1\int_{\mathbb{R}^d}\delta_\mu\phi(\mu_T^{\lambda_1},x)(\mu_T'-\mu_T)(dx)d\lambda_1 \\ &= \sup_{\mu_0'\in B_r^2(\mu_0)}\int_0^1 \mathsf{E}\left[\delta_\mu\phi(\mu_T^{\lambda_1},X_T^{\xi'}) - \delta_\mu\phi(\mu_T^{\lambda_1},X_T^\xi)\right]d\lambda_1 \\ &= \sup_{\mu_0'\in B_r^2(\mu_0)}\int_0^1\int_0^1 \mathsf{E}\left[\left\langle\partial_x\delta_\mu\phi(\mu_T^{\lambda_1},X_T^{\lambda_2}), X_T^{\xi'} - X_T^\xi\right\rangle\right]d\lambda_2 d\lambda_1, \end{aligned}$$

where $\mu_t^{\lambda_1} := \lambda_1(\mu_t' - \mu_t) + \mu_t$, $X_t^{\lambda_2} := \lambda_2(X_t^{\xi'} - X_t^\xi) + X_t^\xi$, $\lambda_1,\lambda_2\in[0,1], t\in[0,T]$. By Theorem 1, we may apply [31, Theorem 4.A, p. 148] to obtain the equality $X_T^{\xi'} - X_T^\xi = \int_0^1\left[D_\xi X_T^{\xi^{\lambda_3}}\right](\xi'-\xi)d\lambda_3$ with $\xi^{\lambda_3} := \lambda_3(\xi'-\xi)+\xi$. Then, setting $\lambda = (\lambda_1,\lambda_2,\lambda_3)$,

$$\begin{aligned} r\Delta_r &= \sup_{\mu_0'\in B_r^2(\mu_0)}\int_{[0,1]^3}\mathsf{E}\left[\left\langle\partial_x\delta_\mu\phi(\mu_T^{\lambda_1},X_T^{\lambda_2}), D_\xi X_T^{\xi^{\lambda_3}}(\xi'-\xi)\right\rangle\right]d\lambda \\ &= \sup_{\mu_0'\in B_r^2(\mu_0)}\int_{[0,1]^3}\mathsf{E}\left[\left\langle\zeta^\lambda,\xi'-\xi\right\rangle\right]d\lambda,\quad\text{with}\quad \zeta^\lambda := \left(D_\xi X_T^{\xi^{\lambda_3}}\right)^*\partial_x\delta_\mu\phi(\mu_T^{\lambda_1},X_T^{\lambda_2}). \end{aligned}\tag{7}$$

By the Hölder inequality, and choosing $\xi',\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d)$ such that $\mathsf{P}_{(\xi',\xi)}\in\Pi(\mu_0,\mu_0')$ is an optimal coupling for $\mu_0' = \mathsf{P}_{\xi'}$, $\mu_0 = \mathsf{P}_\xi$, this implies that

$$\begin{aligned} \Delta_r \le \frac{1}{r}\sup_{\mu_0'\in B_r^2(\mu_0)}\int_{[0,1]^3}\left\|\zeta^\lambda\right\|_{L^2}\left\|\xi'-\xi\right\|_{L^2}d\lambda &= \frac{1}{r}\sup_{\mu_0'\in B_r^2(\mu_0)}\int_{[0,1]^3}\left\|\zeta^\lambda\right\|_{L^2}\mathcal{W}_2(\mu_0',\mu_0)\,d\lambda \\ &\le \sup_{\mu_0'\in B_r^2(\mu_0)}\int_{[0,1]^3}\left\|\zeta^\lambda\right\|_{L^2}\,d\lambda. \end{aligned}\tag{8}$$

We next estimate for all $\mu_0' \in B_r^2(\mu_0)$ the difference

$$E := \left\|\zeta^\lambda - \zeta\right\|_{L^2} \le E_1 + E_2 \text{ where } E_1 := \left\|\left(D_\xi X_T^{\xi^{\lambda_3}}\right)^* \left[\partial_x\delta_\mu\phi(\mu_T^{\lambda_1}, X_T^{\lambda_2}) - \partial_x\delta_\mu\phi(\mu_T, X_T^\xi)\right]\right\|_{L^2}$$
$$E_2 := \left\|\left[\left(D_\xi X_T^{\xi^{\lambda_3}}\right)^* - \left(D_\xi X_T^\xi\right)^*\right]\partial_x\delta_\mu\phi(\mu_T, X_T^\xi)\right\|_{L^2}.$$

Using Theorem 1 (iii), Assumption 1(iii) (taking without loss of generality a concave modulus of continuity ϖ_{μ_T} therein), (4), and Jensen's inequality, we have

$$\begin{aligned}
E_1 &\le C\left[\left\|\partial_x\delta_\mu\phi(\mu_T^{\lambda_1}, X_T^{\lambda_2}) - \partial_x\delta_\mu\phi(\mu_T, X_T^{\lambda_2})\right\|_{L^2} + \left\|\partial_x\delta_\mu\phi(\mu_T, X_T^{\lambda_2}) - \partial_x\delta_\mu\phi(\mu_T, X_T^\xi)\right\|_{L^2}\right]\\
&\le C\left[\varpi\left(\mathcal{W}_2(\mu_T^{\lambda_1}, \mu_T)\right) + \left\|\varpi_{\mu_T}\left(|X_T^{\lambda_2} - X_T^\xi|\right)\right\|_{L^2}\right]\\
&\le C\left[\varpi\left(\left\|X_T^{\lambda_1} - X_T^\xi\right\|_{L^2}\right) + \varpi_{\mu_T}\left(\left\|X_T^{\lambda_2} - X_T^\xi\right\|_{L^2}\right)\right] \le \varpi\left(\|\xi' - \xi\|_{L^2}\right),
\end{aligned}$$

where the moduli of continuity changed from line to line and we dropped the dependence of ϖ from μ_T since the letter is a fixed law (for fixed μ_0). By similar arguments, we may also estimate E_2 noting that, by Theorem 1 (iii), we have $E_2 \le \varpi_\xi(\|\xi^{\lambda_3} - \xi\|_{L^2})$, and we may then conclude that there exists a modulus of continuity $\hat\varpi$ such that

$$E \le \hat\varpi\left(\|\xi' - \xi\|_{L^2}\right) = \hat\varpi(\mathcal{W}_2(\mu_0', \mu_0)) \le \hat\varpi(r), \quad \mu_0' \in B_r^2(\mu_0), \tag{9}$$

where the last equality holds because $\mathsf{P}_{(\xi',\xi)} =: \pi^* \in \Pi(\mu_0, \mu_0')$ is an optimal coupling for $\mu_0' = \mathsf{P}_{\xi'}$ and $\mu_0 = \mathsf{P}_\xi$. Plugging the last estimate in (8), we obtain $\Delta_r \le \hat\varpi(r) + \|\zeta\|_{L^2}$, which induces the required result by taking the $\limsup_{r\to 0}$.

Step 2. We next show that $\|\zeta\|_{L^2} \le \liminf_{r\searrow 0}\Delta_r$ to complete the proof of the theorem. Fix an arbitrary $\xi \in \mathcal{R}_{\mu_0}$, set $\hat\xi := \xi + r\frac{\zeta}{\|\zeta\|_{L^2}}$ and $\hat\mu_0 := \mathsf{P}_{\hat\xi}$. As $\zeta \in L^2(\mathcal{F}_0; \mathbb{R}^d)$, by Lemma 1 and Theorem 1 (iii), we have $\hat\xi \in L^2(\mathcal{F}_0; \mathbb{R}^d)$. Moreover $\mathcal{W}_2(\hat\mu_0, \mu_0) \le \|\hat\xi - \xi\|_{L^2} = r$, so that $\hat\mu_0 \in B_r^2(\mu_0)$. Hence, by (7) and, using the same notations as in Step 1 with $\mu_0' := \hat\mu_0$, we may write

$$\begin{aligned}
\Delta_r &\ge \int_{[0,1]^3}\frac{\mathsf{E}\left[\left\langle\zeta^\lambda, \hat\xi - \xi\right\rangle\right]}{r\|\zeta\|_{L^2}}d\lambda = \int_{[0,1]^3}\frac{\mathsf{E}\left[\left\langle\zeta^\lambda, \zeta\right\rangle\right]}{\|\zeta\|_{L^2}}d\lambda = \|\zeta\|_{L^2} + \int_{[0,1]^3}\frac{\mathsf{E}\left[\left\langle\zeta^\lambda - \zeta, \zeta\right\rangle\right]}{\|\zeta\|_{L^2}}d\lambda\\
&\ge \|\zeta\|_{L^2} - \varpi(\|\hat\xi - \xi\|_{L^2}) = \|\zeta\|_{L^2} - \varpi(r),
\end{aligned}$$

where the last inequality follows from (9). Taking the $\liminf_{r\to 0}$ completes the proof.

