



HAL
open science

Itô's formula for flows of measures on semimartingales

Xin Guo, Huyên Pham, Xiaoli Wei

► To cite this version:

Xin Guo, Huyên Pham, Xiaoli Wei. Itô's formula for flows of measures on semimartingales. *Stochastic Processes and their Applications*, 2023, <10.1016/j.spa.2023.02.004>. <hal-03988009>

HAL Id: hal-03988009

<https://hal.science/hal-03988009v1>

Submitted on 31 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Itô's formula for flows of measures on semimartingales

Xin Guo¹, Huyên Pham^{b,1,*}, Xiaoli Wei^c

^a*Department of Industrial Engineering and Operations Research, UC Berkeley, CA 94720-1777, USA*

^b*LPSM, Université Paris Cité, Building Sophie Germain, Avenue de France, 75013 Paris*

^c*Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, 518055 Shenzhen*

Abstract

We establish Itô's formula along flows of probability measures associated with general semimartingales; this generalizes existing results for flows of measures on Itô processes. Our approach is to first establish Itô's formula for cylindrical functions and then extend it to the general case via function approximation and localization techniques.

This general form of Itô's formula enables the derivation of dynamic programming equations and verification theorems for McKean–Vlasov controls with jump diffusions and for McKean–Vlasov mixed regular-singular control problems. It also allows for generalizing the classical relationship between the maximum principle and the dynamic programming principle to the McKean–Vlasov singular control setting, where the adjoint process is expressed in terms of the derivative of the value function with respect to the probability measures.

Keywords: Itô's formula; Flows of probability measures; Semimartingales; Cylindrical functions; McKean–Vlasov controls

MSC: 60H30, 60K35, 93E20

1. Introduction

Itô's formula is one of the most fundamental building blocks in stochastic calculus. It is the key tool to connect PDEs with functional expectation of diffusion processes. Recent developments in mean-field games and mean-field controls have extended Itô's formula to flows of probability measures. Analogously, this extension is a natural tool for deriving PDEs in the space of probability measures when dealing with SDEs of mean-field type called McKean–Vlasov equations. For instance, the generalized Itô's formula has been used to derive the master equations in mean-field games and the Bellman dynamic programming equation for McKean–Vlasov control problems. (See [13] and the references therein.)

*Corresponding author

Email addresses: xinguo@berkeley.edu (Xin Guo), pham@lpsm.paris (Huyên Pham), xiaoli_wei@sz.tsinghua.edu.cn (Xiaoli Wei)

¹The work of this author is supported by FiME (Finance for Energy Market Research Centre), and the ANR-18-IDEX-0001.

10 There are three approaches to establish Itô's formula along flows of measures associated with
11 Itô processes. The first is to discretize time and mimic the standard proof of Itô's formula [8];
12 this approach can also yield the Itô's formula for flows of measures associated with mean-field
13 jump diffusions [31]. The second is to use the Fokker–Planck approach by assuming the existence
14 of the density for the probability measure [11]. The third, known as the particle approximation
15 approach, is to approximate flows of measures by flows of empirical measures [14, 13]; this has
16 been applied recently in [22], [21] to derive several Itô–Wentzell–Lions formulae on Wiener
17 spaces for real-valued functional random fields that depend on measure flows.

18 *McKean–Vlasov jump diffusion and singular control.* McKean–Vlasov processes, first intro-
19 duced and studied in [33], are stochastic processes governed by SDEs whose coefficients depend
20 on distributions of the solutions. McKean–Vlasov controls are concerned with the optimal con-
21 trol of McKean–Vlasov processes. As in classical control theory, the two main tools to analyze
22 McKean–Vlasov controls are the stochastic maximum principle and the dynamic programming
23 principle. The former has been developed in [2, 12] to study controlled McKean–Vlasov systems
24 in terms of adjoint backward SDEs; it has also been adopted for mean-field games with singular
25 controls [25]. The latter has been developed through a series of works including for Markovian
26 controls [36], open-loop controls [5, 17], Markovian and non-Markovian frameworks [20], and
27 closed-loop controls [41]. None of these works except for [25], however, involves jumps.

28 The dynamic programming principle was developed to analyze the viscosity solution property
29 for a class of jump diffusion processes whose coefficients and control are independent of the state
30 process [9]. Moreover, most existing works on McKean–Vlasov controls deal with Itô processes
31 with continuous paths, except for [30, 25, 29], which considered singular controls and [28, 9],
32 which studied jump diffusions. To the best of our knowledge, there is no prior work on the
33 dynamic programming approach for McKean–Vlasov controls with general semimartingales.
34 The barrier has been the lack of a general form of Itô's formula for flows of measures on
35 semimartingales.

36 *Our work.* We establish Itô's formula for flows of measures associated with general and pos-
37 sibly discontinuous semimartingales (Theorems 3.1 & 3.3). It enables us to derive dynamic
38 programming equations and verification theorems for McKean–Vlasov controls with jump dif-
39 fusions and for McKean–Vlasov mixed regular-singular control problems (Theorems 4.1 & 5.2).
40 It also allows for generalizing the classical relationship between the maximum principle and the
41 dynamic programming principle to the McKean–Vlasov mixed regular-singular control setting,
42 where the adjoint process is expressed in terms of the derivative of the value function with
43 respect to probability measures along the optimal path (Theorem 5.5).

44 Our approach starts by establishing Itô's formula on cylindrical functions [37]. Then, by
45 invoking a general form of Stone–Weierstrass theorem on compact sets in the joint Wasserstein

46 space and Euclidean space, this class of mean-field functions is shown to be dense in the class
47 of twice differentiable functions on the space of probability measures of order two with the
48 topology of $\mathcal{C}^{1,1}$. Finally, with an appropriate localization argument, the general form of Itô's
49 formula is established.

50 There are several key technical ingredients in our work. The first are the properties of
51 cylindrical functions. These are smooth mean-field functions with integrable forms, initially
52 studied in Fleming–Viot processes [23] for modeling population genetics and further developed
53 for analyzing a general class of probability measure-valued processes called polynomial diffu-
54 sions [19]. The second are linear derivatives on the space of probability measures [23, 19]; this
55 form of derivatives enables characterizing the behavior of both the jumps of flows of probability
56 measures and the jumps of semimartingale processes. This is in contrast to the Itô's formula
57 for flows of measures on diffusion processes, which involves only Lions derivatives [32]. Indeed,
58 Lions derivatives, obtained by identifying the Wasserstein space with the Hilbert space of ran-
59 dom variables on atomless probability space [8, 14], appear insufficient for characterizing the
60 infinitesimal changes on the functional of controlled McKean–Vlasov processes with the addition
61 of jumps; this has been observed before in [9].

62 *Comparison with related works.* In an independent work, which has been brought to our at-
63 tention during the review of this paper, [39] obtained a similar form of Itô's formula for flows
64 of measures on càdlàg semimartingales. There are two main differences between our work and
65 theirs. First, they use a time-discretization approach, which is different from our cylindrical
66 function technique. Secondly, the conditions for the Itô's formula in their work are different
67 from ours. They are essentially different sets of conditions needed for the dominated conver-
68 gence theorem to allow for the interchange of the expectation and integration; see Remark 3.14
69 for more detailed comparison.

70 *Cylindrical approach.* The cylindrical function technique appears natural for handling discontin-
71 uous semimartingales and is of independent interest. It has recently been used to obtain Itô's
72 formula for controlled measure-valued martingales [18, Theorem 5.1]. Cylindrical functions
73 have also been used to study viscosity solutions for controlled McKean–Vlasov dynamics [9].

74 Besides the cylindrical function technique and the time-discretization approach, it is natu-
75 rally of interest to explore for instance the particle approximation method. However, it appears
76 more difficult through this route for the case of discontinuous semimartingales, and we leave it
77 for future studies. (See Remark 3.15 for more technical discussions on the particle approximation
78 attempt).

79 *Outline of the paper.* Section 2 introduces the notation and recalls definitions of several forms
80 of derivatives in the Wasserstein space of probability measures. Section 3 is devoted to Itô's

81 formula and its variants for flows of probability measures for semimartingales. Section 4 studies
 82 McKean–Vlasov control with jump diffusions and Section 5 analyzes McKean–Vlasov mixed
 83 regular-singular controls.

84 2. Notations and Preliminaries

85 Throughout the paper, $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ is a fixed filtered probability space satisfying
 86 the usual conditions. That is, $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is complete and the filtration is right continuous.
 87 We assume that there exists a sub- σ -algebra \mathcal{G} of \mathcal{F} , with \mathcal{F} rich enough as will be explained
 88 shortly. Moreover, we will adopt the following notations, unless otherwise specified.

- 89 • Given any normed space $(E, |\cdot|)$, $\mathcal{P}(E)$ is the set of all probability measures on $(E, |\cdot|)$.
 90 For any $p \in \mathbb{N}$, $\mathcal{P}_p(E)$ is the set of all probability measures of p -th order on $(E, |\cdot|)$, defined
 91 as $\mathcal{P}_p(E) = \left\{ \mu \in \mathcal{P}(E) \mid \|\mu\|_p := \left(\int_E |x|^p \mu(dx) \right)^{\frac{1}{p}} < \infty \right\}$. For instance, $\mathcal{P}_2(E)$ is the set of
 92 all probability measures with finite second-order moment. For any probability measures
 93 μ, μ' in $\mathcal{P}_p(E)$, the p -th order Wasserstein distance on $\mathcal{P}_p(E)$ is defined as $W_p(\mu, \mu') =$
 94 $\inf_{\pi} \left(\int_{E \times E} |y - y'|^p \pi(dy, dy') \right)^{\frac{1}{p}}$, where π is a coupling of μ and μ' in the sense that $\pi \in$
 95 $\mathcal{P}(E \times E)$ with marginals μ and μ' . For any probability measure $\mu \in \mathcal{P}(E)$, $\text{supp}(\mu)$ is
 96 the support of μ . $L^2(E)$ is the space of all square-integrable functions $f : E \rightarrow \mathbb{R}$.
- 97 • $L^2_{\mu}(\mathbb{R}^d; \mathbb{R}^d)$ is the space of all Borel mappings $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\int_{\mathbb{R}^d} |f|^2 \mu(dx) < \infty$.
 98 $\nabla \mathcal{C}_c^{\infty}(\mathbb{R}^d) := \{\nabla \varphi : \varphi \in \mathcal{C}_c^{\infty}(\mathbb{R}^d)\}$, where $\mathcal{C}_c^{\infty}(\mathbb{R}^d)$ is the space of smooth real functions with
 99 compact support in \mathbb{R}^d . The closure of $\nabla \mathcal{C}_c^{\infty}(\mathbb{R}^d)$ in $L^2_{\mu}(\mathbb{R}^d; \mathbb{R}^d)$ is denoted as $\mathcal{T}_{\mu} \mathcal{P}_2(\mathbb{R}^d)$,
 100 which is called tangent space of $\mathcal{P}_2(\mathbb{R}^d)$ at a given measure $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.
- 101 • For vectors $a, b \in \mathbb{R}^d$, $d \in \mathbb{N}^+$, $a \cdot b = \sum_{i=1}^d a_i b_i$. For the matrix $C = (C_{ij})_{1 \leq i, j \leq d} \in \mathbb{R}^{d \times d}$,
 102 $\text{Tr}(C) = \sum_{i=1}^d C_{ii}$, the transpose of C is C^{\top} . Denote $\text{diag}(a_1, \dots, a_d)$ for the diagonal
 103 matrix in $\mathbb{R}^{d \times d}$ with diagonal entries $a_i \in \mathbb{R}$, $1 \leq i \leq d$.

104 We denote $L^2(\mathcal{F}; E)$ as the space of all E -valued square integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$;
 105 for any $\vartheta \in L^2(\mathcal{F}; E)$, we set $\|\vartheta\|_{L^2} = \mathbb{E}[|\vartheta|^2]^{\frac{1}{2}}$. We will assume that \mathcal{G} is "rich enough" in a sense
 106 that $\mathcal{P}_2(E) = \{\mathbb{P}_{\xi}; \xi \in L^2(\mathcal{G}; E)\}$, where \mathbb{P}_{ξ} denotes the law of ξ . This is satisfied whenever
 107 the probability space $(\Omega, \mathcal{G}, \mathbb{P})$ is atomless, see page 352, [13]. We may use $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ for an
 108 independent and identical copy of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\bar{\vartheta} \in L^2(\bar{\mathcal{F}}; E)$ for an
 109 independent and identical copy of a random variable $\vartheta \in L^2(\mathcal{F}; E)$.

110 *Differentiability of functions of probability measures.* We first recall two notions of differen-
 111 tiability of functionals with respect to probability measures that will be used throughout the
 112 paper. One is the Lions derivative from the identification of the Wasserstein space with the

113 Hilbert space of random variables on an atomless probability space ([32] and [13]), the other is
 114 the standard/ linear derivative on the Wasserstein space ([23], [19], and [13]).