4. Differentiability of solutions of McKean-Vlasov SDEs with respect to the initial datum

This section is dedicated to the proof of Theorem 1 by essentially revisiting the arguments of [26], pointing out some relevant features and completing them appropriately for our needs.

4.1. *The auxiliary process* $X_t^{x_0,\mathsf{P}_\xi}$

We recall that the idea of [26] to study the differentiability of solutions of McKean-Vlasov SDEs with respect to the initial datum is to disentangle in the McKean-Vlasov SDE the dependence on the initial datum ξ and the dependence of the coefficients on the law $\mathsf{P}_{X_s}^\xi$. So, one starts by freezing the dependence of the coefficients on the flow of measures $(\mathsf{P}_{X_s^\xi})_{s\in[0,T]}$ and introduce the auxiliary (standard) SDE in X_s

$$X_t = x_0 + \int_0^t \hat b\left(s, X_s\right)dZ_s, \quad t \in [0,T], \tag{10}$$

where $x_0 \in \mathbb{R}^d$ and $\hat b(s,x) := b\left(x, \mathsf{P}_{X_s^\xi}\right)$. Clearly, under our assumptions, (10) admits a unique solution in the class $\mathcal{S}^2\left([0,T];\mathbb{R}^d\right)$, that we denote by $X^{x_0,\xi}$. Notice that this solution only depends on the law of ξ: if $\mathsf{P}_\xi = \mathsf{P}_{\xi'}$, we have $\mathsf{P}_{X_t^\xi} = \mathsf{P}_{X_t^\xi}$ for every $t \in [0,T]$, so that $X^{x_0,\xi}$ and $X^{x_0,\xi'}$ are indistinguishable. Then, we can define without ambiguity $X_t^{x_0,\mathsf{P}_\xi} := X_t^{x_0,\xi}$. More generally, given $\hat\xi \in L^2(\mathcal{F}_0)$, we define without ambiguity

$$X_t^{\hat\xi,\xi}(\omega) := X_t^{\hat\xi,\mathsf{P}_\xi}(\omega) := X_t^{\hat x_0,\mathsf{P}_\xi}(\omega)\,|_{\hat x_0=\hat\xi(\omega)} \quad \text{for a.e. } \omega \in \Omega, \quad t \in [0,T]. \tag{11}$$

In particular

$$X_t^{\xi,\xi}(\omega) := X_t^{\xi,\mathsf{P}_\xi}(\omega) = X_t^{x_0,\mathsf{P}_\xi}(\omega)\,|_{x_0=\xi(\omega)} \quad \text{for a.e. } \omega \in \Omega, \quad t \in [0,T]. \tag{12}$$

Clearly, X_t^{ξ,P_ξ} satisfies (2); hence, by pathwise uniqueness of solutions to (2), we have

$$X^{\xi,\mathsf{P}_\xi} = X^\xi, \quad \mathsf{P}\text{-}a.s. \tag{13}$$

Remark 1. Since the increments of the Brownian motion B are independent of $\mathcal{F}_0$, we observe for later use that

$$\mathsf{E}\Big[X_t^{\hat{\xi},\xi}\,\Big|\,\hat{\xi}=\hat{x}_0,\,\xi=x_0\Big]=\mathsf{E}\Big[X_t^{\hat{x}_0,\delta_{x_0}}\Big],\quad t\in[0,T],\ x_0,\hat{x}_0\in\mathbb{R}^d.$$

In particular, taking $\hat{\xi}=\xi$, we obtain

$$\mathsf{E}\Big[X_t^{\xi,\xi}\,\Big|\,\xi=x_0\Big]=\mathsf{E}\Big[X_t^{\xi,\mathsf{P}_\xi}\,\Big|\,\xi=x_0\Big]=\mathsf{E}\Big[X_t^{x_0,\delta_{x_0}}\Big],\quad t\in[0,T],\ x_0\in\mathbb{R}^d.$$

4.1.1. Derivative of $X_t^{x_0,\mathsf{P}_\xi}$ with respect to x_0

In this subsection we construct the derivative of the auxiliary process $X_t^{x_0,\mathsf{P}_\xi}$ with respect to x_0. To this purpose, we introduce the tangent process $\nabla X^{x_0,\mathsf{P}_\xi}$ as the unique solution of

$$\nabla X_t^{x_0,\mathsf{P}_\xi}=I_d+\int_0^t\partial_x b\Big(X_r^{x_0,\mathsf{P}_\xi},\mathsf{P}_{X_r^\xi}\Big)\nabla X_r^{x_0,\mathsf{P}_\xi}\,dZ_r,\tag{14}$$

where I_d denotes the identity matrix of $\mathbb{R}^{d\times d}$. Under Assumption 2, for every $\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d)$, the map $x_0\mapsto X^{x_0,\mathsf{P}_\xi}\in\mathcal{S}^2\big([0,T];\mathbb{R}^d\big)$ is continuously differentiable with respect to x_0 and

$$\lim_{h\to0}\frac{1}{|h|}\Big\|X^{x_0+h,\mathsf{P}_\xi}-X^{x_0,\mathsf{P}_\xi}-\nabla X^{x_0,\mathsf{P}_\xi}h\Big\|_{\mathcal{S}^2}=0.\tag{15}$$

Moreover, there exists $C>0$ such that, for every $t\in[0,T]$, every $x_0,x_0'\in\mathbb{R}^d$, and every $\xi,\xi'\in L^2\big(\mathcal{F}_0;\mathbb{R}^d\big)$, we have

$$\Big\|\nabla X^{x_0,\mathsf{P}_\xi}\Big\|_{\mathcal{S}^2}^2\le C,\tag{16}$$

and

$$\Big\|\nabla X^{x_0,\mathsf{P}_\xi}-\nabla X^{x_0',\mathsf{P}_{\xi'}}\Big\|_{\mathcal{S}^2}^2\le C\Big(\Big|x_0-x_0'\Big|^2+\mathcal{W}_2(\mathsf{P}_\xi,\mathsf{P}_{\xi'})^2\Big),\tag{17}$$

as proved in [26, Lemma 4.1]. Hence, the limit in (15) is uniform in x_0,ξ and, in particular, there exists $C>0$ such that,

$$\Big\|X^{x_0+h,\mathsf{P}_\xi}-X^{x_0,\mathsf{P}_\xi}-\nabla X^{x_0,\mathsf{P}_\xi}h\Big\|_{\mathcal{S}^2}\le C|h|^2,\quad \xi\in L^2(\mathcal{F}_0;\mathbb{R}^d),\ x_0,h\in\mathbb{R}^d.\tag{18}$$

As observed in [26, Remark 4.1], the process $\nabla X^{\xi,\mathsf{P}_\xi}:=\nabla X^{x_0,\mathsf{P}_\xi}\Big|_{x_0=\xi}$ is the unique solution in $\mathcal{S}^2\big([0,T];\mathbb{R}^{d\times d}\big)$ to the SDE

$$\nabla X_t^{\xi,\mathsf{P}_\xi}=I_d+\int_0^t\partial_x b\Big(X_r^\xi,\mathsf{P}_{X_r^\xi}\Big)\nabla X_r^{\xi,\mathsf{P}_\xi}\,dZ_r.\tag{19}$$

4.1.2. Derivative of $X_t^{x_0,\mathsf{P}_\xi}$ with respect to ξ

Let us consider a copy of $(\Omega,\mathcal{F},\mathbb{F},\mathsf{P})$, which we denote by $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{F}},\tilde{\mathsf{P}})$. In this setting, given any random variable X defined on $(\Omega,\mathcal{F},\mathbb{F},\mathsf{P})$, we may consider a random variable $\tilde{X}$, which is identical to X, but it is defined on $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{F}},\tilde{\mathsf{P}})$; hence, the expectation $\tilde{\mathsf{E}}[\tilde{X}]:=\int_{\tilde{\Omega}}\tilde{X}\,d\tilde{P}$ will apply only to $\tilde{X}$.

For $\xi,\eta\in L^2(\mathcal{F}_0;\mathbb{R}^d)$, consider the family of mean-field SDEs with values in $\mathbb{R}^d$

$$Y_t^\xi(\eta)=\int_0^t\Big(\partial_x b\Big(X_r^\xi,\mathsf{P}_{X_r^\xi}\Big)Y_r^\xi(\eta)+\tilde{\mathbb{E}}\Big[\partial_{\tilde{x}}\delta_\mu b\Big(X_r^{\xi,\mathsf{P}_\xi},\mathsf{P}_{X_r^\xi},\tilde{X}_r^{\tilde{\xi}}\Big)\Big(\nabla\tilde{X}_r^{\tilde{\xi},\mathsf{P}_\xi}\tilde{\eta}+\tilde{Y}_r^{\tilde{\xi}}(\tilde{\eta})\Big)\Big]\Big)dZ_r,\tag{20}$$

where $(\tilde{\xi},\tilde{\eta},\tilde{B},\tilde{X}^{\tilde{\xi}},\tilde{Y}^{\tilde{\xi}}(\tilde{\eta}))$ is a copy of $(\xi,\eta,B,X^\xi,Y^\xi(\eta))$ defined on the probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{F}},\tilde{\mathsf{P}})$. In addition, given a copy $\tilde{Y}_r^{\tilde{\xi}}(\tilde{\eta})$ of a solution $Y_r^\xi(\eta)$ to (20), we also consider the standard SDE for all $x_0\in\mathbb{R}^d$:

$$Y_t^{x_0,\xi}(\eta)=\int_0^t\Big(\partial_x b\Big(X_r^{x_0,\xi},\mathsf{P}_{X_r^\xi}\Big)Y_r^{x_0,\xi}(\eta)+\tilde{\mathbb{E}}\Big[\partial_{\tilde{x}}\delta_\mu b\Big(X_r^{x_0,\mathsf{P}_\xi},\mathsf{P}_{X_r^\xi},\tilde{X}_r^{\tilde{\xi}}\Big)\Big(\nabla\tilde{X}_r^{\tilde{\xi},\mathsf{P}_\xi}\tilde{\eta}+\tilde{Y}_r^{\tilde{\xi}}(\tilde{\eta})\Big)\Big]\Big)dZ_r,$$

As in [26, Eqs. (4.15)–(4.16)], due to the boundedness of the derivatives of b,σ (see Assumption 2), the system of equations above admits a unique solution $\big(Y^\xi(\eta),Y^{x_0,\xi}(\eta)\big)\in\mathcal{S}^2\big([0,T];\mathbb{R}^d\times\mathbb{R}^d\big)$. By uniqueness of solutions to (20), we have

$$Y_t^{\xi,\xi}(\eta):=Y_t^{x_0,\xi}(\eta)\,|_{x_0=\xi}=Y_t^\xi(\eta).\tag{21}$$

Notice that $Y_t^\xi(\eta),Y_t^{x_0,\xi}(\eta)$ are linear in η (i.e. $\alpha_1Y_t^\xi(\eta)+\alpha_2Y_t^\xi(\eta')$ satisfies (20) for $\alpha_1\eta+\alpha_2\eta'$ and then, thanks to uniqueness of solutions we have $\alpha_1Y_t^\xi(\eta)+\alpha_2Y_t^\xi(\eta')=Y_t^{x_0,\xi}(\alpha_1\eta+\alpha_2\eta')$); hence we will write $Y_t^\xi\eta$ and $Y_t^{x_0,\xi}\eta$. Moreover, as in [26, Eq. (4.17)], there exists $C>0$ such that for each $x_0\in\mathbb{R}^d$ and $\xi,\eta\in L^2(\Omega;\mathbb{R}^d)$, one has

$$\|Y_t^{x_0,\xi}\eta\|_{L^2}\le\Big\|Y^{x_0,\xi}\eta\Big\|_{\mathcal{S}^2}\le C\|\eta\|_{L^2},$$

for every $t\in[0,T]$; this implies $Y_t^{x_0,\xi}\in\mathcal{L}\big(L^2(\mathcal{F}_0;\mathbb{R}^d),L^2(\mathcal{F}_t;\mathbb{R}^d)\big)$, for every $t\in[0,T]$.

Now, let $x_0,y\in\mathbb{R}^d,\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d)$ and consider also the family of mean-field SDEs with values in $\mathbb{R}^{d\times d}$

$$U_t^{\xi}(y)=\int_0^t\left(\partial_x b\Big(X_r^{\xi},\mathsf{P}_{X_r^{\xi}}\Big)U_r^{\xi}(y)+\tilde{\mathsf{E}}\Big[\partial_{\tilde{x}}\delta_\mu b\Big(X_r^{\xi},\mathsf{P}_{X_r^{\xi}},\tilde{X}_r^{y,\mathsf{P}_\xi}\Big)\Big(\nabla\tilde{X}_r^{y,\mathsf{P}_\xi}+\tilde{U}_r^{\tilde{\xi}}(y)\Big)\Big]\right)dZ_r,$$

where $\left(\tilde{\xi},\tilde{B},\tilde{X}^{x_0,\mathsf{P}_\xi},\tilde{U}^{\tilde{\xi}}(y)\right)$ is a copy of $\left(\xi,B,X^{x_0,\mathsf{P}_\xi},U^{\xi}(y),\right)$ defined on the probability space $(\tilde{\Omega},\tilde{\mathcal{F}},(\tilde{\mathcal{F}}_t)_{t\in[0,T]},\tilde{\mathsf{P}})$. Given a copy $\tilde{U}_t^{\tilde{\xi}}(y)$ of a solution $U_t^{\xi}(y)$ to such equation, consider also the standard SDE

$$U_t^{x_0,\xi}(y)=\int_0^t\left(\partial_x b\Big(X_r^{x_0,\mathsf{P}_\xi},\mathsf{P}_{X_r^{\xi}}\Big)U_r^{x_0,\xi}(y)+\tilde{\mathsf{E}}\Big[\partial_{\tilde{x}}\delta_\mu b\Big(X_r^{x_0,\mathsf{P}_\xi},\mathsf{P}_{X_r^{\xi}},\tilde{X}_r^{y,\mathsf{P}_\xi}\Big)\Big(\nabla\tilde{X}_r^{y,\mathsf{P}_\xi}+\tilde{U}_r^{\tilde{\xi}}(y)\Big)\Big]\right)dZ_r,$$

As in [26, Eqs. (4.48)–(4.49)], the system of equations above admits a unique solution $\left(U_\cdot^{\xi}(\eta),U_\cdot^{x_0,\xi}(y)\right)\in\mathcal{S}^2\left([0,T];\mathbb{R}^{d\times d}\times\mathbb{R}^{d\times d}\right)$ and we have again

$$U_t^{\xi,\xi}(y):=U_t^{x_0,\xi}(y)\Big|_{x_0=\xi}=U_t^{\xi}(y).\tag{22}$$

By [26, Lemma 4.3], under Assumption 2, there exists $C>0$ such that for every $t\in[0,T],x_0,x_0'\in\mathbb{R}^d$, $\xi,\xi'\in L^2(\mathcal{F}_0;\mathbb{R}^d)$,

$$\left\|U^{x_0,\xi}(y)\right\|_{\mathcal{S}^2}\le C,\quad\left\|U^{x_0,\xi}(y)-U^{x_0',\xi'}(y')\right\|_{\mathcal{S}^2}\le C\Big(\left|x_0-x_0'\right|+|y-y'|+\mathcal{W}_2\big(\mathsf{P}_\xi,\mathsf{P}_{\xi'}\big)\Big).\tag{23}$$

Due to the result above, we can define without ambiguity

$$U_t^{x_0,\mathsf{P}_\xi}(y):=U_t^{x_0,\xi}(y).$$

We now have all ingredients to state the Fréchet differentiability of $X_t^{x_0,\mathsf{P}_\xi}$ with respect to ξ.