The Lions derivative of a functional f , introduced in [32], is defined through the lift of f .
 The idea is to take the function $f : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, and let $\tilde{f} : L^2(\mathcal{F}; \mathbb{R}^d) \rightarrow \mathbb{R}$ be a lift of f such
 that $\tilde{f}(\vartheta) = f(\mathbb{P}_\vartheta)$ for any $\vartheta \in L^2(\mathcal{F}; \mathbb{R}^d)$. Then, \tilde{f} is differentiable in the Fréchet sense at ϑ_0 if
 there exists a linear continuous mapping $D\tilde{f}(\vartheta_0) : L^2(\mathcal{F}; \mathbb{R}^d) \rightarrow \mathbb{R}$ such that

$$\tilde{f}(\vartheta) - \tilde{f}(\vartheta_0) = \mathbb{E}[D\tilde{f}(\vartheta_0) \cdot (\vartheta - \vartheta_0)] + o(\|\vartheta - \vartheta_0\|_{L^2}),$$

115 as $\|\vartheta - \vartheta_0\|_{L^2} \rightarrow 0$. It has been shown that when \tilde{f} is the lift of a function f in $\mathcal{P}_2(\mathbb{R}^d)$, the law
 116 of $D\tilde{f}(\vartheta_0)$ depends on ϑ_0 only via its law \mathbb{P}_{ϑ_0} , and

$$D\tilde{f}(\vartheta_0) = h_0(\vartheta_0), \tag{2.1}$$

117 for some Borel function $h_0 : \mathbb{R}^d \rightarrow \mathbb{R}^d$. (See e.g., [13], Chapter 5 and [26]). The Lions derivative
 118 is thus well defined:

119 **Definition 2.1** *f is differentiable at $\mu_0 = \mathbb{P}_{\vartheta_0} \in \mathcal{P}_2(\mathbb{R}^d)$ if its lift function \tilde{f} is Fréchet differ-*
 120 *entiable at ϑ_0 ; and in this case, the function h_0 in (2.1) is called the Lions derivative of f at*
 121 *μ_0 , and denoted as $(\partial_\mu f)(\mu_0, \cdot)$.*

122 The definition of the linear derivative on the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$ ([23], [19] and [13],
 123 Chapter 5) is more straightforward:

124 **Definition 2.2** *For $f : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, the linear derivative of f at μ is a function $\frac{\delta f}{\delta \mu}$ on $\mathcal{P}_2(\mathbb{R}^d)$*
 125 *$\times \mathbb{R}^d$ such that for every $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$*

$$f(\mu) - f(\mu') = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta f}{\delta \mu}(h\mu + (1-h)\mu', x)(\mu - \mu')(dx)dh.$$

126 Clearly, the linear derivative $\frac{\delta f}{\delta \mu}$ is defined up to an additive constant. It is very closely
 127 related to the Gâteaux derivative, denoted by $\frac{\partial f}{\partial \mu}$, which usually assumes the existence of the
 128 density function for a probability measure (see [6]).

129 Under suitable regularity conditions on f , the Lions derivative in Definition 2.1, and the
 130 linear derivative and the Gâteaux derivative in Definition 2.2 are connected through the following
 131 equation, according to Proposition 5.48 of [13],

$$\partial_\mu f(\mu, x) = \partial_x \frac{\delta f}{\delta \mu}(\mu, x) = \partial_x \frac{\partial f}{\partial \mu}(\mu, x). \tag{2.2}$$

132 For instance, consider the linear $\Phi(\mu) = \int_{\mathbb{R}^d} g(x)\mu(dx) =: \langle g, \mu \rangle$, where the derivative of g has
 133 linear growth, then $\partial_\mu \Phi(\mu, x) = \partial_x g(x) = \partial_x \frac{\delta \Phi}{\delta \mu}(\mu, x)$.

134 **Remark 2.3** *In addition to the Lions derivative and the linear derivative, there is an intrinsic*
 135 *notion of derivative on the Wasserstein space used in optimal transport and gradient flows ([1],*
 136 *[26]). Precisely, one says that $f : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is differentiable at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists a*
 137 *unique $\xi \in \mathcal{T}_\mu \mathcal{P}_2(\mathbb{R}^d)$, such that*

$$\lim_{n \rightarrow \infty} \frac{f(\mu_n) - f(\mu) - \int_{\mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)} \xi(x)(y-x)\pi_n(dx, dy)}{W_2(\mu, \mu_n)} = 0$$

138 *for any sequence $\{\mu_n\}_n \subset \mathcal{P}_2(\mathbb{R}^d)$ such that 2-Wasserstein distance $W_2(\mu_n, \mu) \rightarrow 0$ and any*
 139 *sequence of optimal plans $\{\pi_n\}_n \subset \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ_n, μ such that $W_2(\mu_n, \mu) =$
 140 $\left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi_n(dx, dy) \right)^{\frac{1}{2}}$ for any n . Such ξ is called Wasserstein gradient and denoted as
 141 $\nabla_w f(\mu)$. It is shown in [26] that Wasserstein gradient $\nabla_w f(\mu)$ is equivalent to Lions derivative
 142 $\partial_\mu f(\mu, \cdot)$ in Definition 2.1.*

143 Our Itô's formula (Theorem 3.1) will be first established on the space of differential functions
 144 $\mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$ in $\mathcal{P}_2(\mathbb{R}^d)$, first introduced in [8], [14].

145 **Definition 2.4** *We say a function $f \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, if its lift \tilde{f} is Fréchet differentiable, and*
 146 *if there exists a continuous version of $\partial_\mu f(\mu, x)$ such that*

- 147 • *the mapping $(\mu, x) \mapsto \partial_\mu f(\mu, x)$ is jointly continuous with respect to (μ, x) and there is a*
 148 *constant $C > 0$ such that*

$$|\partial_\mu f(\mu, x)| \leq C, \tag{2.3}$$

149 *for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and any $x \in \mathbb{R}^d$;*

- 150 • *for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, the mapping $x \mapsto \partial_\mu f(\mu, x)$ is continuously differentiable. Its deriva-*
 151 *tive, denoted by $\partial_x \partial_\mu f(\mu, x)$, is jointly continuous with respect to (μ, x) ; and there is a*
 152 *constant $C > 0$ such that*

$$|\partial_x \partial_\mu f(\mu, x)| \leq C, \tag{2.4}$$

153 *for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and any $x \in \mathbb{R}^d$.*

154 We also introduce an alternative class $\mathcal{C}_{relax}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$ of differential functions where we relax
 155 the boundedness conditions in $\mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$ by square integrability growth conditions, in view
 156 of Itô's formula in Theorem 3.3.

157 **Definition 2.5** *We say a function $f \in \mathcal{C}_{relax}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, if its lift \tilde{f} is Fréchet differentiable, and*
 158 *if there exists a continuous version of $\partial_\mu f(\mu, x)$ such that*

- 159 • *the mapping $(\mu, x) \mapsto \partial_\mu f(\mu, x)$ is jointly continuous with respect to (μ, x) ;*
- 160 • *for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, the mapping $x \mapsto \partial_\mu f(\mu, x)$ is continuously differentiable. Its deriva-*
 161 *tive, denoted by $\partial_x \partial_\mu f(\mu, x)$, is jointly continuous with respect to (μ, x) ;*

162 • for any compact set \mathcal{K} in $\mathcal{P}_2(\mathbb{R}^d)$,

$$\sup_{\mu \in \mathcal{K}} \left[\int_{\mathbb{R}^d} \left\{ |\partial_\mu f(\mu, x)|^2 + |\partial_x \partial_\mu f(\mu, x)|^2 \right\} \mu(dx) \right] < \infty. \quad (2.5)$$

163 Note that in the above Definitions, the continuity in μ is with respect to the Wasserstein
 164 distance W_2 . Moreover, if $f \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$ or $\mathcal{C}_{relax}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, then the linear derivative $\frac{\delta f}{\delta \mu}$ in
 165 Definition 2.2 exists and is jointly continuous with respect to $(\mu, x) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$.

166 3. Itô's Formula for Flows of Measures on Semimartingales

167 3.1. Itô's Formula and Several Variants

168 In this section, we will present Itô's formula and several of its variants for flows of measures
 169 on a class of semimartingales $X = (X_t)_{0 \leq t \leq T}$ ($T > 0$), which are square integrable and satisfy
 170 the **(H)** condition, following the notation in [37]. This condition is similar to the integrability
 171 conditions on the drift and diffusion coefficients imposed for Itô's diffusion processes in [8] and
 172 [14].

173 **(H)** There exists one decomposition $X = X(0) + V + L$, where V is càdlàg process of finite
 174 variation with $V(0) = 0$, and L is a local martingale with $L(0) = 0$ such that

$$\mathbb{E}[\text{Var}(V)_T] < \infty, \quad \mathbb{E} \left[\sum_{0 < t \leq T} |\Delta X_t| \right] < \infty, \quad \mathbb{E}[[X, X]_T^c] < \infty. \quad (3.1)$$

175 Here $\text{Var}(V)_T$ denotes the variation of the process V on $[0, T]$, X_{t-} the left limit of X at t ,
 176 $\Delta X_t = X_t - X_{t-}$ the jump of X at t , $[X, X]$ the quadratic variation of X , and $[X, X]_t^c =$
 177 $[X^c, X^c] = [X, X]_t - \sum_{0 \leq r \leq t} |\Delta X_r|^2$ the continuous part of $[X, X]$, with X^c the continuous
 178 part of X .

179 **Theorem 3.1 (Itô's formula)** *Given a semimartingale X satisfying **(H)** and $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow$
 180 \mathbb{R} . If $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, then for any $0 \leq t < s \leq T$,*

$$\begin{aligned} & \Phi(\mathbb{P}_{X_s}) - \Phi(\mathbb{P}_{X_t}) \\ &= \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \cdot dX_r + \frac{1}{2} \text{Tr} \left(\partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right) \right. \\ & \quad + \sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \cdot \Delta X_r \right. \\ & \quad \left. \left. + \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right\} \right]. \end{aligned} \quad (3.2)$$

181 **Remark 3.2** *Note that when $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, the jump remainder in the RHS of (3.2) is
 182 finite. That is,*

$$\begin{aligned} & \sum_{t < r \leq s} \left| \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right| < \infty, \quad \mathbb{E} \left[\left| \sum_{t < r \leq s} \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \cdot \Delta X_r \right| \right] < \infty, \\ & \mathbb{E} \left[\sum_{t < r \leq s} \left| \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right| \right] < \infty. \end{aligned}$$

183 *Indeed, by the definition of the Lions derivative in Definition 2.1 and its relation with the linear*
 184 *derivative (2.2), we see*

$$\begin{aligned}\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) &= \int_0^1 \frac{d}{dh} \Phi(\mathbb{P}_{X_{r-} + h\Delta X_r}) dh \\ &= \int_0^1 \mathbb{E}[\partial_\mu \Phi(\mathbb{P}_{X_{r-} + h\Delta X_r}, X_{r-} + h\Delta X_r) \cdot \Delta X_r] dh, \quad (3.3)\end{aligned}$$

$$\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) = \int_0^1 \partial_\mu \Phi(\mathbb{P}_{X_r}, X_{r-} + h\Delta X_r) \cdot \Delta X_r dh. \quad (3.4)$$

185 *Thus, by (3.3)-(3.4) (or the Lipschitz continuity of Φ) and (2.3),*

$$\left| \Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right| \leq C \mathbb{E}[|\Delta X_r|], \quad \left| \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right| \leq C |\Delta X_r|. \quad (3.5)$$

186 *The claim is now clear by (3.1) from assumption **(H)** and (3.5).*

187 Under such alternative conditions on the functional Φ and the semimartingale X , one can
 188 still derive the same form of Itô's formula (3.2). For instance, one can relax the condition on Φ
 189 and instead impose a slightly stronger integrability condition on the semimartingale X .

190 **(H)_{strict}** There exists one decomposition $X = X(0) + V + L$, where V is càdlàg process of
 191 finite variation with $V(0) = 0$, and L is a local martingale with $L(0) = 0$ such that

$$\mathbb{E}[|\text{Var}(V)_T|^2] < \infty, \quad \mathbb{E}\left[\left(\sum_{0 < t \leq T} |\Delta X_t|\right)^2\right] < \infty, \quad \mathbb{E}[|[X, X]_T^c|^2] < \infty, \quad (3.6)$$

192 where $\text{Var}(V)_T$, X_{t-} , ΔX_t , $[X, X]$ and $[X, X]_t^c$ are given in **(H)**.

193 **Theorem 3.3 (Itô's formula (II))** *Given a semimartingale X satisfying **(H)_{strict}** and a*
 194 **functional $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ in $\mathcal{C}_{relax}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, Itô's formula (3.2) holds.**

195 In the special case when the semimartingale X is continuous, we have the following corollary,
 196 recovering earlier results from [13] and [14].

197 **Corollary 3.4** *Suppose that the semimartingale X is continuous (and hence \mathbb{P}_{X_t} is continuous*
 198 *in time), and that X satisfies*

$$\mathbb{E}[|\text{Var}(V)_T|^2] < \infty, \quad \mathbb{E}[|[X, X]_T^c|^2] < \infty.$$

199 *Let Φ be a function from $\mathcal{P}_2(\mathbb{R}^d)$ into \mathbb{R} satisfying*

$$\sup_{\mu \in \mathcal{K}} \left[\int_{\mathbb{R}^d} |\partial_x \partial_\mu \Phi(\mu, x)|^2 \mu(dx) \right] < +\infty,$$

200 *for any compact set $\mathcal{K} \subset \mathcal{P}_2(\mathbb{R}^d)$. Then, for all $0 \leq t < s \leq T$,*

$$\Phi(\mathbb{P}_{X_s}) - \Phi(\mathbb{P}_{X_t}) = \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \cdot dX_r + \frac{1}{2} \text{Tr} \left(\partial_x \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) d[X, X]_r^c \right) \right].$$

201 To avoid confusion, for the rest of the paper, we will always refer to the first version of Itô's
 202 formula in Theorem 3.1 unless otherwise specified.

203 To extend Itô's formula to the time-space-measure-dependent case, let us first define the
 204 space $\mathcal{C}^{1,2,(1,1)}([0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ of continuous functions Φ on $[0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ such that

- 205 • $\partial_t \Phi(t, x, \mu)$, $\partial_x \Phi(t, x, \mu)$ and $\partial_{xx} \Phi(t, x, \mu)$ exist and are jointly continuous with respect to
 206 $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$;
- 207 • both $\partial_\mu \Phi(t, x, \mu, x')$ and $\partial_{x'} \partial_\mu \Phi(t, x, \mu, x')$ exist and are jointly continuous with respect
 208 to $(t, x, \mu, x') \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$; moreover, there exists $C > 0$ such that
 209 $|\partial_\mu \Phi(t, x, \mu, x')| \leq C$ and $|\partial_{x'} \partial_\mu \Phi(t, x, \mu, x')| \leq C$.