Lemma 2. *Let Assumption 2 hold.*

(i) *For all* $(t,x_0)\in[0,T]\times\mathbb{R}^d$*, the map* $\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d)\longmapsto X_t^{x_0,\mathsf{P}_\xi}\in L^2(\mathcal{F}_t;\mathbb{R}^d)$ *is continuously Fréchet differentiable; i.e. there exists a continuous map* $\xi\mapsto\partial_\xi X_t^{x_0,\mathsf{P}_\xi}\in\mathcal{L}\left(L^2(\mathcal{F}_0;\mathbb{R}^d);L^2(\mathcal{F}_t;\mathbb{R}^d)\right)$*, s.t.*

$$\lim_{\|\eta\|_{L^2}\to0}\frac{1}{\|\eta\|_{L^2}}\left\|X_t^{x_0,\mathsf{P}_{\xi+\eta}}-X_t^{x_0,\mathsf{P}_\xi}-\partial_\xi X_t^{x_0,\mathsf{P}_\xi}\eta\right\|_{L^2}=0,\quad\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d).$$

(ii) *The Fréchet derivative* $\partial_\xi X_t^{x_0,\mathsf{P}_\xi}$ *satisfies*

$$\partial_\xi X_t^{x_0,\mathsf{P}_\xi}\eta=Y_t^{x_0,\xi}\eta=\tilde{\mathsf{E}}\Big[U_t^{x_0,\mathsf{P}_\xi}(\tilde{\xi})\tilde{\eta}\Big]\quad t\in[0,T],\ \eta\in L^2(\mathcal{F}_0;\mathbb{R}^d).\tag{24}$$

(iii) *There exists* $C>0$ *such that*

$$\left\|\partial_\xi X_t^{x_0,\mathsf{P}_\xi}\right\|_{\mathcal{L}}\le C,\qquad t\in[0,T],\ x_0\in\mathbb{R}^d,\ \xi\in L^2(\mathcal{F}_0;\mathbb{R}^d).$$

(iv) *There exists* $C>0$ *such that for every* $t\in[0,T],\ x_0,x_0'\in\mathbb{R}^d,\ \xi,\xi'\in L^2(\mathcal{F}_0;\mathbb{R}^d)$*,*

$$\left\|\partial_\xi X_t^{x_0,\mathsf{P}_\xi}-\partial_\xi X_t^{x_0',\mathsf{P}_{\xi'}}\right\|_{\mathcal{L}}\le C\big(|x_0-x_0'|+\mathcal{W}_2\big(\mathsf{P}_\xi,\mathsf{P}_{\xi'}\big)\big).$$

In particular, it follows that the limit in (i) is uniform in t,x_0,ξ *and there exists* $C>0$ *such that, for every* $x_0\in\mathbb{R}^d$*,* $t\in[0,T]$*, and* $\xi\in L^2(\mathcal{F}_0;\mathbb{R}^d)$*,*

$$\left\|X_t^{x_0,\mathsf{P}_{\xi+\eta}}-X_t^{x_0,\mathsf{P}_\xi}-\partial_\xi X_t^{x_0,\mathsf{P}_\xi}\eta\right\|_{L^2}\le C\|\eta\|_{L^2}^2.$$

Proof. (i), (ii), and (iii) follows directly from [26, Proposition 4.2]. As for (iv), notice that the estimate in (iv) with $x_0=x_0'$ has been shown in the proof of [26, Proposition 4.2], using 23. Such calculation can be extended to prove (iv) for general $x_0\neq x_0'$.

With usual notations, we have

$$\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta:=\partial_\xi X_t^{x_0,\mathsf{P}_\xi}\Big|_{x_0=\xi}\eta=\tilde{\mathsf{E}}\Big[U_t^{x_0,\mathsf{P}_\xi}(\tilde{\xi})\tilde{\eta}\Big]\Big|_{x_0=\xi}=\tilde{\mathsf{E}}\big[U_t^{\xi}(\tilde{\xi})\tilde{\eta}\big].\tag{25}$$

Remark 2. As for Remark 1, for every $x_0,h\in\mathbb{R}^d$, we have

$$\mathsf{E}\Big[\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\,\Big|(\xi,\eta)=(x_0,h)\Big]=\mathsf{E}\Big[\partial_\xi X_t^{x_0,\delta_{x_0}}h\Big]=\mathsf{E}\big[Y_t^{x_0}h\big]=\mathsf{E}\big[\tilde{E}[U_t^{x_0}(x_0)h]\big].\tag{26}$$

In view of the equality $X_t^{\xi}=X_t^{\xi,\mathsf{P}_\xi}$, we warn the reader not to confuse $\partial_\xi X_t^{\xi,\mathsf{P}_\xi}$, i.e. the (partial) derivative with respect to ξ of $X_t^{x_0,\mathsf{P}_\xi}$ evaluated at $x_0=\xi$, with the (total) derivative of $X_t^{\xi}=X_t^{\xi,\mathsf{P}_\xi}$ with respect to ξ, which we will denote by $D_\xi X_t^{\xi}$.

4.2. Derivative of X_t^ξ with respect to ξ

Recall (12) and consider the map

$$\xi \in L^2(\mathcal{F}_0;\mathbb{R}^d) \longmapsto X_t^\xi = X_t^{\xi,\mathsf{P}_\xi} \in L^2(\mathcal{F}_t;\mathbb{R}^d). \tag{27}$$

Formally, computing the derivative of this map, it is natural to expect (see also [26, Eq. (4.8), p. 840]) that it is the operator $D_\xi X_t^\xi \in \mathcal{L}\big(L^2(\mathcal{F}_0;\mathbb{R}^d); L^2(\mathcal{F}_t;\mathbb{R}^d)\big)$ given by

$$\Big([D_\xi X_t^\xi]\eta\Big)(\omega) := \nabla X_t^{\xi,\mathsf{P}_\xi}(\omega)\eta(\omega) + \Big[\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\Big](\omega), \quad \eta \in L^2(\mathcal{F}_0;\mathbb{R}^d). \tag{28}$$

In this section, we will show that the operator $D_\xi X_t^\xi$ defined above is indeed the Gateaux derivative of (27). Let $\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$. By (19) and (20), we have, for every $\xi \in L^2(\mathcal{F}_0;\mathbb{R}^d)$,

$$D_\xi X_t^\xi \eta = \eta + \int_0^t \Big(\partial_x b(X_s^\xi, \mathsf{P}_{X_s^\xi}) D_\xi X_s^\xi \eta + \tilde{\mathsf{E}}\Big[\partial_{\tilde{x}}\delta_\mu b(X_s^\xi, \mathsf{P}_{X_s^\xi}, \tilde{X}_s^{\tilde{\xi}}) D_\xi \tilde{X}_s^{\tilde{\xi}}\tilde{\eta}\Big]\Big) dZ_s, \tag{29}$$

where $\tilde{X}_s^{\tilde{\xi}}$ is the solution to (2) on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathcal{F}}_t, \tilde{B}_t, \tilde{\mathsf{P}})_{t\in[0,T]}$ with initial condition $\tilde{\xi}$ such that $\tilde{\mathsf{P}}_{\tilde{\xi}} = \mu_0$. Moreover, due to boundedness of the derivatives of b (cf. Assumption 2), we have that $D_\xi X_t^\xi \eta$ is the unique solution to the equation above.