210 **Corollary 3.5 (Time-space-measure-dependent Itô's formula)** *Let X be a semimartin-*
 211 *gale satisfying (H) (with $t \mapsto \mathbb{P}_{X_t}$ not necessarily continuous). For any $\Phi \in \mathcal{C}^{1,2,(1,1)}([0, T] \times$
 212 $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, we have for $0 \leq t < s \leq T$:*

$$\begin{aligned}
 & \Phi(s, X_s, \mathbb{P}_{X_s}) = \Phi(t, X_t, \mathbb{P}_{X_t}) + \int_t^s \partial_r \Phi(r, X_r, \mathbb{P}_{X_r}) dr \\
 & + \mathbb{E} \left[\int_t^s \partial_\mu \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) d\bar{X}_r + \frac{1}{2} \text{Tr} \left(\partial_x \partial_\mu \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) d[\bar{X}, \bar{X}]_r^c \right) \right. \\
 & + \left. \sum_{t < r \leq s} \left\{ \left(\Phi(r, X_{r-}, \mathbb{P}_{X_r}) - \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) \cdot \Delta \bar{X}_r \right\} \right] \\
 & + \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(r, X_{r-}, \mathbb{P}_{X_r}, \bar{X}_r) - \frac{\delta \Phi}{\delta \mu}(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right] \\
 & + \sum_{i=1}^d \int_t^s \partial_{x_i} \Phi(r, X_{r-}, \mathbb{P}_{X_r}) dX_r^i + \frac{1}{2} \sum_{i,j=1}^d \int_t^s \partial_{x_i} \partial_{x_j} \Phi(r, X_{r-}, \mathbb{P}_{X_r}) d[X^i, X^j]_r^c \\
 & + \sum_{t < r \leq s} \left\{ \Phi(r, X_r, \mathbb{P}_{X_r}) - \Phi(r, X_{r-}, \mathbb{P}_{X_r}) - \sum_{i=1}^d \partial_{x_i} \Phi(r, X_{r-}, \mathbb{P}_{X_r}) \Delta X_r^i \right\}.
 \end{aligned}$$

213 Here \bar{X} denotes an independent and identical copy of X on a copy of the probability space
 214 $(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. To see how it follows from Theorem 3.1, fix an $m \in \mathbb{N}$. Let $\pi_{t,s}^m = (t = t_0^m < t_1^m < \dots < t_{m+1}^m = s)$ be any partition of $[t, s]$.

$$\begin{aligned}
 & \Phi(s, X_s, \mathbb{P}_{X_s}) - \Phi(t, X_t, \mathbb{P}_{X_t}) \\
 & = \sum_{j=0}^m \left\{ \Phi(t_{j+1}^m, X_{t_{j+1}^m}, \mathbb{P}_{X_{t_{j+1}^m}}) - \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_j^m}}) \right\} \\
 & = \sum_{j=0}^m \left\{ \Phi(t_{j+1}^m, X_{t_{j+1}^m}, \mathbb{P}_{X_{t_{j+1}^m}}) - \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_{j+1}^m}}) + \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_{j+1}^m}}) - \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_j^m}}) \right\} \\
 & = \sum_{j=0}^m \left\{ \Phi(t_{j+1}^m, X_{t_{j+1}^m}, \mathbb{P}_{X_{t_{j+1}^m}}) - \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_{j+1}^m}}) \right\} + \sum_{j=0}^m \left\{ \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_{j+1}^m}}) - \Phi(t_j^m, X_{t_j^m}, \mathbb{P}_{X_{t_j^m}}) \right\}.
 \end{aligned}$$

Let $m \rightarrow \infty$, then, similar as proving the classical Itô's formula for semimartingales, we see that the first sum converges in probability to

$$\int_t^s \partial_r \Phi(r, X_r, \mathbb{P}_{X_r}) dr + \sum_{i=1}^d \int_t^s \partial_{x_i} \Phi(r, X_{r-}, \mathbb{P}_{X_r}) dX_r^i + \frac{1}{2} \sum_{i,j=1}^d \int_t^s \partial_{x_i x_j}^2 \Phi(r, X_{r-}, \mathbb{P}_{X_r}) d[X^i, X^j]_r^c$$

$$+ \sum_{t < r \leq s} \left\{ \Phi(r, X_r, \mathbb{P}_{X_r}) - \Phi(r, X_{r-}, \mathbb{P}_{X_r}) - \sum_{i=1}^d \partial_{x_i} \Phi(r, X_{r-}, \mathbb{P}_{X_r}) \Delta X_r^i \right\}.$$

According to Theorem 3.1, the second sum converges to

$$\bar{\mathbb{E}} \left[\int_t^s \partial_\mu \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) \cdot d\bar{X}_r + \int_t^s \text{Tr} \left(\partial_x \partial_\mu \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) d[\bar{X}, \bar{X}]_r^c \right) \right]$$

$$+ \sum_{t < r \leq s} \left(\Phi(r, X_{r-}, \mathbb{P}_{X_r}) - \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \bar{\mathbb{E}} \left[\sum_{t < r \leq s} \partial_\mu \Phi(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) \cdot \Delta \bar{X}_r \right]$$

$$+ \bar{\mathbb{E}} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_r) - \frac{\delta \Phi}{\delta \mu}(r, X_{r-}, \mathbb{P}_{X_{r-}}, \bar{X}_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right].$$

215

□

216 Itô's formula in Theorem 3.1 takes some special forms when applied to a class of semimartin-
217 gales driven by the McKean–Vlasov SDEs or jump diffusion processes which are useful for the
218 subsequent analysis of McKean–Vlasov controls in Sections 4 and 5.

219 To see this, first let $b = (b_s)_{s \geq 0}$ and $\sigma = (\sigma_s)_{s \geq 0}$ be \mathbb{F} -adapted processes valued in \mathbb{R}^d and
220 $\mathbb{R}^{d \times d}$, respectively, and $\beta = (\beta_s(\theta))_{s \geq 0}$ be \mathbb{F} -predictable processes valued in \mathbb{R}^d for $\theta \in \mathbb{R}^q$.
221 Consider the following jump-diffusion process $(X_s)_{s \geq 0}$

$$dX_s = b_s ds + \sigma_s dW_s + \int_{\theta \in \mathbb{R}^q} \beta_s(\theta) N(ds, d\theta), \quad X_{t-} = \xi \in L^2(\mathcal{F}_t; \mathbb{R}^d), \quad (3.7)$$

where $(W_s)_{s \geq 0}$ is Brownian motion, and $N(ds, d\theta)$ is a Poisson random measure with a finite
intensity measure ν , with its compensated Poisson random measure $\tilde{N}(ds, d\theta) := N(ds, d\theta) -$
 $\nu(d\theta)ds$. It is easy to check that such $(X_s)_{s \geq t}$ satisfies assumption **(H)** if

$$\mathbb{E} \left[\int_0^T \left(|b_s|^2 + |\sigma_s|^2 + \int_{\mathbb{R}^q} |\beta_s(\theta)| \nu(d\theta) \right) ds \right] < \infty.$$

222 Note that such process X has jumps induced by the Poisson random measure, whereas \mathbb{P}_X is
223 continuous in time. (See, for instance [27]).

224 **Corollary 3.6 (Itô's formula for jump process)** *Given a semimartingale X defined by (3.7),*
225 *then for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$, we have*

$$\Phi(\mathbb{P}_{X_s}) = \Phi(\mathbb{P}_\xi) + \int_t^s \mathbb{E} \left[\partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \cdot b_r + \frac{1}{2} \text{Tr} \left(\sigma_r \sigma_r^\top \partial_x \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \right) \right. \\ \left. + \int_{\mathbb{R}^q} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r + \beta_r(\theta)) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) \right) \nu(d\theta) \right] dr,$$

226 for all $0 \leq t < s \leq T$.

227 **Proof.** Apply Theorem 3.1 to $\Phi(\mathbb{P}_{X_r})$ with X_r given in (3.7) between t and s ,

$$\begin{aligned} \Phi(\mathbb{P}_{X_s}) - \Phi(\mathbb{P}_{X_{t-}}) &= \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \cdot b_r + \frac{1}{2} \text{Tr} \left(\sigma_r \sigma_r^\top \partial_x \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \right) dr \right. \\ &\quad + \sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right. \\ &\quad \left. \left. + \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right\} \right]. \end{aligned} \quad (3.8)$$

228 Now it suffices to compute the jump term in (3.8). As the distribution \mathbb{P}_{X_r} is continuous in
229 time, we have

$$\sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right\} = 0.$$

230 Note that at the time r , X_{r-} has a jump $\beta_r(\theta)$ with Poisson random measure $N(dr, d\theta)$. Thus,

$$\begin{aligned} &\mathbb{E} \left[\sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right. \right. \\ &\quad \left. \left. + \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right\} \right] \\ &= \mathbb{E} \left[\int_t^s \int_{\theta \in \mathbb{R}^q} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r + \beta_r(\theta)) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) \right) \nu(d\theta) dr \right]. \end{aligned}$$

231

□

232 *Relation to Fokker–Planck equation.* One can check that this form of Itô’s formula for (3.7) is
233 consistent with its Fokker–Planck equation. To see this, suppose that \mathbb{P}_{X_t} has a PDF p_t such
234 that $b_t = b(X_t, p_t)$, $\sigma_t = \sigma(X_t, p_t)$, $\beta_t(\theta) = \theta$, and N is a compound Poisson Process with
235 intensity $\lambda \in \mathbb{R}^d$ and jump PDF γ , then under mild conditions, the time evolution of p_t can be
236 prescribed by an integral differential equation

$$\begin{aligned} \partial_t p_t &= -\nabla \cdot (p_t b(x, p_t)) + \frac{1}{2} \nabla \cdot \nabla \cdot (\sigma(x, p_t) \sigma(x, p_t)^\top p_t) \\ &\quad + \lambda \left(\int_{\mathbb{R}^d} p_t(\theta) \gamma(x - \theta) d\theta - p_t(x) \right), \end{aligned} \quad (3.9)$$

237 where $\nabla \cdot$ denotes the divergence. One can further obtain the time evolution of $\phi(p_t)$ for a
238 smooth function $\phi : L^2(\mathbb{R}^d) \rightarrow \mathbb{R}$ with respect to the Gâteaux derivative ([6]),

$$\begin{aligned} d\phi(p_t) &= \int_{\mathbb{R}^d} \frac{\partial \phi}{\partial p}(p_t, x) \partial_t p_t dx dt \\ &= \mathbb{E} \left[\partial_x \frac{\partial \phi}{\partial p}(p_t, x) \cdot b(X_t, p_t) + \frac{1}{2} \text{Tr} \left(\sigma \sigma^\top(X_t, p_t) \partial_{xx} \frac{\partial \phi}{\partial p}(p_t, X_t) \right) \right. \\ &\quad \left. + \lambda \int_{\mathbb{R}^d} \left(\frac{\partial \phi}{\partial p}(p_t, X_t + \theta) - \frac{\partial \phi}{\partial p}(p_t, X_t) \right) \gamma(\theta) d\theta \right] dt, \end{aligned}$$

239 where the second inequality is by (3.9) and from integration by parts. Now the consistency
240 follows from the connection (2.2) between the Gâteaux derivative and the linear derivative in
241 Definition 2.2.

242 Next, let us consider processes that arise from control problems of singular type. Let (η_t)
 243 $= (\eta_t^i)_{1 \leq i \leq d}$ be finite variation processes valued in \mathbb{R}^d , and denote $\lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$ with λ_i
 244 nonnegative constant in \mathbb{R} . Take the following càdlàg process $X = (X_s^1, \dots, X_s^d)_{s \geq t}$,

$$dX_s = b_s ds + \sigma_s dW_s + \lambda d\eta_s, \quad s \geq t, \quad X_{t-} = \xi \in L^2(\mathcal{F}_t; \mathbb{R}^d), \quad (3.10)$$

245 where $(W_s)_{s \geq 0}$ is a d -dimensional Brownian motion. Assuming that $\mathbb{E}[\text{Var}(\eta)_T] < \infty$ and
 246 $\mathbb{E}[\int_0^T (|b_s|^2 + |\sigma_s|^2) ds] < \infty$, one can easily check that X satisfies assumption **(H)**, and we then
 247 have,

248 **Corollary 3.7** *Given a semimartingale X in (3.10) (with $t \mapsto \mathbb{P}_{X_t}$ not necessarily continuous).*

249 *For any $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$,*

$$\begin{aligned} & \Phi(\mathbb{P}_{X_s}) - \Phi(\mathbb{P}_\xi) \\ = & \mathbb{E} \left[\int_t^s \left\{ \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \cdot b_r + \frac{1}{2} \text{Tr} \left(\partial_x \partial_\mu \Phi(\mathbb{P}_{X_r}, X_r) \sigma_r \sigma_r^\top \right) \right\} dr \right. \\ & + \sum_{i=1}^d \int_t^s \lambda_i \partial_\mu \Phi_i(\mathbb{P}_{X_{r-}}, X_{r-}) d\eta_r^i \\ & + \sum_{t \leq r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \sum_{i=1}^d \lambda_i \partial_\mu \Phi_i(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta \eta_r^i \right\} \\ & \left. + \mathbb{E} \left[\sum_{t \leq r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right], \quad 0 \leq t \leq s \leq T. \right. \end{aligned}$$

250 3.2. Proof of Theorem 3.1

251 We will establish Itô's formula for a class of cylindrical functions called $\mathcal{G}(\mathcal{M})$. We will
 252 then prove the general case by applying the general form of Stone–Weierstrass theorem on
 253 compact sets in the joint Wasserstein space and Euclidean space and by appropriate localization
 254 argument. For ease of exposition, we will prove for the case of $d = 1$ without loss of generality.
 255 We will discuss (see Remark 3.13) how to adapt the arguments to the general case of $d > 1$.

256 Throughout the proof of Theorem 3.1, we adopt a generic constant C for ease of exposition,
 257 unless otherwise specified.

258 3.2.1. Proof of Theorem 3.1 for $\mathcal{G}(\mathcal{K}_K)$

259 Let us first define the general space of cylindrical functions $\mathcal{G}(\mathcal{M})$ for $\mathcal{M} \subset \mathcal{P}_2(\mathbb{R})$.

Definition 3.8 *Given $\mathcal{M} \subset \mathcal{P}_2(\mathbb{R})$, define*

$$\begin{aligned} \mathcal{G}(\mathcal{M}) := & \{ \mathcal{M} \ni \mu \mapsto \Phi(\mu) = f(\langle g_1, \mu \rangle, \dots, \langle g_n, \mu \rangle) \text{ for some } n \in \mathbb{N}, \\ & \text{and polynomial } f : \mathbb{R}^n \rightarrow \mathbb{R} \text{ and } g_1, \dots, g_n : \mathbb{R} \rightarrow \mathbb{R} \text{ polynomials} \}. \end{aligned} \quad (3.11)$$

Observe that $\mathcal{G}(\mathcal{M})$ is an algebra as it is closed under pointwise addition, multiplication, and scalar multiplication. Moreover, $\mathcal{G}(\mathcal{M})$ can be rewritten in the following form

$$\mathcal{G}(\mathcal{M}) = \text{span}\{\langle g, \mu^n \rangle : n \geq 1, \mu \in \mathcal{M}, g : \mathbb{R}^n \rightarrow \mathbb{R} \text{ is monomial}\},$$

260 where $\langle g, \mu^n \rangle := \int_{\mathbb{R}^k} g(x_1, \dots, x_n) \mu(dx_1) \dots \mu(dx_n)$ with $n \geq 1$ is a basis of $\mathcal{G}(\mathcal{M})$, as each
 261 $\Phi \in \mathcal{G}(\mathcal{M})$ is a linear combination of monomials $\langle g, \mu^k \rangle$. (See [19] for more discussions of this
 262 function space).