Proof of Theorem 1. Step 1. We start by showing the estimates of Claim (i) for the operator $D_\xi X_t^\xi$ defined in (29). Indeed, we have

$$\left\|D_\xi X_t^\xi \eta\right\|_{L^2} \le \left\|\nabla X_t^{\xi,\mathsf{P}_\xi}\eta\right\|_{L^2} + \left\|\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\right\|_{L^2}, \quad \eta \in L^2(\mathcal{F}_0;\mathbb{R}^d).$$

For the first term on the right-hand-side we have

$$\left\|\nabla X_t^{\xi,\mathsf{P}_\xi}\eta\right\|_{L^2} = \left(\mathsf{E}\left[\mathsf{E}\left[\left|\nabla X_t^{\xi,\mathsf{P}_\xi}\eta\right|^2\middle|(\xi,\eta)\right]\right]\right)^{1/2}.$$

By the independence of (ξ,η) (which are $\mathcal{F}_0$−measurable) and the increments of B_t (which are independent of $\mathcal{F}_0$), and using (16), for all $x_0, h \in \mathbb{R}^d$,

$$\mathsf{E}\left[\left|\nabla X_t^{\xi,\mathsf{P}_\xi}\eta\right|^2\middle|(\xi,\eta) = (x_0,h)\right] = \mathsf{E}\left[\left|\nabla X_t^{x_0,\delta_{x_0}} h\right|^2\right] \le C|h|^2,$$

so that $\left\|\nabla X_t^{\xi,\mathsf{P}_\xi}\eta\right\|_{L^2} \le C\|\eta\|_{L^2}$. For the second term on the right-hand-side, we have

$$\left\|\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\right\|_{L^2} = \left(\mathsf{E}\left[\mathsf{E}\left[\left|\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\right|^2\middle|(\xi,\eta)\right]\right]\right)^{1/2}$$

By Remark 2 and Lemma 2, for all $x_0, h \in \mathbb{R}^d$

$$\mathsf{E}\left[\left|\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\right|^2\middle|(\xi,\eta) = (x_0,h)\right] = \mathsf{E}\left[\left|\partial_\xi X_t^{x_0,\delta_{x_0}} h\right|^2\right] \le C|h|^2,$$

so that $\left\|\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\right\|_{L^2} \le C\|\eta\|_{L^2}$. Hence

$$\left\|D_\xi X_t^\xi \eta\right\|_{L^2} \le C\|\eta\|_{L^2},$$

so that we have the claim.

Next, we prove the uniform strong continuity estimate. For $\eta, \xi, \xi' \in L^2(\mathcal{F}_0;\mathbb{R}^d)$, we have:

$$\begin{aligned}\left\|\left(D_\xi X_t^\xi - D_\xi X_t^{\xi'}\right)\eta\right\|_{L^2} \le E_t^1(\eta) + E_t^2(\eta), \text{ where } E_t^1(\eta) &:= \left\|\left(\nabla X_t^{\xi,\mathsf{P}_\xi} - \nabla X_t^{\xi',\mathsf{P}_{\xi'}}\right)\eta\right\|_{L^2} \\ E_t^2(\eta) &:= \left\|\left(\partial_\xi X_t^{\xi,\mathsf{P}_\xi} - \partial_\xi X_t^{\xi',\mathsf{P}_{\xi'}}\right)\eta\right\|_{L^2}.\end{aligned} \tag{30}$$

- Consider $E_t^1(\eta)$. By the independence of (ξ,ξ') and the increments of B_t, it follows from (17) that for all $x_0, x_0' \in \mathbb{R}^d$

$$\begin{aligned}\mathsf{E}\left[\left|\nabla X_t^{\xi,\mathsf{P}_\xi} - \nabla X_t^{\xi',\mathsf{P}_{\xi'}}\right|^2\middle|(\xi,\xi') = (x_0,x_0')\right] &= \mathsf{E}\left[\left|\nabla X_t^{x_0,\delta_{x_0}} - \nabla X_t^{x_0',\delta_{x_0'}}\right|^2\right] \\ &\le C\left(\left|x_0 - x_0'\right|^2 + \mathcal{W}_2\left(\delta_{x_0},\delta_{x_0'}\right)^2\right) \le C\left|x_0 - x_0'\right|^2,\end{aligned}$$

  which implies that

$$\left\|\nabla X_t^{\xi,\mathsf{P}_\xi} - \nabla X_t^{\xi',\mathsf{P}_{\xi'}}\right\|_{L^2} \le C\|\xi - \xi'\|_{L^2}, \tag{31}$$

and therefore for $\eta \in L^\infty(\mathcal{F}_0;\mathbb{R}^d)$:

$$E_t^1(\eta) \le \left\| \nabla X_t^{\xi,\mathsf{P}_\xi} - \nabla X_t^{\xi',\mathsf{P}_{\xi'}} \right\|_{L^2} \|\eta\|_{L^\infty} \le C\|\xi - \xi'\|_{L^2}\|\eta\|_{L^\infty}. \tag{32}$$

Let now $\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$. By considering $\eta_N := \eta I_{|\eta|\le N} \in L^\infty(\mathcal{F}_0;\mathbb{R}^d)$ (so that $\|\eta_N - \eta\|_{L^2} \to 0$), we have

$$E_t^1(\eta) \le E_t^1(\eta - \eta_N) + E_t^1(\eta_N).$$

Using (16), we have $E_t^1(\eta - \eta_N) \le C\|\eta_N - \eta\|_{L^2}$. Moreover, by (32), we have $E_t^1(\eta_N) \le C\|\xi - \xi'\|_{L^2}\|\eta_N\|_{L^\infty}$. Hence

$$E_t^1(\eta) \le C\|\eta_N - \eta\|_{L^2} + C\|\xi - \xi'\|_{L^2}\|\eta_N\|_{L^\infty}.$$

Let $\epsilon > 0$. Take first $\bar N$ such that $C\|\eta_{\bar N} - \eta\|_{L^2} < \epsilon/2$. Then there exists $\delta > 0$ such that $C\|\xi - \xi'\|_{L^2}\|\eta_{\bar N}\|_{L^\infty} < \epsilon/2$, for all $\|\xi - \xi'\|_{L^2} < \delta$. Hence, for all $\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$ there exists a modulus of continuity ϖ_η such that $E_t^1(\eta) \le \varpi_\eta(\|\xi - \xi'\|_{L^2})$.

- Consider $E_t^2(\eta)$. Let $\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$. By (25), we have

$$\begin{aligned} E_t^2(\eta) &= \left\| \tilde{\mathbb{E}}\left[\left(U_t^{\xi,\mathsf{P}_\xi}(\tilde\xi) - U_t^{\xi',\mathsf{P}_{\xi'}}(\tilde\xi') \right)\tilde\eta \right] \right\|_{L^2} \le \left\| \left(\tilde{\mathbb{E}}\left[\left| U_t^{\xi,\mathsf{P}_\xi}(\tilde\xi) - U_t^{\xi',\mathsf{P}_{\xi'}}(\tilde\xi') \right|^2 \right] \right)^{1/2} \right\|_{L^2} \|\tilde\eta\|_{L^2} \\ &= \left(\mathbb{E}\left[\mathbb{E}\left[\tilde{\mathbb{E}}\left[\left| U_t^{\xi,\mathsf{P}_\xi}(\tilde\xi) - U_t^{\xi',\mathsf{P}_{\xi'}}(\tilde\xi') \right|^2 \right] \middle| (\xi,\xi') \right] \right] \right)^{1/2} \|\eta\|_{L^2}. \end{aligned}$$

As in Remark 2 and using Lemma 2, for all $x_0, x_0' \in \mathbb{R}^d$,

$$\mathbb{E}\left[\tilde{\mathbb{E}}\left[\left| U_t^{\xi,\mathsf{P}_\xi}(\tilde\xi) - U_t^{\xi',\mathsf{P}_{\xi'}}(\tilde\xi') \right|^2 \right] \middle| (\xi,\xi') = (x_0,x_0') \right] = \mathbb{E}\left[\tilde{\mathbb{E}}\left[\left| U_t^{x_0,\delta_{x_0}}(x_0) - U_t^{x_0',\delta_{x_0'}}(x_0') \right|^2 \right] \right] \le C|x_0 - x_0'|^2.$$

It follows $E_t^2(\eta) \le C\|\xi - \xi'\|_{L^2}\|\eta\|_{L^2}$. Notice that we have actually proved that

$$\left\| \partial_\xi X_t^{\xi,\mathsf{P}_\xi} - \partial_\xi X_t^{\xi',\mathsf{P}_{\xi'}} \right\|_{\mathcal{L}} \le C\|\xi - \xi'\|_{L^2}, \tag{33}$$

a stronger estimate implying the required claim.