Now, consider the cylindrical function $\Phi(\mu) = f(\langle g_1, \mu \rangle, \dots, \langle g_n, \mu \rangle)$, whose Lions derivative is given by

$$\partial_\mu \Phi(\mu, x) = \sum_{k=1}^n \partial_{y_k} f(\langle g_1, \mu \rangle, \dots, \langle g_n, \mu \rangle) \partial_x g_k(x).$$

263 Note, however, the Lions derivative and the mixed second order derivative of cylindrical func-
 264 tions are not necessarily in $\mathcal{G}(\mathcal{M})$. Instead they belong to a bigger algebraic space defined
 265 below.

Definition 3.9 Given $\mathcal{M} \subset \mathcal{P}_2(\mathbb{R})$ and $U \subset \mathbb{R}$. Let

$$\mathcal{H}(\mathcal{M} \times U) := \{\mathcal{M} \times U \ni (\mu, x) \mapsto \Phi(\mu, x) = \sum_{k=1}^n f_k(\langle g_k, \mu^k \rangle) h_k(x), \text{ for some } n \in \mathbb{N},$$

$$f_k, h_k : \mathbb{R} \rightarrow \mathbb{R} \text{ and } g_k : \mathbb{R}^k \rightarrow \mathbb{R} \text{ are monomials}\}.$$

One can see that $\mathcal{H}(\mathcal{M} \times U)$ is also an algebra and can be rewritten in the following form

$$\mathcal{H}(\mathcal{M} \times U) = \text{span}\left\{\langle g, \mu^k \rangle x^l : k \geq 1, l \geq 0, \mu \in \mathcal{M}, x \in U, g : \mathbb{R}^k \rightarrow \mathbb{R} \text{ is monomial}\right\}.$$

266 Moreover, $\mathcal{G}(\mathcal{M})$ can be viewed as a subalgebra of $\mathcal{H}(\mathcal{M} \times U)$.

267 We will next consider a particular choice of $\mathcal{G}(\mathcal{M})$ with $\mathcal{M} := \mathcal{K}_K$, and

$$\mathcal{K}_K := \left\{ \mu \in \mathcal{P}_2(\mathbb{R}) \mid \text{supp}(\mu) \subset [-K, K] \right\}.$$

268 Clearly \mathcal{K}_K is not empty since the Dirac measure δ_0 is in \mathcal{K}_K . Moreover, by Lemma 5.7 and
 269 Proposition 5.3 in [10], \mathcal{K}_K is compact under the W_p distance for any $p \geq 1$. Now we can
 270 establish Itô's formula for any $\Phi \in \mathcal{G}(\mathcal{K}_K)$.

271 **Lemma 3.10** Given a semimartingale X satisfying assumption **(H)** and $|X_t| \leq K$ \mathbb{P} -a.s., for
 272 some $K > 0$ and any $t \in [0, T]$. Then Itô's lemma in the form of equation (3.2) holds for
 273 $\Phi(\mathbb{P}_{X_t})$, with $\Phi \in \mathcal{G}(\mathcal{K}_K)$.

274 **Proof.** Given $\Phi \in \mathcal{G}(\mathcal{K}_K)$, $\Phi(\mu) = f(\langle g_1, \mu \rangle, \dots, \langle g_n, \mu \rangle)$ by (3.11), where f and g_k , $1 \leq k \leq n$
 275 are polynomials. Therefore,

$$\Phi(\mathbb{P}_{X_r}) = f(\mathbb{E}[g_1(X_r)], \dots, \mathbb{E}[g_n(X_r)]),$$

where the process $r \mapsto X_r$ is a semimartingale and $|X_r| \leq K$ \mathbb{P} -a.s. Clearly, if we define $Y_r^k = \mathbb{E}[g_k(X_r)]$, $k = 1, \dots, n$ for any $r \in [0, T]$, then $\Phi(\mathbb{P}_{X_r}) = f(Y_r^1, \dots, Y_r^n)$. Since X_r is a semimartingale and g_k is polynomial, by the classical Itô's formula for semimartingale, $g_k(X_r)$ and (hence) $Y_r^k = \mathbb{E}[g_k(X_r)]$ are bounded semimartingales. Now, setting $g = (g_1, \dots, g_n)$ and $Y_r = (Y_r^1, \dots, Y_r^n)$, $0 \leq r \leq T$, and applying Itô's formula to $f(Y_r)$ between t and s , $0 \leq t < s \leq T$, we see

$$\begin{aligned}
\Phi(\mathbb{P}_{X_s}) - \Phi(\mathbb{P}_{X_t}) &= f(Y_s) - f(Y_t) \\
&= \underbrace{\sum_{k=1}^n \int_t^s \partial_{y_k} f(Y_{r-}) dY_r^k}_I + \underbrace{\frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n \int_t^s \partial_{y_k y_j} f(Y_{r-}) d[Y^j, Y^k]_r^c}_{II} \\
&\quad + \underbrace{\sum_{t < r \leq s} \left\{ f(Y_r) - f(Y_{r-}) - \sum_{k=1}^n \partial_{y_k} f(Y_{r-}) \Delta Y_r^k \right\}}_{III}.
\end{aligned} \tag{3.12}$$

276 Let us now compute the terms I , II , and III separately.

277 Applying the classical Itô's formula to $g_k(X_r)$, $1 \leq k \leq n$, we have

$$\begin{aligned}
g_k(X_s) - g_k(X_t) &= \int_t^s \partial_x g_k(X_{r-}) dX_r + \frac{1}{2} \partial_{xx} g_k(X_{r-}) d[X, X]_r^c \\
&\quad + \sum_{t < r \leq s} \left\{ g_k(X_r) - g_k(X_{r-}) - \partial_x g_k(X_{r-}) \Delta X_r \right\}.
\end{aligned} \tag{3.13}$$

278 Since $|X_r| \leq K$, \mathbb{P} -a.s., there exists constant $C > 0$ such that \mathbb{P} -a.s.

$$\max \left(|g_k(X_r)|, |\partial_x g_k(X_{r-})|, |\partial_{xx} g_k(X_{r-})| \right) \leq C.$$

279 Taking expectation on both sides of (3.13), we see

$$\begin{aligned}
Y_s^k - Y_t^k &= \mathbb{E} \left[\int_t^s \partial_x g_k(X_{r-}) dX_r + \frac{1}{2} \partial_{xx} g_k(X_{r-}) d[X, X]_r^c \right. \\
&\quad \left. + \sum_{t < r \leq s} \left\{ g_k(X_r) - g_k(X_{r-}) - \partial_x g_k(X_{r-}) \Delta X_r \right\} \right].
\end{aligned} \tag{3.14}$$

280 By the definition of the integral,

$$I := \sum_{k=1}^n \int_t^s \partial_{y_k} f(Y_{r-}) dY_r^k = \lim_{m \rightarrow \infty} \sum_{\pi_{t,s}} \sum_{k=1}^n \partial_{y_k} f(Y_{t_j^m}) (Y_{t_{j+1}^m}^k - Y_{t_j^m}^k),$$

281 holds for any arbitrary partition $\pi_{t,s}^m$ of $[t, s]$. Therefore, in order to calculate the term I , we may

282 take a partition $\bar{\pi}_{t,s}^m = (t = t_0^m < t_1^m < \dots < t_{m+1}^m = s)$, $m \in \mathbb{N}$ such that $\max_j |t_j^m - t_{j+1}^m| \rightarrow 0$

283 and $\max_j \sup_{r \in (t_j^m, t_{j+1}^m)} |Y_{r-} - Y_{t_j^m}| \rightarrow 0$ as $m \rightarrow \infty$. The existence of $\bar{\pi}_{t,s}^m$ is ensured by the fact

284 that $(Y_r)_{t \leq r \leq s}$ is a càdlàg function of finite variation. Therefore, I is well-defined and can be

285 approximated by the partition $\bar{\pi}_{t,s}^m$:

$$\begin{aligned}
I &= \lim_{m \rightarrow \infty} \sum_{\bar{\pi}_{t,s}^m} \sum_{k=1}^n \partial_{y_k} f(Y_{t_j^m}) (Y_{t_{j+1}^m}^k - Y_{t_j^m}^k) \\
&= \lim_{m \rightarrow \infty} \sum_{j=0}^m \mathbb{E} \left[\int_{t_j^m}^{t_{j+1}^m} \sum_{k=1}^n \partial_{y_k} f(Y_{t_j^m}) \partial_x g_k(X_{r-}) dX_r \right. \\
&\quad + \frac{1}{2} \int_{t_j^m}^{t_{j+1}^m} \sum_{k=1}^n \partial_{y_k} f(Y_{t_j^m}) \partial_{xx} g_k(X_{r-}) d[X, X]_r^c \\
&\quad \left. + \sum_{t_j^m < r \leq t_{j+1}^m} \sum_{k=1}^n \partial_{y_k} f(Y_{t_j^m}) \left\{ g_k(X_r) - g_k(X_{r-}) - \partial_x g_k(X_{r-}) \Delta X_r \right\} \right] \\
&= \mathbb{E} \left[\int_t^s (\partial_\mu \Phi)(\mathbb{P}_{X_{r-}}, X_{r-}) dX_r + \frac{1}{2} \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right. \\
&\quad \left. + \sum_{t < r \leq s} \left\{ \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right\} \right],
\end{aligned}$$

286 where in the second equality, (3.14) is applied to Y_t^k between t_j^m and t_{j+1}^m , and in the last equality,
287 $Y_{t_j^m}$ in the second inequality is replaced by Y_{r-} , $r \in (t_j^m, t_{j+1}^m]$ as $\max_j \sup_{r \in (t_j^m, t_{j+1}^m]} |Y_{r-} -$
288 $Y_{t_j^m}| \rightarrow 0$ as $m \rightarrow \infty$.

289 Now, since Y^k is deterministic, $[Y^k, Y^j]_r^c = 0$, and

$$II := \frac{1}{2} \sum_{k=1}^n \sum_{j=1}^n \int_t^s \partial_{y_k y_j} f(Y_{r-}) d[Y^k, Y^j]_r^c = 0.$$

290 Moreover, notice that the third term of (3.12) is

$$\begin{aligned}
III &= \sum_{t < r \leq s} \left\{ \Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) - \sum_{k=1}^n \mathbb{E} \left[\partial_{y_k} f(Y_{r-}) (g_k(X_r) - g_k(X_{r-})) \right] \right\} \\
&= \sum_{t < r \leq s} \left\{ \Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) - \mathbb{E} \left[\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right] \right\}.
\end{aligned}$$

Summing up these three terms I , II , and III , and substituting them into (3.12), we obtain

$$\begin{aligned}
&\Phi(\mathbb{P}_{X_s}) - \Phi(\mathbb{P}_{X_t}) \\
&= \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dX_r + \frac{1}{2} \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right] \\
&\quad + \sum_{t < r \leq s} \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \mathbb{E} \left[\sum_{t < r \leq s} \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right] \\
&\quad + \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right\} \right] \\
&\quad - \sum_{t < r \leq s} \mathbb{E} \left[\left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \right].
\end{aligned}$$

291 As $(\mathbb{P}_{X_r})_{t < r \leq s}$ is a càdlàg function, let us divide $[t, s]$ into countable set $\{r \in [t, s] : \mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}$

292 and uncountable set $\{r \in [t, s] : \mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}$, we have

$$\begin{aligned} & \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \right] \\ &= \sum_{t < r \leq s} \mathbb{E} \left[\left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right] \\ & \quad + \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right], \end{aligned}$$

293 where we switch the order of the expectation and the summation on countable set $\{r \in [t, s] :$

294 $\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}$. Meanwhile,

$$\begin{aligned} & \sum_{t < r \leq s} \mathbb{E} \left[\left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \right] \\ &= \sum_{t < r \leq s} \mathbb{E} \left[\left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right]. \end{aligned}$$

295 We conclude that

$$\begin{aligned} & \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right\} \right] - \sum_{t < r \leq s} \mathbb{E} \left[\left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \right] \\ &= \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right\} \mathbf{1}_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right]. \end{aligned}$$

296 Note that $|\partial_\mu \Phi(\mu, x)|$ is bounded on the compact set $\mathcal{K}_K \times [-K, K]$, and by the same argu-
 297 ment as in Remark 3.2, one can see that the terms $\sum_{t < r \leq s} \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}}$,
 298 $\mathbb{E} \left[\sum_{t < r \leq s} \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right]$, and $\mathbb{E} \left[\sum_{t < r \leq s} \left\{ \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right\} \mathbf{1}_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right]$
 299 are all finite. \square

300 3.2.2. Proof of Theorem 3.1 for $\mathcal{C}^{1,1}(\mathcal{K}_K)$

301 We next establish the Itô's lemma for $\mathcal{C}^{1,1}(\mathcal{K}_K)$. To this end, we first need:

302 **Lemma 3.11 (Stone-Weierstrass)** *Take a compact Hausdorff space H , and let $\mathcal{C}(H)$ be the*
 303 *algebra of real-valued continuous functions on H , with the topology of uniform convergence. Let*
 304 *\mathcal{A} be a subalgebra of $\mathcal{C}(H)$. If \mathcal{A} separates points on H and vanishes at no point on H , then \mathcal{A}*
 305 *is dense in $\mathcal{C}(H)$.*

306 Then we will establish the following Lemma showing that cylindrical functions restricted on
 307 \mathcal{K}_K is dense in $\mathcal{C}^{1,1}(\mathcal{K}_K)$ with appropriate choices of norms.

308 **Lemma 3.12** *$\mathcal{G}(\mathcal{K}_K)$ is dense in $\mathcal{C}^{1,1}(\mathcal{K}_K)$, the collection of all $\mathcal{C}^{1,1}$ functions on \mathcal{K}_K with the*
 309 *supremum norm of derivatives of all orders:*

$$\|\Phi\|_{\mathcal{K}_K} := \sup_{(\mu, x) \in \mathcal{K}_K \times [-K, K]} \left(|\Phi(\mu)| + |\partial_\mu \Phi(\mu, x)| + |\partial_x \partial_\mu \Phi(\mu, x)| \right).$$

310 **Proof.** The proof consists of two steps, and we will adopt the linear derivative.

311 *Step 1.* If $\Phi \in \mathcal{C}^{1,1}(\mathcal{K}_K)$, then $\partial_{xx} \frac{\delta\Phi}{\delta\mu}(\mu, x) \in \mathcal{C}(\mathcal{K}_K \times [-K, K])$. Since we are concerned about
 312 the support of μ , we restrict $\partial_{xx} \frac{\delta\Phi}{\delta\mu}$ to $\mathcal{K}_K \times [-K, K]$. We see that

- 313 • $\mathcal{H}(\mathcal{K}_K \times [-K, K]) =: \mathcal{H}_K$ separates points on $\mathcal{K}_K \times [-K, K]$. To see this, take $(\mu, x) \neq$
 314 $(\mu', x') \in \mathcal{K}_K \times [-K, K]$, we have either $\mu \neq \mu'$ or $x \neq x'$. If $\mu \neq \mu'$, from Theorem 30.1 in
 315 [7], there exists $j_0 \in \mathbb{N}$ such that $\int_{\mathbb{R}} x^{j_0}(\mu - \mu')(dx) \neq 0$, otherwise, $\mu = \mu'$. In this case,
 316 define $p(\mu, x) = \langle x^{j_0}, \mu \rangle \in \mathcal{H}_K$, then $p(\mu, x) \neq p(\mu', x)$. If $\mu = \mu'$, $x \neq x'$, take $p(\mu, x) =$
 317 x , then $p(\mu, x) \neq p(\mu', x')$. In either case, \mathcal{H}_K separates points on $\mathcal{K}_K \times [-K, K]$.
- 318 • \mathcal{H}_K vanishes at no point on $\mathcal{K}_K \times [-K, K]$. This is obvious as one can always choose a
 319 nonzero constant function in \mathcal{H}_K .

320 It thus follows from the Stone-Weierstrass lemma that $\mathcal{H}(\mathcal{K}_K \times [-K, K])$ is dense in $\mathcal{C}(\mathcal{K}_K \times$
 321 $[-K, K])$ with the topology of uniform convergence. Therefore, one can find a sequence of
 322 functions $p_n, \tilde{p}_n \in \mathcal{H}_K$ such that for any $\epsilon > 0$, there exists $N(\epsilon)$ such that for $n \geq N(\epsilon)$

$$\sup_{(\mu, x) \in \mathcal{K}_K \times [-K, K]} \left| p_n(\mu, x) - \partial_{xx} \frac{\delta\Phi}{\delta\mu}(\mu, x) \right| \leq \epsilon, \quad \sup_{\mu \in \mathcal{K}_K} \left| \tilde{p}_n(\mu) - \frac{\delta\Phi}{\delta\mu}(\mu, 0) \right| \leq \epsilon, \quad (3.15)$$

323 *Step 2.* Let

$$\begin{aligned} P_n(\mu, x) &:= \tilde{p}_n(\mu) + \int_0^x \int_0^y p_n(\mu, z) dz dy, \\ \Phi_n(\mu) &:= \Phi(\delta_0) + \int_0^1 \int_{[-K, K]} P_n(\lambda\mu + (1-\lambda)\delta_0, x) (\mu - \delta_0)(dx) d\lambda. \end{aligned}$$

324 One can easily check that $\Phi_n \in \mathcal{G}(\mathcal{K}_K)$ by the above construction. Now we have

$$\begin{aligned} &P_n(\mu, x) - \frac{\delta\Phi}{\delta\mu}(\mu, x) \\ &= \left(\tilde{p}_n(\mu) + \int_0^x \int_0^y p_n(\mu, z) dz dy \right) - \left(\frac{\delta\Phi}{\delta\mu}(\mu, 0) + \int_0^x \int_0^y \partial_{xx} \frac{\delta\Phi}{\delta\mu}(\mu, z) dz dy \right) \\ &= \tilde{p}_n(\mu) - \frac{\delta\Phi(\mu, 0)}{\delta\mu} + \int_0^x \int_0^y (p_n(\mu, z) - \partial_{xx} \frac{\delta\Phi}{\delta\mu}(\mu, z)) dz dy. \end{aligned}$$

325 Thus by (3.15),

$$\begin{aligned} &\sup_{\mathcal{K}_K \times [-K, K]} |\partial_x P_n(\mu, x) - \partial_\mu \Phi(\mu, x)| \leq K \cdot \epsilon, \\ &\sup_{\mathcal{K}_K \times [-K, K]} |P_n(\mu, x) - \frac{\delta\Phi}{\delta\mu}(\mu, x)| \leq (1 + K^2) \cdot \epsilon. \end{aligned}$$

326 Moreover,

$$\begin{aligned} &\Phi_n(\mu) - \Phi(\mu) \\ &= \left(\Phi(\delta_0) + \int_0^1 \int_{[-K, K]} P_n(\lambda\mu + (1-\lambda)\delta_0, x) (\mu - \delta_0)(dx) d\lambda \right) \\ &\quad - \left(\Phi(\delta_0) + \int_0^1 \int_{[-K, K]} \frac{\delta\Phi}{\delta\mu}(\lambda\mu + (1-\lambda)\delta_0, x) (\mu - \delta_0)(dx) d\lambda \right) \\ &= \int_0^1 \int_{[-K, K]} \left(P_n(\lambda\mu + (1-\lambda)\delta_0, x) - \frac{\delta\Phi}{\delta\mu}(\lambda\mu + (1-\lambda)\delta_0, x) \right) (\mu - \delta_0)(dx) d\lambda. \end{aligned}$$

327 Hence

$$\sup_{\mathcal{K}_K} |\Phi_n(\mu) - \Phi(\mu)| \leq 2(1 + K^2)\epsilon.$$

328 Therefore,

$$\|\Phi_n - \Phi\|_{\mathcal{K}_K} \leq \epsilon \left(1 + K + 2(1 + K^2)\right),$$

329 with $\Phi_n \in \mathcal{G}(\mathcal{K}_K)$, which is shown to be dense in $\mathcal{C}^{1,1}(\mathcal{K}_K)$. \square

330 We are now ready to show that Itô's formula (3.2) holds for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{K}_K)$. Without loss
 331 of generality, assume that $X_0 = 0$. Fix $K > 0$ and suppose $|X_r| \leq K$ \mathbb{P} -a.s. for any $r \in [0, T]$.
 332 By Lemma 3.12, for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{K}_K)$, there is a sequence of cylindrical functions $\{\Phi_n\}_{n=1}^\infty$ in
 333 $\mathcal{G}(\mathcal{K}_K)$ such that

$$\|\Phi_n - \Phi\|_{\mathcal{K}_K} \rightarrow 0. \quad (3.16)$$

By the definition of topology of $\mathcal{C}^{1,1}$ in (3.16), it is easy to deduce that

$$\begin{aligned} \sup_{\mu \in \mathcal{K}_K} |\Phi_n(\mu) - \Phi(\mu)| \rightarrow 0, \quad \sup_{\mu \in \mathcal{K}_K, X \sim \mu} \left\{ |\partial_\mu \Phi_n(\mu, X) - \partial_\mu \Phi(\mu, X)| \right. \\ \left. + |\partial_x \partial_\mu \Phi_n(\mu, X) - \partial_x \partial_\mu \Phi(\mu, X)| \right\} \rightarrow 0 \quad \mathbb{P} - a.s. \end{aligned} \quad (3.17)$$

334 From (2.3), (2.4), and (3.17), there exists a K -dependent constant C such that for all $n \in \mathbb{N}$,

$$\sup_{\mu \in \mathcal{K}_K, X \sim \mu} \left(|\partial_\mu \Phi_n(\mu, X)|, |\partial_\mu \Phi(\mu, X)|, |\partial_x \partial_\mu \Phi_n(\mu, X)|, |\partial_x \partial_\mu \Phi(\mu, X)| \right) \leq C. \quad (3.18)$$

335 By Lemma 3.10, Itô's formula holds for $\{\Phi_n\}_{n \geq 1}$

$$\begin{aligned} & \Phi_n(\mathbb{P}_{X_s}) - \Phi_n(\mathbb{P}_{X_t}) \\ &= \mathbb{E} \left[\int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dX_r + \frac{1}{2} \partial_x \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right] \\ &+ \sum_{t < r \leq s} \left\{ \Phi_n(\mathbb{P}_{X_r}) - \Phi_n(\mathbb{P}_{X_{r-}}) \right\} \mathbf{1}_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \mathbb{E} \left[\sum_{t < r \leq s} \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right] \\ &+ \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi_n}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi_n}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right]. \end{aligned} \quad (3.19)$$

336 To establish a similar equation for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{K}_K)$, let us check that each term in the RHS
 337 of (3.19) converges to its suitable limit.

338 • *First term of RHS of (3.19).* Note from (3.18) that the integrand is bounded. By applying
 339 the Dominated Convergence Theorem for stochastic integrals

$$\begin{aligned} \int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r &\xrightarrow{\text{a.s.}} \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r, \\ \int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r &\xrightarrow{\text{p}} \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r, \end{aligned}$$

340 where $\xrightarrow{a.s.}$ and \xrightarrow{P} denote respectively the convergence \mathbb{P} -a.s. and in probability.

341 On one hand, from (3.1)

$$\left| \int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r \right| \leq C \text{Var}(V)_s;$$

342 and by (3.1) and the Dominated Convergence Theorem

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r \right] = \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r \right]. \quad (3.20)$$

343 On the other hand, by Itô's isometry, (3.1), and (3.18)

$$\begin{aligned} \sup_{n \geq 1} \mathbb{E} \left[\left| \int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r \right|^2 \right] &= \sup_{n \geq 1} \mathbb{E} \left[\int_t^s \left| \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) \right|^2 d[X, X]_r \right] \\ &\leq C \mathbb{E}[[X, X]_s] < \infty. \end{aligned}$$

344 This implies the uniform integrability of $\left\{ \int_t^s (\partial_\mu \Phi_n)(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r \right\}_{n \geq 1}$ and

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r \right] = \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r \right]. \quad (3.21)$$

345 Summing up (3.20) and (3.21), we deduce that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_t^s \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) dX_r \right] = \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dX_r \right].$$

346 • *Second term of RHS of (3.19).* By (3.18), $\left| \int_t^s \partial_x \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right| \leq C[X, X]_s^c$. By
347 the Dominated Convergence Theorem,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_t^s \partial_x \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right] = \mathbb{E} \left[\int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right].$$

348 • *Third term of RHS of (3.19).* Note that $|\Delta X_r| \leq 2K$ \mathbb{P} -a.s.. By (3.17),

$$\begin{aligned} &\lim_{n \rightarrow \infty} \left\{ \left(\Phi_n(\mathbb{P}_{X_r}) - \Phi_n(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right\} \\ &= \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r, \quad \mathbb{P} - a.s.. \end{aligned}$$

349 Moreover, by Remark 3.2 and (3.18)

$$\begin{aligned} &\sup_{n \geq 1} \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \left(\Phi_n(\mathbb{P}_{X_r}) - \Phi_n(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right\} \right] \\ &\leq 2 \sup_{n \geq 1} \sup_{(\mu, x) \in \mathcal{H}_K} |\partial_\mu \Phi_n(\mu, x)| \mathbb{E} \left[\sum_{t < r \leq s} |\Delta X_r| \right] \leq C \mathbb{E} \left[\sum_{t < r \leq s} |\Delta X_r| \right]. \end{aligned}$$

350 Applying the Dominated Convergence Theorem yields,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \left(\Phi_n(\mathbb{P}_{X_r}) - \Phi_n(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi_n(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right\} \right] \\ &= \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right\} \right]. \end{aligned}$$

• *Fourth term of RHS of (3.19).* By (3.18),

$$\sup_n \mathbb{E} \left[\sum_{t < r \leq s} \left| \left(\frac{\delta \Phi_n}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi_n}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) \right| 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right] < \infty.$$

351 By the Dominated Convergence Theorem,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{t < r \leq s} \left| \left(\frac{\delta \Phi_n}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi_n}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) \right| 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right] \\ &= \mathbb{E} \left[\sum_{t < r \leq s} \left| \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) \right| 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right]. \end{aligned}$$

352 Now taking the limit on both sides of (3.19), we establish Itô's formula (3.2) for any $\Phi \in$
353 $\mathcal{C}^{1,1}(\mathcal{K}_K)$.

354 3.2.3. Proof of Theorem 3.1

355 We now finish the proof of Theorem 3.1, using a localization argument for general functions
356 over $\mathcal{P}_2(\mathbb{R})$.

357 Take a general process $X = (X_t)_{t \in [0, T]}$, and note that $X_- := (X_{t-})_{t \in [0, T]}$ is locally bounded.
358 Fix $K > 0$, denote $\tau_K = \inf\{0 < t \leq T, |X_{t-}| > K\}$, and define the truncated process

$$X_t^K = X_{t-}^{\tau_K} = X_{t- \wedge \tau_K}.$$

359 Then $|X_t^K| \leq K$ \mathbb{P} -a.s.. Now, for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}))$, one can apply Itô's formula to $\Phi|_{\mathcal{K}_K}$,
360 the restriction of Φ over \mathcal{K}_K :

$$\begin{aligned} & \Phi(\mathbb{P}_{X_s^K}) - \Phi(\mathbb{P}_{X_t^K}) \tag{3.22} \\ &= \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) dX_r^K + \frac{1}{2} \int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) d[X^K, X^K]_r^c \right] \\ &+ \sum_{t < r \leq s} \left\{ \Phi(\mathbb{P}_{X_r^K}) - \Phi(\mathbb{P}_{X_{r-}^K}) \right\} 1_{\{\mathbb{P}_{X_r^K} \neq \mathbb{P}_{X_{r-}^K}\}} - \mathbb{E} \left[\sum_{t < r \leq s} \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) \Delta X_r^K \right] \\ &+ \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r^K}, X_r^K) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r^K}, X_{r-}^K) \right) 1_{\{\mathbb{P}_{X_r^K} = \mathbb{P}_{X_{r-}^K}\}} \right]. \end{aligned}$$

361 To see that Itô's formula (3.2) holds for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}))$, let us check each term of the RHS
362 of (3.22).