Step 2. We now complete the proof of Claim (i), i.e. that we have, for $D_\xi X_t^\xi$ defined in (29),

$$I_t(|r|,\eta) := \frac{1}{|r|}\left\| X_t^{\xi + r\eta} - X_t^\xi - rD_\xi X_t^\xi \eta \right\|_{L^2} \xrightarrow{r\to 0} 0, \quad \eta \in L^2(\mathcal{F}_0;\mathbb{R}^d). \tag{34}$$

We first provide an estimate when $\eta \in L^4(\mathcal{F}_0;\mathbb{R}^d)$ and then use a density argument. Given $r \in \mathbb{R}$ and $\eta \in L^4(\mathcal{F}_0;\mathbb{R}^d)$, we have

$$\begin{aligned} I_t(|r|,\eta) &= \frac{1}{|r|}\left\| X_t^{\xi+r\eta,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_\xi} - \left[r\nabla X_t^{\xi,\mathsf{P}_\xi}\eta + r\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta \right] \right\|_{L^2} \\ &\le \frac{1}{|r|}\left\| X_t^{\xi+r\eta,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_{\xi+r\eta}} - r\nabla X_t^{\xi,\mathsf{P}_\xi}\eta \right\|_{L^2} \\ &\quad + \frac{1}{|r|}\left\| X_t^{\xi,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_\xi} - r\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta \right\|_{L^2} =: I_t^1(|r|,\eta) + I_t^2(|r|,\eta). \end{aligned}$$

- As for $I_t^1(|r|,\eta)$, we have

$$I_t^1(|r|,\eta) = \frac{1}{|r|}\left(\mathsf{E}\left[\mathsf{E}\left[\left| X_t^{\xi+r\eta,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_{\xi+r\eta}} - r\nabla X_t^{\xi,\mathsf{P}_{\xi+r\eta}}\eta \right|^2 \middle| (\xi,\eta) \right] \right] \right)^{1/2}.$$

Since (ξ,η) is independent of B_t and using Remarks 1, 2, as wells as (15), (17), we have for every $x_0, h \in \mathbb{R}^d$

$$\begin{aligned} &\mathsf{E}\left[\left| X_t^{\xi+r\eta,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_{\xi+r\eta}} - r\nabla X_t^{\xi,\mathsf{P}_\xi}\eta \right|^2 \middle| (\xi,\eta) = (x_0,h) \right] \\ &= \mathsf{E}\left[\left| X_t^{x_0+rh,\mathsf{P}_{x_0+rh}} - X_t^{x_0,\mathsf{P}_{x_0+rh}} - r\nabla X_t^{x_0,\delta_{x_0}} h \right|^2 \right] \\ &= \mathsf{E}\left[\left| \int_0^1 \nabla X_t^{x_0+\theta rh,\mathsf{P}_{x_0+rh}} rh\, d\theta - r\nabla X_t^{x_0,\delta_{x_0}} h \right|^2 \right] \\ &\le C\left(|r|^2|h|^2 + \mathcal{W}_2(\mathsf{P}_{x_0+rh},\delta_{x_0})^2 \right)|r|^2|h|^2 \le C|r|^4|h|^4, \end{aligned}$$

for $C > 0$ independent of $x_0, |r|, |h|$. Hence $I_t^1(|r|,\eta) \le C|r|\left(\mathsf{E}\left[|\eta|^4\right]\right)^{1/2} = C|r|\|\eta\|_{L^4}^2$.

- As for $I_t^2(|r|,\eta)$, we have

$$I_t^2(|r|,\eta) = \frac{1}{|r|}\left(\mathsf{E}\left[\mathsf{E}\left[\left| X_t^{\xi,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_\xi} - r\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta \right|^2 \middle| (\xi,\eta) \right] \right] \right)^{1/2}.$$

As before, using now Remarks 1, 2, and Lemma 2, we have for every $x_0, h \in \mathbb{R}^d$

$$\mathsf{E}\left[\left|X_t^{\xi,\mathsf{P}_{\xi+r\eta}} - X_t^{\xi,\mathsf{P}_\xi} - r\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\eta\right|^2 \middle| (\xi,\eta) = (x_0,h)\right]$$
$$= \mathsf{E}\left[\left|X_t^{x_0,\mathsf{P}_{x_0+rh}} - X_t^{x_0,\delta_{x_0}} - r\partial_\xi X_t^{x_0,\delta_{x_0}} h\right|^2\right] \le C|r|^4|h|^4,$$

where C is independent of $x_0, |r|, |h|$. Hence, $I_t^2(|r|,\eta) \le C|r|\big(\mathsf{E}\big[|\eta|^4\big]\big)^{1/2} = C|r|\|\eta\|_{L^4}^2$.

We therefore conclude that there exists $C > 0$ (independent of η) such that

$$I_t(|r|,\eta) \le C|r|\|\eta\|_{L^4}^2. \tag{35}$$

Now, let $\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$ and define $\eta_N := \eta I_{|\eta|\le N} \in L^\infty(\mathcal{F}_0;\mathbb{R}^d)$. Clearly, we have $\|\eta_N - \eta\|_{L^2} \to 0$ as $N \to \infty$. By (4), the estimate on the operator norm of $D_\xi X_t^\xi$, and (35), we have

$$\begin{aligned} I_t(|r|,\eta) &\le \frac{1}{|r|}\left\|X_t^{\xi+r\eta} - X_t^\xi - rD_\xi X_t^\xi \eta - \left\{X_t^{\xi+r\eta_N} - X_t^\xi - rD_\xi X_t^\xi \eta_N\right\}\right\|_{L^2} + I_t(|r|,\eta_N) \\ &\le \frac{1}{|r|}\left\|X_t^{\xi+r\eta} - X_t^{\xi+r\eta_N}\right\|_{L^2} + \left\|D_\xi X_t^\xi(\eta-\eta_N)\right\|_{L^2} + I_t(|r|,\eta_N) \\ &\le 2C\|\eta-\eta_N\|_{L^2} + C|r|\|\eta_N\|_{L^4}^2. \end{aligned}$$

Take now $\varepsilon > 0$ and choose first $N \ge 1$ such that $2C\|\eta-\eta_N\|_{L^2} < \varepsilon/2$ and then $\delta > 0$ such that $C\delta\|\eta_N\|_{L^4}^2 < \varepsilon/2$. Then, we get $I_t(|r|,\eta) < \varepsilon$ for every $|r| < \delta$. By arbitrariness of ε, we have proved (34).

Step 3. We now prove (ii). Clearly, the adjoint operator $\left(D_\xi X_t^\xi\right)^* \in \mathcal{L}(L^2(\mathcal{F}_t;\mathbb{R}^d); L^2(\mathcal{F}_0;\mathbb{R}^d))$ satisfies

$$\left(D_\xi X_t^\xi\right)^* = \left(\partial_{x_0} X_t^{\xi,\mathsf{P}_\xi}\right)^* + \left(\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\right)^*.$$

Notice that $\left(\nabla X_t^\xi\right)^* \in \mathcal{L}\big(L^2(\mathcal{F}_t;\mathbb{R}^d); L^2(\mathcal{F}_0;\mathbb{R}^d)\big)$ satisfies

$$\left\langle \left(\nabla X_t^\xi\right)^*\xi_t, \eta\right\rangle_{L^2} = \left\langle \xi_t, \nabla X_t^\xi \eta\right\rangle_{L^2} = \mathsf{E}\left[\left\langle \xi_t, \nabla X_t^\xi\eta\right\rangle_{\mathbb{R}^d}\right] = \mathsf{E}\left[\left\langle \left(\nabla X_t^\xi\right)^T\xi_t, \eta\right\rangle_{\mathbb{R}^d}\right] = \left\langle \left(\nabla X_t^\xi\right)^T\xi_t, \eta\right\rangle_{L^2},$$

for every $\xi_t \in L^2(\mathcal{F}_t;\mathbb{R}^d), \eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$. Hence, $\left(D_\xi X_t^\xi\right)^*$ is characterized by

$$\left[\left(D_\xi X_t^\xi\right)^*\eta\right](\omega) = \left(\partial_{x_0} X_t^{\xi,\mathsf{P}_\xi}\right)^T(\omega)\eta(\omega) + \left[\left(\partial_\xi X_t^{\xi,\mathsf{P}_\xi}\right)^*\eta\right](\omega). \tag{36}$$

As for the estimate on the operator norm of $\left(D_\xi X_t^\xi\right)^*$, it directly follows from the corresponding estimate for $D_\xi X_t^\xi$.

Finally, we prove the uniform strong continuity estimate of $\left(D_\xi X_t^\xi\right)^*$. Indeed, we have

$$\begin{aligned} \left\|\left(D_\xi X_t^\xi - D_\xi X_t^{\xi'}\right)^*\eta\right\|_{L^2} &\le \left\|\left(\nabla X_t^{\xi,\mathsf{P}_\xi} - \nabla X_t^{\xi',\mathsf{P}_{\xi'}}\right)^T\eta\right\|_{L^2} \\ &\quad + \left\|\left(\partial_\xi X_t^{\xi,\mathsf{P}_\xi} - \partial_\xi X_t^{\xi',\mathsf{P}_{\xi'}}\right)^*\eta\right\|_{L^2} =: \tilde{E}_t^1(\eta) + \tilde{E}_t^2(\eta). \end{aligned}$$

First, by (33) we have

$$\tilde{E}_t^2(\eta) \le \left\|\left(\partial_\xi X_t^{\xi,\mathsf{P}_\xi} - \partial_\xi X_t^{\xi',\mathsf{P}_{\xi'}}\right)^*\right\|_{\mathcal{L}} \|\xi-\xi'\|_{L^2} \le C\|\xi-\xi'\|_{L^2}.$$

Next, for every $\eta \in L^\infty((\Omega,\mathcal{F}_t,\mathsf{P});\mathbb{R}^d)$, by (31) we have

$$\tilde{E}_t^1(\eta) \le \left\|\left(\nabla X_t^{\xi'} - \nabla X_t^\xi\right)^T\right\|_{L^2} \|\eta\|_{L^\infty} \le C\|\xi'-\xi\|_{L^2}\|\eta\|_{L^\infty}.$$

Hence, for $\eta \in L^2(\mathcal{F}_t;\mathbb{R}^d)$, we can proceed as we did with E_t^1 in the previous step to get $\tilde{E}_t^1(\eta) \le \varpi_\eta(\|\xi'-\xi\|_{L^2})$. The claim follows.

Remark 3. (i) We report here an argument communicated to us by R. Buckdahn and J. Li, which shows that in general we may not expect that the map $L^2(\mathcal{F}_0;\mathbb{R}^d) \to L^2(\mathcal{F}_t;\mathbb{R}^d)$, $\xi \mapsto X_t^\xi$ is Fréchet differentiable. Consider (2) with $b^0(x,\mu) = b^0(x)$, $b^1(x,\mu) = b^1(x)$ and $b^0, b^1 \in C^2$ with bounded and Lipschitz first and second order derivatives. Fix $t \in [0,T]$ and let $\xi,\eta \in L^2(\mathcal{F}_0;\mathbb{R}^d)$. Then, with usual notations, we have $X_t^\xi = X_t^x\,|_{x=\xi}$, $X_t^{\xi+\eta} - X_t^\xi = \int_0^1 \nabla X_t^{\xi+\theta\eta}\eta d\theta$, and

$$X_t^{\xi+\eta} - X_t^\xi - \nabla X_t^\xi\eta = \int_0^1 \left[\nabla X_t^{\xi+\theta\eta} - \nabla X_t^\xi\right] d\theta\ \eta = \left[\int_0^1 \int_0^\theta \partial_{x_0^2} X_t^{\xi+\lambda\eta} d\lambda d\theta\ \eta\right]\eta,$$

where the second order derivative process $\partial_{x_0^2} X_t^\xi$ was defined in [26, Section 5]. Hence, in general we may only expect

$$\left\| X_t^{\xi+\eta} - X_t^\xi - \nabla X_t^\xi \eta \right\|_{L^2} = \left(\mathsf{E}\left\{ \left| \left[\int_0^1 \int_0^\theta \partial_{x_0^2} X_t^{\xi+\lambda\eta} d\lambda d\theta \; \eta \right] \eta \right|^2 \right\} \right)^{1/2} \approx \left(\mathsf{E}\left[|\eta|^4 \right] \right)^{1/2}, \tag{37}$$

and not a behavior of the type $o\Big(\left(\mathsf{E}\left[|\eta|^2 \right] \right)^{1/2} \Big)$ for $\left(\mathsf{E}\left[|\eta|^2 \right] \right)^{1/2} \to 0$, as required by the definition of Fréchet differentiability.

5. DRO Sensitivity for a systemic risk model

We consider an extension of the systemic risk model presented in [22, Section 2] to the case of multiplicative noise. Consider a model of interbank borrowing and lending of N banks, where the log-monetary reserve of each bank i given by the following 1-dimensional SDE

$$\mathrm{d}X_s^i = \frac{a}{N} \sum_{j=1}^N (X_t^j - X_t^i)\,\mathrm{d}s + \sigma(X_t^i)\,\mathrm{d}B_s^i, \qquad X_t^i = \xi^i, \tag{38}$$

where $a > 0$ measures the rate of borrowing/lending between bank i and bank j, $\sigma \in C_b^{1,1}(\mathbb{R})$ is the volatility coefficient of the bank reserve, B^i are standard real-valued i.i.d. Wiener processes, and $\xi^i \sim \mu_0 \in \mathcal{P}_2(\mathbb{R})$ are i.i.d random variables. When $N \to \infty$, the log-monetary reserve of a representative bank is provided by the McKean-Vlasov SDE

$$\mathrm{d}X_t = a\big(\mathsf{E}[X_t] - X_t\big)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \qquad X_0 = \xi \in L^2(\mathcal{F}_0),\ \mathsf{P}_\xi = \mu_0 \in \mathcal{P}_2(\mathbb{R}), \tag{39}$$

which is covered by Example 3, with

$$b(x,\mu) := \left(a\left(\int_{\mathbb{R}} y\mu(dy) - x \right), \sigma(x) \right), \quad \partial_x b(x,\mu) = (-a, \partial_x \sigma(x)), \quad \partial_{\tilde{x}} \partial_\mu b(x,\mu,\tilde{x}) = (a,0).$$

Moreover, we have

$$\mathsf{E}\left[X_t^\xi \right] = \mathsf{E}[\xi], \quad t \geq 0. \tag{40}$$

In order to evaluate systemic risk, according to the criterion in [22], we evaluate the variance of the log monetary reserve at some time $T > 0$, i.e.

$$\phi(\mu_T) = \mathsf{Var}[X_T^\xi] = \mathsf{E}[(X_T^\xi)^2] - (\mathsf{E}[X_T^\xi])^2,$$

where

$$\phi : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}, \;\; \phi(\mu) := \int_{\mathbb{R}} x^2 \mu(dx) - \left(\int_{\mathbb{R}} x\mu(dx) \right)^2.$$