363 • *First term of RHS of (3.22).* By the construction of X^K and the stopping rule for stochastic
364 integral

$$\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) dX_r^K = \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) 1_{[0, \tau_K]} dX_r,$$

365 Since $|(\partial_\mu \Phi)(\mathbb{P}_{X_{r-}}, X_{r-}) 1_{[0, \tau_K]}| \leq |(\partial_\mu \Phi)(\mathbb{P}_{X_{r-}}, X_{r-})| \leq C$, \mathbb{P} -a.s., by the Dominated Conver-
366 gence Theorem for stochastic integrals (Proposition 2.74 [34])

$$\begin{aligned} & \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) 1_{[0, \tau_K]} dV_r \xrightarrow{a.s.} \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r, \\ & \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) 1_{[0, \tau_K]} dL_r \xrightarrow{p} \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dL_r. \end{aligned}$$

367 Repeating the same argument as the proof for the first term of RHS of (3.19),

$$\lim_{K \rightarrow \infty} \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) dX_r^K \right] = \mathbb{E} \left[\int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dX_r \right]. \quad (3.23)$$

368 • *Second term of RHS of (3.22).* Note that we have in the pathwise sense

$$\begin{aligned} \int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) d[X^K, X^K]_r^c &= \int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) 1_{[0, \tau_K]} d[X, X]_r^c \\ &\xrightarrow{a.s.} \int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c. \end{aligned}$$

369 Since $|\partial_x \partial_\mu \Phi(\mu, x)| \leq C$ and $\mathbb{E}[[X, X]_T^c] < \infty$ for any $T > 0$, Dominated Convergence Theorem
370 implies

$$\lim_{K \rightarrow \infty} \mathbb{E} \left[\int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) d[X^K, X^K]_r^c \right] = \mathbb{E} \left[\int_t^s \partial_x \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r^c \right].$$

371 • *Third term of RHS of (3.22).* By the construction of X_t^K ,

$$\begin{aligned} &\left(\Phi(\mathbb{P}_{X_r^K}) - \Phi(\mathbb{P}_{X_{r-}^K}) \right) 1_{\{\mathbb{P}_{X_r^K} \neq \mathbb{P}_{X_{r-}^K}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) \Delta X_r^K \\ &\xrightarrow{a.s.} \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r. \end{aligned}$$

372 By Remark 3.2, RHS of the above equation is integrable under $\mathbb{E}[\sum_{t < r \leq s} \cdot]$, and by the Domi-
373 nated Convergence Theorem

$$\begin{aligned} &\lim_{K \rightarrow \infty} \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r^K}) - \Phi(\mathbb{P}_{X_{r-}^K}) \right) 1_{\{\mathbb{P}_{X_r^K} \neq \mathbb{P}_{X_{r-}^K}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) \Delta X_r^K \right\} \right] \\ &= \mathbb{E} \left[\sum_{t < r \leq s} \left\{ \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \Delta X_r \right\} \right]. \end{aligned}$$

374 • *Fourth term of RHS of (3.22).* As

$$\left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r^K}, X_r^K) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) \right) 1_{\{\mathbb{P}_{X_r^K} = \mathbb{P}_{X_{r-}^K}\}} \xrightarrow{a.s.} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}},$$

375 by the Dominated Convergence Theorem,

$$\begin{aligned} &\lim_{K \rightarrow \infty} \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r^K}, X_r^K) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}^K}, X_{r-}^K) \right) 1_{\{\mathbb{P}_{X_r^K} \neq \mathbb{P}_{X_{r-}^K}\}} \right] \\ &= \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_{r-}}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \right]. \quad (3.24) \end{aligned}$$

376 Now, taking limit on both sides of (3.22) and $\Phi(\mathbb{P}_{X_t^K}) \rightarrow \Phi(\mathbb{P}_{X_t})$ as $K \rightarrow \infty$, Itô's formula (3.2)
377 holds for any $\Phi \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}))$. \square

378 3.3. Proof of Theorem 3.3

379 Note that all the arguments of Theorem 3.1 remain essentially the same here. The only
380 modification is to ensure the Dominated Convergence Theorem holds whenever appropriate.
381 Specifically, instead of relying on $|\partial_\mu \Phi(\mu, x)| \leq C$ and $|\partial_x \partial_\mu \Phi(\mu, x)| \leq C$ in Definition 2.4

382 for (3.23)-(3.24) in the proof of Theorem 3.1, one first needs to check that conditions (2.5) in
 383 Definition 2.5 and (3.6) in $(\mathbf{H})_{strict}$ are sufficient for the Dominated Convergence Theorem.
 384 That is to check

$$\begin{aligned} & \mathbb{E} \left[\left| \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) dV_r \right|^2 \right] \leq \sup_{0 < r \leq T} \mathbb{E} \left[|\partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-})|^2 \right] \cdot \mathbb{E} [\text{Var}(V)_T^2] \\ & \mathbb{E} \left[\left| \int_t^s \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) d[X, X]_r \right|^2 \right] \leq \sup_{0 < r \leq T} \mathbb{E} \left[|\partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-})|^2 \right] \cdot \mathbb{E} [|[X, X]_s|^2], \\ & \mathbb{E} \left[\sum_{t < r \leq s} \left| \left(\Phi(\mathbb{P}_{X_r}) - \Phi(\mathbb{P}_{X_{r-}}) \right) 1_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} - \partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-}) \cdot \Delta X_r \right|^2 \right] \\ & \leq 4 \sup_{0 < r \leq T} \mathbb{E} \left[|\partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-})|^2 \right] \mathbb{E} \left[\left(\sum_{0 < t \leq T} |\Delta X_t| \right)^2 \right], \end{aligned}$$

385 and

$$\begin{aligned} & \mathbb{E} \left[\sum_{t < r \leq s} \left(\frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta \Phi}{\delta \mu}(\mathbb{P}_{X_r}, X_{r-}) \right) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right]^2 \\ & \leq \sup_{0 < r \leq T} \mathbb{E} \left[|\partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-})|^2 \right] \cdot \mathbb{E} \left[\left(\sum_{0 < t \leq T} |\Delta X_t| \right)^2 \right], \end{aligned}$$

386 where (3.3)-(3.4) in Remark 3.2 are used. Now, it remains to verify conditions (3.6)-(2.5) are
 387 satisfied: given a semimartingale X , $\widehat{\mathcal{K}} := \mathcal{K}_X \cup \mathcal{K}_{X-}$ is compact in $\mathcal{P}_2(\mathbb{R})$, with $\mathcal{K}_X = \{\mathbb{P}_{X_r}, 0 \leq$
 388 $r \leq T\}$ and $\mathcal{K}_{X-} = \{\mathbb{P}_{X_{r-}}, 0 < r \leq T\}$. To see this, let $\{\mu_{r_n}\}_{n \geq 1}$ be any sequence in $\widehat{\mathcal{K}}$, where
 389 μ_{r_n} is either $\mathbb{P}_{X_{r_n}} \in \mathcal{K}_X$ or $\mathbb{P}_{X_{r_n-}} \in \mathcal{K}_{X-}$. Since $\{r_n\}_{n \geq 1}$ has a convergent subsequence in $[0, T]$,
 390 denoted as $\{r_{n_k}\}_{k \geq 1}$, one can show that $\mu_{r_{n_k}}$ also converges with its limit in $\widehat{\mathcal{K}}$. Consequently
 391 by (2.5), $\sup_{0 < r \leq T} \mathbb{E} \left[|\partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-})|^2 \right] \leq \sup_{\widehat{\mathcal{K}}} \mathbb{E} \left[|\partial_\mu \Phi(\mathbb{P}_{X_{r-}}, X_{r-})|^2 \right] < \infty$. \square

392 **Remark 3.13 (Generalization to $d > 1$)** For ease of exposition, the proof for Theorem 3.1
 393 is given for $d = 1$. Nevertheless, its adaptation for the case of $d > 1$ is straightforward, including
 394 Definition 3.8, Definition 3.9, Lemma 3.10, and Lemma 3.12. Indeed by [19], the underlying
 395 space in $\mathcal{G}(\mathcal{M})$ in Definition 3.8 and $\mathcal{H}(\mathcal{M}, \mathcal{U})$ in Definition 3.9 can be any locally compact
 396 Polish space. When the underlying space is \mathbb{R}^d , the compact set in $\mathcal{P}_2(\mathbb{R}^d)$ can be defined as
 397 $\mathcal{K}_K := \{\mu \in \mathcal{P}_2(\mathbb{R}^d) | \text{supp}(\mu) \subset [-K, K]^d\}$. One can also check that the proof of Lemma 3.10
 398 remains unchanged except for the polynomial g_k , $1 \leq k \leq n$, now defined from \mathbb{R}^d to \mathbb{R} ; hence
 399 the derivative of g_k is replaced by its partial derivative with respect to x_i , $1 \leq i \leq d$.

400 Finally, for Lemma 3.12, one can find $\mathbf{p}_n(\mu, x) := (p_n^{i,j}(\mu, x))_{1 \leq i, j \leq d}$ with $p_n^{i,j}(\mu, x) =$
 401 $p_n^{j,i}(\mu, x)$ such that $|p_n^{i,j}(\mu, x) - \partial_{x_i} \partial_{x_j} \frac{\delta \Phi}{\delta \mu}(\mu, x)| \leq \epsilon$, and then obtain $P_n(\mu, x)$ by integrat-
 402 ing $\mathbf{p}_n(\mu, x)$ with respect to $dz = dz^1 \dots dz^d$ and $dy = dy^1 \dots dy^d$. The rest proceeds similarly
 403 as the proof of Lemma 3.12.

404 **Remark 3.14** Note that we propose two different sets of conditions for Itô's formula: one in
 405 Theorem 3.1 with stronger conditions on the functional and weaker integrability conditions on

406 the semimartingale, the other in Theorem 3.3 with weaker conditions on the functional and
 407 stronger integrability condition on the semimartingale. Note that conditions in [39] are similar
 408 to ours in Theorem 3.3, as both sets of conditions allow for the interchange of the expectation
 409 and integration.

410 Comparing conditions (2.5) and (3.6) with Definition 3.1 and equation (3.2) in [39], it is
 411 clear that our condition (3.6) on the semimartingale X is from the Doob-Meyer decomposition
 412 of the semimartingale X while equation (3.2) in [39] is from the Doob-Meyer decomposition
 413 of the continuous part of the semimartingale X . Note that we assume $\partial_\mu\Phi$ and $\partial_x\partial_\mu\Phi$ in
 414 (2.5) (Definition 2.5) to be bounded in L^2 , locally uniform in μ , whereas [39] assumes that (see
 415 their Definition 3.1) $\frac{\delta\Phi}{\delta\mu}$ has quadratic growth in x , $\partial_\mu\Phi$ has linear growth in x , and $\partial_x\partial_\mu\Phi$ is
 416 bounded, all locally uniform in μ ; The assumptions in Definition 3.1, [39] provide with sufficient
 417 conditions for our assumptions in (2.5).

418 **Remark 3.15** We would also like to point out that there appears to have some essential diffi-
 419 culties to adopt the particle approximation approach for general semimartingales. For instance,
 420 there are two terms for jumps in Itô's formula in Theorems 3.1 & 3.3: the first jump term
 421 sums over the set of time points at which the law of the semimartingale is not continuous, and
 422 the second jump term sums over the set of time points at which the law of the semimartingale
 423 is continuous while the semimartingale itself has jumps. Using the particle approximation ap-
 424 proach, the first jump term in the standard Itô's formula of the empirical projection converges
 425 to the first jump term, as expected. However, it is unclear whether the second jump term in
 426 the standard Itô's formula of the empirical projection converges to the second jump term asso-
 427 ciated with the linear derivative in Theorem 3.1 or Theorem 3.3. To see this more precisely,
 428 suppose that $((X_t^n)_{t \geq 0})_{n \geq 1}$ is a sequence of i.i.d. copies of the semimartingale $(X_t)_{t \geq 0}$. Denote
 429 by $\mu_t^N = \frac{1}{N} \sum_{n=1}^N \delta_{X_t^n}$ as the empirical measure of $((X_t^n)_{t \geq 0})_{n \geq 1}$. Given an integer $N \geq 1$, the
 430 empirical projection of $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ onto \mathbb{R}^d is defined as

$$\phi^N : (\mathbb{R}^d)^N \ni (x_1, \dots, x_N) \mapsto \Phi\left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i}\right).$$

431 The strategy of the particle approximation approach is then to expand $\phi^N : (\mathbb{R}^d)^N \ni (x_1, \dots, x_N)$
 432 using classical Itô's formula and take the limit. However, one cannot see clearly in the expan-
 433 sion of ϕ^N whether the term $\mathbb{E}[\sum_{t < r \leq s} (\phi^N(X_r^1, \dots, X_r^N) - \phi^N(X_{r-}^1, \dots, X_{r-}^N)) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}}]$
 434 converges, as $N \rightarrow \infty$, to $\mathbb{E}[\sum_{t < r \leq s} (\frac{\delta\Phi}{\delta\mu}(\mathbb{P}_{X_r}, X_r) - \frac{\delta\Phi}{\delta\mu}(\mathbb{P}_{X_r}, X_{r-})) 1_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}}]$ in (3.2).

435 The remaining part of this paper demonstrates how Itô's formula enables us to derive dy-
 436 namic programming equations and verification theorems for McKean–Vlasov controls with jump
 437 diffusions and for McKean–Vlasov mixed regular-singular control problems. It also allows for
 438 generalizing the classical relation between the maximum principle and the dynamic program-

439 ming principle to the McKean–Vlasov singular control setting, where the adjoint process is
 440 expressed in terms of the derivative of the value function with respect to probability measures.