Notice that both the functional $\int_{\mathbb{R}} x^2 \mu(dx)$ and $\left(\int_{\mathbb{R}} x\mu(dx) \right)^2$ satisfy our assumptions (see Examples 1 and 2) and we have $\partial_x \delta_\mu \phi(\mu, x) = 2x - 2\int_{\mathbb{R}} y\mu(dy)$. Hence, with the notations of Section 3, we may consider

$$\Phi(\mu_0, \delta) := \sup_{\mu_0' \in B_\delta^2(\mu_0)} \phi\left(\mu_T'\right), \qquad \delta \geq 0,$$

and apply Theorem 2 to evaluate the sensitivity with respect to the initial distribution, measuring the change in the variance of the log-monetary reserve with respect to the initial distribution. We get

$$\frac{\partial \Phi}{\partial \delta}(\mu_0, 0) = 2 \left\| \left(D_\xi X_T^\xi \right)^* \left(X_T^\xi - \mathsf{E}\left[X_T^\xi \right] \right) \right\|_{L^2} = 2 \left\| \left(D_\xi X_T^\xi \right)^* \left(X_T^\xi - \mathsf{E}[\xi] \right) \right\|_{L^2}, \tag{41}$$

where we have used (40) and where $D_\xi X_t^\xi \in \mathcal{L}\big(L^2(\mathcal{F}_0); L^2(\mathcal{F}_t)\big)$, $t \in [0,T]$, is provided by Theorem 1 and (29). Here, the latter is characterized as solution to

$$\begin{aligned} D_\xi X_t^\xi \eta &= \eta + \int_0^t \left(-a D_\xi X_s^\xi \eta + a\mathsf{E}\left[D_\xi X_s^\xi \eta \right] \right) ds + \int_0^t \partial_x \sigma(X_s^\xi) D_\xi X_s^\xi \eta \, dB_s \\ &= \eta + \int_0^t \left(-a D_\xi X_s^\xi \eta + a\mathsf{E}[\eta] \right) ds + \int_0^t \partial_x \sigma(X_s^\xi) D_\xi X_s^\xi \eta \, dB_s, \qquad \eta \in L^2(\mathcal{F}_0), \end{aligned}$$

where we used that $\mathsf{E}\left[D_\xi X_t^\xi \eta \right] = \mathsf{E}[\eta]$.

Funding

Filippo de Feo and Fausto Gozzi were supported by the Italian Ministry of University and Research (MIUR), in the framework of PRIN projects 2017FKHBA8 001 (The Time-Space Evolution of Economic Activities: Mathematical Models and Empirical Applications) and 20223PNJ8K (Impact of the Human Activities on the Environment and Economic Decision Making in a Heterogeneous Setting: Mathematical Models and Policy Implications). Filippo de Feo acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – CRC/TRR 388 "Rough Analysis, Stochastic Dynamics and Related Fields" – Project ID 516748464 and by INdAM (Instituto Nazionale di Alta Matematica F. Severi) - GNAMPA (Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni). Salvatore Federico is a member of the GNAMPA project "Problemi di controllo ottimo stocastico con memoria ed informazione parziale" funded by INdAM and of the PRIN project 2022BEMMLZ "Stochastic control and games and the role of information" - CUP: D53D23005780006 funded by the Italian Ministry of University and Research (MUR). Nizar Touzi is partially supported by NSF grant #DMS-2508581.

Acknowledgements

The authors are very grateful to R. Buckdahn and J. Li for the content of Remark 3. In particular, Filippo de Feo is very grateful to R. Buckdahn for valuable discussions concerning the differentiability with respect to the initial datum of the solution of the McKean-Vlasov SDE and for the hospitality at Universite de Bretagne Occidentale, Brest.

References

[1] H. Lam, Robust sensitivity analysis for stochastic systems, 41, 2016. Mathematics of Operations Research.
[2] F. Farokhi, Distributionally robust optimization with noisy data for discrete uncertainties using total variation distance, 2023, pp. 1494–1499.
[3] E. Bayraktar, T. Chen, Nonparametric adaptive robust control under model uncertainty, SIAM J. Control Optim. 61 (2023) 2737–2760.
[4] P. Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations, Math. Program. 171 (2018) 115–166.
[5] J. Blanchet, K. Murthy, Quantifying distributional model risk via optimal transport, 44, Mathematics of Operations Research, 2019.
[6] A. Neufeld, M.N.C. En, Y. Zhang, Robust SGLD algorithm for solving non-convex distributionally robust optimisation problems. 2024, Tech. Report arXiv preprint arXiv:2403.09532.
[7] S. Fuhrmann, M. Kupper, M. Nendel, Wasserstein perturbations of markovian transition semigroups, Annal. de l'Instit. Henri Poincaré 59 (2023). Probabilités Et Statistiques.
[8] J. Blanchet, A. Shapiro, Statistical limit theorems in distributionally robust optimization, in: 2023 Winter Simulation Conference (WSC), 2023, pp. 31–45.
[9] M. Nendel, A. Sgarabottolo, A parametric approach to the estimation of convex risk functionals based on Wasserstein distance. 2022, Techn. Report arXiv preprint arXiv:2210.14340.
[10] J. Blanchet, L. Chen, X.Y. Zhou, Distributionally robust mean-variance portfolio selection with wasserstein distances, Manage. Sci. 68 (2022) 6382–6410.
[11] J. Blanchet, Y. Kang, K. Murthy, Robust wasserstein profile inference and applications to machine learning, J. Appl. Probab. 56 (2019) 830–857.
[12] J. Blanchet, J. Li, S. Lin, X. Zhang, Distributionally robust optimization and robust statistics. 2024, Techn. Report arXiv preprint arXiv:2401.14655.
[13] N. Lanzetti, S. Bolognani, F. Dörfler, First-order conditions for optimization in the Wasserstein space. 2022, Techn. Report arXiv preprint arXiv:2209.12197.
[14] M. Yue, D. Kuhn, W. Wiesemann, On linear optimization over wasserstein balls, Math. Program. 195 (2022) 1107–1122.
[15] D. Bartl, S. Drapeau, J.O. ój, J. Wiesel, Sensitivity analysis of wasserstein distributionally robust optimization problems, in: Proceedings of the Royal Society a. 477.2256, the Royal Society A. 477.2256, 2021, p. 20210176.
[16] D. Bartl, J. Wiesel, Sensitivity of multiperiod optimization problems with respect to the adapted wasserstein distance, SIAM J. Financ. Math. 14 (2023) 704–720.
[17] Y. Jiang, J. Obloj, Sensitivity of causal distributionally robust optimization. 2024, Tech. Report arXiv preprint arXiv:2408.17109.
[18] N. Sauldubois, N. Touzi, First order Martingale model risk and semi-static hedging. 2024, Tech. Report arXiv preprint arXiv:2410.06906.
[19] J.M. Lasry, P.L. Lions, Mean field games, Japanese J. Math. 2 (1) (2007) 229–260.
[20] M. Huang, R.P. Malhamé, P.E. Caines, Nash equilibria for large-population linear stochastic systems of weakly coupled agents". Analysis, control and optimization of complex dynamic systems, Springer US, Boston, MA, Boston, MA, 2005.
[21] R. Carmona, F. Delarue, Probabilistic theory of mean field games with applications I-II, Springer, 2018.
[22] R. Carmona, J.P. Fouque, L.H. Sun, Mean field games and systemic risk, Commun. Math. Sci. textbf 13 (2015) 911–933.
[23] S. Nadtochiy, M. Shkolnikov, Mean field systems on networks, with singular interaction through hitting times, Annals Probab. 48 (3) (2020) 1520–1556.
[24] M.F. Djete, N. Touzi, Mean field game of mutual holding, Annals Appl. Probabil. 34 (6) (2024) 4999–5031.
[25] A.D. Crescenzo, F. de Feo, H. Pham, Linear-quadratic optimal control for non-exchangeable mean-field SDEs and applications to systemic risk. 2025, arXiv preprint arXiv:2503.03318.
[26] R. Buckdahn, J. Li, S. Peng, C. Rainer, Mean-field stochastic differential equations and associated PDEs, Ann. Probab. 45 (2017) 824–878.
[27] J. Chassagneux, D. Crisan, F. Delarue, A probabilistic approach to classical solutions of the master equation for large population equilibria, 280, American Mathematical Society, 1379.
[28] C. Villani, Optimal transport: old and new, 338, Springer, 2009.
[29] T. Hao, J. Li, Mean-field SDEs with jumps and nonlocal integral-PDEs, in: Nonlinear Differential Equations and Applications NoDEA, 2016, pp. 1–51.
[30] R. Buckdahn, J. Li, C. Xing, Doubly-stochastic interpretation for nonlocal semi-linear backward stochastic partial differential equations, J. Differ. Equ. 375 (2023) 1–81.
[31] E. Zeidler, Nonlinear functional analysis and its applications, in: Fixed-point Theorems, 1986, p. 897.