441 4. McKean–Vlasov Control with Jump-diffusion Process

442 4.1. Problem Formulation and HJB Equation

On a given probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, we consider a d -dimensional Brownian motion $W = (W_t)_{t \geq 0}$, and an independent Poisson random measure N with a finite intensity measure ν , with $\tilde{N}(d\theta, dt) := N(d\theta, dt) - \nu(d\theta)dt$ the compensated Poisson random measure. Fix $T < \infty$, consider the following McKean–Vlasov control problem with jump-diffusion process, where the \mathbb{R}^d -valued state variable X_s ($s \geq t$) follows the dynamics starting from $t \in [0, T]$,

$$\begin{cases} dX_s &= b(X_s, \alpha_s, \mathbb{P}_{X_s})ds + \sigma(X_s, \alpha_s, \mathbb{P}_{X_s})dW_s \\ &+ \int_{\mathbb{R}^q} \beta(X_{s-}, \alpha_{s-}, \mathbb{P}_{X_{s-}}, \theta)N(ds, d\theta), \quad s \geq t, \\ X_{t-} &= \xi \in L^2(\mathcal{F}_t; \mathbb{R}^d), \end{cases} \quad (4.1)$$

443 Here the control $\alpha = (\alpha_s)_{s \geq t}$ is a càdlàg \mathbb{F} -adapted process taking values in a subset A of \mathbb{R}^m ,
 444 satisfying the square integrability condition: $\mathbb{E}[\int_t^T |\alpha_s|^2 ds] < \infty$, and denoted by $\alpha \in \mathcal{A}_t$.

445 Define a cost functional

$$J(t, \xi, \alpha) = \mathbb{E} \left[\int_t^T f(X_s^{t, \xi, \alpha}, \alpha_s, \mathbb{P}_{X_s^{t, \xi, \alpha}})ds + g(X_T^{t, \xi, \alpha}, \mathbb{P}_{X_T^{t, \xi, \alpha}}) \right], \quad (4.2)$$

446 then the McKean–Vlasov control problem is to find an optimal control (if exists) to minimize
 447 the cost functional J . That is to solve for

$$V(t, \xi) = \inf_{\alpha \in \mathcal{A}_t} J(t, \xi, \alpha), \quad (4.3)$$

448 subject to the jump diffusion (4.1).

449 To ensure the well-definedness of the control problem, the following conditions on coefficients
 450 b , σ and β are imposed.

451 **(H1)** There exists a constant $C_{b, \sigma, \beta} > 0$ such that the coefficients b , σ defined from $\mathbb{R}^d \times A \times$
 452 $\mathcal{P}_2(\mathbb{R}^d)$ to \mathbb{R}^d , $\mathbb{R}^{d \times d}$ respectively and β from $\mathbb{R}^d \times A \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^q$ to \mathbb{R}^d satisfy:

$$|b(x, a, \mu)| + |\sigma(x, a, \mu)| + \int_{\mathbb{R}^q} |\beta(x, a, \mu, \theta)|\nu(d\theta) \leq C_{b, \sigma, \beta} \left(1 + |x| + |a| + \|\mu\|_2 \right),$$

453 for any $x \in \mathbb{R}^d$, $a \in A$, and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Moreover, there exists a constant $C'_{b, \sigma, \beta} > 0$ such that

$$\begin{aligned} & |b(x, a, \mu) - b(x', a', \mu')| + |\sigma(x, a, \mu) - \sigma(x', a', \mu')| + \int_{\mathbb{R}^q} |\beta(x, a, \mu, \theta) - \beta(x', a', \mu', \theta)|\nu(d\theta) \\ & \leq C'_{b, \sigma, \beta} \left(|x - x'| + |a - a'| + W_2(\mu, \mu') \right), \end{aligned}$$

454 for any $x, x' \in \mathbb{R}^d$, $a, a' \in A$, and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$.

455 One can easily check that if $\alpha \in \mathcal{A}_t$, the dynamics (4.1) under conditions **(H1)** has a unique
 456 square integrable solution, and we denote by X the solution to (4.1) and omit the superscripts
 457 of $X^{t,\xi,\alpha}$ whenever there is no confusion.

458 Furthermore, to ensure that the cost functional in (4.2) is well defined and finite, we will
 459 make the following assumptions on coefficients f and g .

460 **(H2)** There exists $C_{f,g} > 0$ such that the coefficients $f : \mathbb{R}^d \times A \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ and $g :$
 461 $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ satisfy:

$$|f(x, a, \mu)| + |g(x, \mu)| \leq C_{f,g} \left(1 + |x|^2 + |a|^2 + \|\mu\|_2^2 \right),$$

462 for all $x \in \mathbb{R}^d$, $a \in A$, and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

463 Now, proceeding by the same argument as in [17], we have the dynamic programming prin-
 464 ciple (PP) for this McKean–Vlasov control problem (4.1)-(4.3).

465 *DPP.* Under assumption **(H1)**, the value function in (4.3) is law-invariant. Moreover, for $\mu =$
 466 $\mathbb{P}_\xi \in \mathcal{P}_2(\mathbb{R}^d)$, with a slight abuse of notation, write

$$V(t, \mu) = V(t, \xi) = \inf_{\alpha \in \mathcal{A}_t} J(t, \xi, \alpha). \quad (4.4)$$

Then, the dynamic programming principle holds such that

$$V(t, \mu) = \inf_{\alpha \in \mathcal{A}_t} \left\{ \mathbb{E} \left[\int_t^s f(X_r^{t,\xi,\alpha}, \alpha_r, \mathbb{P}_{X_r^{t,\xi,\alpha}}) dr \right] + V(s, \mathbb{P}_{X_s^{t,\xi,\alpha}}) \right\}, \quad (\text{MKV-Jump-DPP})$$

467 for all $t, s \in [0, T]$, with $t < s$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and any $\xi \in L^2(\mathcal{F}_t; \mathbb{R}^d)$ such that $\mathbb{P}_\xi = \mu$.

468 Next, define an operator H on $\mathbb{R}^d \times A \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^{d \times d} \times \mathbb{L}_\nu$, with \mathbb{L}_ν the set of all
 469 deterministic integrable functions with respect to ν , such that

$$\begin{aligned} H(x, a, \mu, p, M, r(\cdot)) &= f(x, a, \mu) + b(x, a, \mu) \cdot p + \frac{1}{2} \text{Tr}(\sigma \sigma^\top(x, a, \mu) M) \\ &+ \int_{\mathbb{R}^q} (r(x + \beta(x, a, \mu, \theta)) - r(x)) \nu(d\theta). \end{aligned} \quad (4.5)$$

470 Then, if the value function in (4.3) is sufficiently smooth, applying Corollary 3.6 to (MKV-
 471 Jump-DPP) yields the following HJB equation

$$\begin{cases} \partial_t V(t, \mu) + \mathbb{E} \left[\inf_{\alpha \in A} H(\xi, a, \mu, \partial_\mu V(t, \mu, \xi), \partial_x \partial_\mu V(t, \mu, \xi), \frac{\delta V}{\delta \mu}(t, \mu, \xi)) \right] = 0, \\ V(T, \mu) = \langle g(\cdot, \mu), \mu \rangle. \end{cases} \quad (4.6)$$

472 One can further connects the value function $V(t, \mu)$ in (MKV-Jump-DPP) with the HJB
 473 equation (4.6) in the following verification theorem.

474 4.2. Verification Theorem

475 **Theorem 4.1 (Verification Theorem)** Let $\widehat{V} : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ be a function in $\mathcal{C}^{1,(1,1)}([0, T] \times$
 476 $\mathcal{P}_2(\mathbb{R}^d))$.

477 (a) Suppose that \widehat{V} is a solution to the HJB equation (4.6), then $\widehat{V}(t, \mu) \leq V(t, \mu)$ for any
 478 $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, with V given in (4.4).

479 (b) Moreover, suppose that there exists a continuous function $\hat{a}(t, x, \mu)$ for $(t, x, \mu) \in [0, T] \times$
 480 $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, valued in A such that

$$\hat{a}(t, x, \mu) \in \arg \min_{a \in A} H(x, a, \mu, \partial_\mu \widehat{V}(t, \mu, x), \partial_x \partial_\mu \widehat{V}(t, \mu, x), \frac{\delta \widehat{V}}{\delta \mu}(t, \mu, x)); \quad (4.7)$$

481 suppose also that the corresponding controlled dynamics under (4.7)

$$\begin{cases} dX_s &= b(X_s, \hat{a}(s, X_s, \mathbb{P}_{X_s}), \mathbb{P}_{X_s})ds + \sigma(X_s, \hat{a}(s, X_s, \mathbb{P}_{X_s}), \mathbb{P}_{X_s})dW_s \\ &+ \int_{\mathbb{R}^q} \beta(X_{s-}, \hat{a}(s, X_{s-}, \mathbb{P}_{X_s}), \mathbb{P}_{X_s}, \theta)N(ds, d\theta), \quad t \leq s \leq T, \\ X_{t-} &= \xi. \end{cases}$$

482 admits a unique solution denoted as X_s^* , and the control $\alpha^* = \{\alpha_s^* := \hat{a}(s, X_s^*, \mathbb{P}_{X_s^*}), t \leq$
 483 $s < T\} \in \mathcal{A}_t$. Then $\widehat{V} = V$, with V given in (4.4), and α^* is an optimal Markovian
 484 control.

485 **Proof.** (a) Fix $(t, \xi) \in [0, T] \times L^2(\mathcal{F}_t; \mathbb{R}^d)$, with $\mathbb{P}_\xi = \mu \in \mathcal{P}_2(\mathbb{R}^d)$, and consider $X = X^{t, \xi, \alpha}$, a
 486 solution to SDE (4.1) for an arbitrary control $\alpha \in \mathcal{A}_t$. By assumption **(H1)**,

$$\mathbb{E} \left[\sup_{t \leq s \leq T} |X_s^{t, \xi, \alpha}|^2 \right] \leq C_T (1 + \mathbb{E}[|\xi|^2]) < \infty,$$

487 which implies that assumption **(H)** holds, i.e.,

$$\mathbb{E} \left[\int_t^T \left\{ |b(X_s, \alpha_s, \mathbb{P}_{X_s})|^2 + |\sigma(X_s, \alpha_s, \mathbb{P}_{X_s})|^2 + \int_{\mathbb{R}^d} |\beta(X_s, \alpha_s, \mathbb{P}_{X_s}, \theta)| \nu(d\theta) \right\} ds \right] < \infty.$$

488 Now applying Corollary 3.6 to $\widehat{V}(s, \mathbb{P}_{X_s^{t, \xi, \alpha}})$ between t and T yields

$$\begin{aligned} \widehat{V}(t, \mu) &= \widehat{V}(T, \mathbb{P}_{X_T}) - \int_t^T \left\{ \partial_s \widehat{V}(s, \mathbb{P}_{X_s}) + \mathbb{E} \left[\partial_\mu \widehat{V}(\mathbb{P}_{X_s}, X_s) \cdot b(X_s, \alpha_s, \mathbb{P}_{X_s}) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \text{Tr}(\sigma \sigma^\top(X_s, \alpha_s, \mathbb{P}_{X_s}) \partial_x \partial_\mu \widehat{V}(\mathbb{P}_{X_s}, X_s)) \right. \right. \\ &\quad \left. \left. + \int_{\mathbb{R}^q} \left(\frac{\delta \widehat{V}}{\delta \mu}(\mathbb{P}_{X_s}, X_s + \beta(X_s, \alpha_s, \mathbb{P}_{X_s}, \theta)) - \frac{\delta \widehat{V}}{\delta \mu}(\mathbb{P}_{X_s}, X_s) \right) \nu(d\theta) \right] \right\} ds \\ &= \widehat{V}(T, \mathbb{P}_{X_T}) - \int_t^T \left\{ \partial_s \widehat{V}(s, \mathbb{P}_{X_s}) - \mathbb{E} \left[f(X_s, \alpha_s, \mathbb{P}_{X_s}) \right] \right. \\ &\quad \left. + \mathbb{E} \left[H(X_s, \alpha_s, \mathbb{P}_{X_s}, \partial_\mu \widehat{V}(s, \mathbb{P}_{X_s}, X_s), \partial_x \partial_\mu \widehat{V}(s, \mathbb{P}_{X_s}, X_s), \frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_s}, X_s)) \right] \right\} ds \\ &\leq \widehat{V}(T, \mathbb{P}_{X_T}) + \mathbb{E} \left[\int_t^T f(X_s, \alpha_s, \mathbb{P}_{X_s}) ds \right] \\ &= \mathbb{E} \left[g(X_T, \mathbb{P}_{X_T}) + \int_t^T f(X_s, \alpha_s, \mathbb{P}_{X_s}) ds \right] \\ &\leq V(t, \mu), \end{aligned}$$

489 where the first inequality is from the HJB equation (4.6) for \widehat{V} , the last equality is from the
 490 terminal condition for \widehat{V} , and the final inequality is by the definition of V .

491 (b) Now we apply the same argument again with the control $\alpha^* \in \mathcal{A}_t$ given by $\alpha_s^* = \hat{a}(s, X_s^*, \mathbb{P}_{X_s^*})$,
 492 $t \leq s \leq T$, with \hat{a} attaining the infimum in (4.7), and obtain

$$\widehat{V}(t, \mu) = \widehat{V}(T, \mathbb{P}_{X_T^{t, \xi, \alpha^*}}) + \mathbb{E} \left[\int_t^T f(X_s^{t, \xi, \alpha^*}, \alpha_s^*, \mathbb{P}_{X_s^{t, \xi, \alpha^*}}) dr \right],$$

493 which implies that $\widehat{V}(t, \mu) = V(t, \mu)$. □

494 **Remark 4.2** *In the case when there is no smooth solution to the HJB equation (4.6), then its*
 495 *solution should be interpreted in the viscosity sense as in [9], [16], [15]. Notice that a particular*
 496 *form of Itô's formula for cylindrical functions has also been used in [9] in order to derive the viscosity*
 497 *solution property of the value function.*

4.3. Example: McKean–Vlasov Linear Quadratic Control Problem with Jump-diffusion Process

499 In the particular case of a McKean–Vlasov linear quadratic (LQ) control problem with jump
 500 diffusion, we will illustrate how to apply the Bellman equation (4.6) and the verification theorem
 501 4.1 to derive explicit solutions. In fact, a similar problem has been solved in [40] via a duality
 502 approach where the jump is driven by a Poisson random martingale measure, and also studied in
 503 [4] which considers jump-diffusion-regime switching state dynamics of multiple decisions makers.

504 For simplicity, take $d = 1$ and $A = \mathbb{R}$ as the analysis can be easily generalized to the
 505 multivariate case. In the LQ case, coefficients of dynamics in (4.1) are specified as

$$\begin{aligned} b(x, a, \mu) &= b_0 + b_1 x + \bar{b}_1 \bar{\mu} + b_2 a, \\ \sigma(x, a, \mu) &= \sigma_0 + \sigma_1 x + \bar{\sigma}_1 \bar{\mu} + \sigma_2 a, \\ \beta(x, a, \mu, \theta) &= \beta_0(\theta) + \beta_1(\theta) x + \bar{\beta}_1(\theta) \bar{\mu} + \beta_2(\theta) a, \end{aligned}$$

506 for $(x, a, \mu) \in \mathbb{R} \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R})$, where $\bar{\mu} := \int_{\mathbb{R}} x \mu(dx)$; the cost functional in (4.2) takes the form
 507 of

$$\begin{aligned} f(x, a, \mu) &= f_1 x^2 + \bar{f}_1 \bar{\mu}^2 + f_2 a^2, \\ g(x, \mu) &= g_1 x^2 + \bar{g}_1 \bar{\mu}^2, \end{aligned}$$

508 where $\text{Var}(\mu) := \int x^2 \mu(dx) - \bar{\mu}^2$. Here $b_0, \sigma_0, b_i, \sigma_i, (i = 1, 2), \bar{b}_1, \bar{\sigma}_1, f_1, \bar{f}_1, f_2, g_1, \bar{g}_1$, are con-
 509 stants in \mathbb{R} ; and $\beta_0, \beta_1, \bar{\beta}_1, \beta_2$ are deterministic functions of θ .

510 We start by guessing the form of the solution to the HJB equation (4.6) with

$$\widehat{V}(t, \mu) = A(t) \text{Var}(\mu) + B(t) \bar{\mu}^2 + C(t) \bar{\mu} + D(t) \tag{4.8}$$

511 for some time-dependent deterministic functions $A(t), B(t), C(t)$ and $D(t)$. Then

$$\begin{aligned}\frac{\delta \widehat{V}}{\delta \mu}(t, \mu, x) &= A(t)x^2 + 2(B(t) - A(t))\bar{\mu}x + C(t)x + c, \\ \partial_\mu \widehat{V}(t, \mu, x) &= 2A(t)x + 2(B(t) - A(t))\bar{\mu} + C(t), \\ \partial_x \partial_\mu \widehat{V}(t, \mu, x) &= 2A(t),\end{aligned}$$

512 where c is a constant. Suppose that \widehat{V} in (4.8) satisfies the HJB equation (4.6), by some straight-
513 forward calculations, one can decompose H in (4.5) into two parts, with one part involving a ,
514 denoted as $G(a)$, and the other part independent of a , such that

$$\begin{aligned}& \inf_{a \in \mathbb{R}} H(x, a, \mu, \partial_\mu \widehat{V}(t, \mu), \partial_x \partial_\mu \widehat{V}(t, \mu), \frac{\delta \widehat{V}}{\delta \mu}(t, \mu)) \\ &= \inf_{a \in \mathbb{R}} G(a) + \left(f_1 + 2b_1 A(t) + |\sigma_1|^2 A(t) + \langle 2\beta_1 + |\beta_1|^2, \nu \rangle A(t) \right) x^2 \\ &+ \left(2(B(t) - A(t))b_1 + 2\bar{b}_1 A(t) + 2\sigma_1 \bar{\sigma}_1 A(t) + 2\langle \beta_1 \bar{\beta}_1 + \bar{\beta}_1, \nu \rangle A(t) + 2\langle \beta_1, \nu \rangle (B(t) - A(t)) \right) x \bar{\mu} \\ &+ \left(\bar{f}_1 + 2\bar{b}_1 (B(t) - A(t)) + |\bar{\sigma}_1|^2 A(t) + \langle |\bar{\beta}_1|^2, \nu \rangle A(t) + 2\langle \bar{\beta}_1, \nu \rangle (B(t) - A(t)) \right) \bar{\mu}^2 \\ &+ \left(2b_0 A(t) + b_1 C(t) + 2\sigma_0 \sigma_1 A(t) + 2\langle \beta_0 \beta_1 + \beta_0, \nu \rangle A(t) + \langle \beta_1, \nu \rangle C(t) \right) x \\ &+ \left(\bar{b}_1 C(t) + 2b_0 (B(t) - A(t)) + 2\sigma_0 \bar{\sigma}_1 A(t) + 2\langle \beta_0 \bar{\beta}_1, \nu \rangle A(t) + 2\langle \beta_0, \nu \rangle (B(t) - A(t)) \right. \\ &\left. + \langle \bar{\beta}_1, \nu \rangle C(t) \right) \bar{\mu} + b_0 C(t) + \sigma_0^2 A(t) + \langle |\beta_0|^2, \nu \rangle A(t) + \langle \beta_0, \nu \rangle C(t),\end{aligned}$$

515 where $G(a)$ takes the following quadratic form

$$\begin{aligned}G(a) &= U(t)a^2 + 2S(t)xa + 2(Z(t) - S(t))\bar{\mu}a + Y(t)a \\ &= U(t) \left(a + \frac{S(t)}{U(t)}(x - \bar{\mu}) + \frac{Z(t)}{U(t)}\bar{\mu} + \frac{Y(t)}{2U(t)} \right)^2 \\ &\quad - \frac{S(t)^2}{U(t)}(x - \bar{\mu})^2 - \frac{Z(t)^2}{U(t)}\bar{\mu}^2 - \frac{Z(t)Y(t)}{U(t)}\bar{\mu} - \frac{Y(t)^2}{4U(t)} \\ &\quad - \frac{2S(t)Z(t)}{U(t)}(x - \bar{\mu})\bar{\mu} - \frac{S(t)Y(t)}{U(t)}(x - \bar{\mu}),\end{aligned} \tag{4.9}$$

516 with

$$\begin{cases} U(t) &= f_2 + \sigma_2^2 A(t) + \langle \beta_2^2, \nu \rangle A(t), \\ S(t) &= b_2 A(t) + \sigma_1 \sigma_2 A(t) + \langle \beta_1 \beta_2 + \beta_2, \nu \rangle A(t), \\ Z(t) &= b_2 B(t) + (\sigma_1 + \bar{\sigma}_1) \sigma_2 A(t) + \langle (\beta_1 + \bar{\beta}_1) \beta_2, \nu \rangle A(t) + \langle \beta_2, \nu \rangle B(t), \\ Y(t) &= b_2 C(t) + 2\sigma_0 \sigma_2 A(t) + 2\langle \beta_0 \beta_2, \nu \rangle A(t) + \langle \beta_2, \nu \rangle C(t). \end{cases}$$

517 It is clear by (4.9) that $G(a)$ attains the infimum at

$$\hat{a}(t, x, \mu) = -\frac{S(t)}{U(t)}(x - \bar{\mu}) - \frac{Z(t)}{U(t)}\bar{\mu} - \frac{Y(t)}{2U(t)}. \tag{4.10}$$

518 Substituting (4.9) with \hat{a} into (4.6), we obtain

$$\begin{aligned}
0 &= \partial_t \widehat{V}(t, \mu) + \mathbb{E} \left[\inf_{a \in \mathbb{R}} H(\xi, a, \mu, \partial_\mu \widehat{V}(t, \mu, \xi), \partial_x \partial_\mu \widehat{V}(t, \mu, \xi), \frac{\delta \widehat{V}}{\delta \mu}(t, \mu, \xi)) \right] \quad (4.11) \\
&= \left(\dot{A}(t) + f_1 + 2b_1 A(t) + |\sigma_1|^2 A(t) + \langle 2\beta_1 + |\beta_1|^2, \nu \rangle A(t) - \frac{S(t)^2}{U(t)} \right) \text{Var}(\mu) \\
&\quad + \left(\dot{B}(t) + f_1 + \bar{f}_1 + 2(b_1 + \bar{b}_1)B(t) + (\sigma_1 + \bar{\sigma}_1)^2 A(t) + \langle (\beta_1 + \bar{\beta}_1)^2, \nu \rangle A(t) \right. \\
&\quad \quad \left. + 2\langle (\beta_1 + \bar{\beta}_1), \nu \rangle B(t) - \frac{Z(t)^2}{U(t)} \right) \bar{\mu}^2 \\
&\quad + \left(\dot{C}(t) + 2(b_1 + \bar{b}_1)C(t) + 2b_0 B(t) + 2\sigma_0(\sigma_1 + \bar{\sigma}_1)A(t) + 2\langle \beta_0(\beta_1 + \bar{\beta}_1), \nu \rangle A(t) \right. \\
&\quad \quad \left. + 2\langle \beta_0, \nu \rangle B(t) + \langle \beta_1 + \bar{\beta}_1, \nu \rangle C(t) \right) \bar{\mu} \\
&\quad + \dot{D}(t) + b_0 C(t) + \sigma_0^2 A(t) + \langle |\beta_0|^2, \nu \rangle A(t) - \frac{Y(t)^2}{4U(t)}.
\end{aligned}$$

519 Now comparing terms in $\text{Var}(\mu)$, $\bar{\mu}^2$, $\bar{\mu}$ in (4.11), we get the following ODEs system for $A(t)$,
520 $B(t)$, $C(t)$ and $D(t)$,

$$\left\{ \begin{array}{l} \dot{A}(t) + f_1 + 2b_1 A(t) + |\sigma_1|^2 A(t) + \langle 2\beta_1 + |\beta_1|^2, \nu \rangle A(t) - \frac{S(t)^2}{U(t)^2} = 0, \\ A(T) = g_1, \end{array} \right.$$

521

$$\left\{ \begin{array}{l} \dot{B}(t) + f_1 + \bar{f}_1 + 2(b_1 + \bar{b}_1)B(t) + (\sigma_1 + \bar{\sigma}_1)^2 A(t) + \langle (\beta_1 + \bar{\beta}_1)^2, \nu \rangle A(t) \\ \quad + 2\langle (\beta_1 + \bar{\beta}_1), \nu \rangle B(t) - \frac{Z(t)^2}{U(t)} = 0, \\ B(T) = g_1 + \bar{g}_1, \end{array} \right.$$

522

$$\left\{ \begin{array}{l} \dot{C}(t) + 2(b_1 + \bar{b}_1)C(t) + 2b_0 B(t) + 2\sigma_0(\sigma_1 + \bar{\sigma}_1)A(t) + 2\langle \beta_0(\beta_1 + \bar{\beta}_1), \nu \rangle A(t) \\ \quad + 2\langle \beta_0, \nu \rangle B(t) + \langle \beta_1 + \bar{\beta}_1, \nu \rangle C(t) = 0, \\ C(T) = 0, \end{array} \right.$$

523

$$\left\{ \begin{array}{l} \dot{D}(t) + 2b_0 C(t) + \sigma_0^2 A(t) + \langle |\beta_0|^2, \nu \rangle A(t) + \langle \beta_0, \nu \rangle C(t) - \frac{Y(t)^2}{4U(t)} = 0, \\ D(T) = 0. \end{array} \right.$$

524 By the Verification Theorem 4.1, the solution for the McKean–Vlasov LQ control problem
525 is given by the solution of the Riccati equations for $A(t)$ and $B(t)$ and those of the linear
526 equations for $C(t)$ and $D(t)$ given $A(t)$ and $B(t)$. See [42] for sufficient conditions on the
527 existence of solutions to Riccati equations.

528 From (4.10), the optimal control is given by

$$\alpha_t^* = -\frac{S(t)}{U(t)}(X_t^* - \mathbb{E}[X_t^*]) - \frac{Z(t)}{U(t)}\mathbb{E}[X_t^*] - \frac{Y(t)}{2U(t)},$$

529 where X_t^* is the controlled dynamic associated with α_t^* . In fact, one can further compute $\mathbb{E}[X_t^*]$
530 so that

$$d\mathbb{E}[X_t^*] = R(t)\mathbb{E}[X_t^*]dt + Q(t),$$

531 with

$$\begin{aligned} R(t) &= b_1 + \bar{b}_1 + \langle \beta_1 + \bar{\beta}_1, \nu \rangle - \frac{Z(t)}{U(t)}(b_2 + \langle \beta_2, \nu \rangle), \\ Q(t) &= b_0 + \langle \beta_0, \nu \rangle - \frac{Y(t)}{2U(t)}(b_2 + \langle \beta_2, \nu \rangle). \end{aligned}$$

532 Note that the path of X_t^* is discontinuous, yet $\mathbb{E}[X_t^*]$ is continuous.

533 5. McKean–Vlasov Mixed Regular-Singular Control Problem

534 5.1. Problem Formulation and Dynamic Programming Equation

On a given probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, we consider a d -dimensional Brownian motion $W = (W_t)_{t \geq 0}$. A mixed regular and singular controlled McKean–Vlasov dynamics can be formulated as the following SDE starting from $t \in [0, T]$

$$\begin{cases} dX_s &= b(X_s, \alpha_s, \mathbb{P}_{X_s})ds + \sigma(X_s, \alpha_s, \mathbb{P}_{X_s})dW_s + \lambda d\eta_s, \quad t \leq s \leq T, \\ X_{t-} &= \xi \in L^2(\mathcal{F}_t; \mathbb{R}^d), \end{cases} \quad (5.1)$$

535 for some measurable functions $b : \mathbb{R}^d \times \mathbb{R}^m \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R}^m \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times d}$, and
536 some nonnegative constant $\lambda = \text{diag}(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d \times d}$. Here the mixed regular and singular
537 control pair $(\alpha_s, \eta_s)_{t \leq s \leq T}$ is a pair of processes valued in $\mathbb{R}^m \times [0, \infty)^d$ such that

- 538 • $(\alpha_s)_{t \leq s \leq T} \in \mathcal{A}_t := \left\{ \alpha : [t, T] \rightarrow \mathbb{R}^m \mid \alpha \text{ is } \mathbb{F}\text{-adapted, and } \mathbb{E} \left[\int_t^T |\alpha_s|^2 ds \right] < \infty \right\}$,
- 539 • $(\eta_s)_{t \leq s \leq T} \in \mathcal{U}_t$ is a nondecreasing càdlàg \mathbb{F} -adapted process satisfying $\mathbb{E}[\text{Var}(V)_T] < \infty$.

Given the dynamics (5.1), the McKean–Vlasov regular-singular control problem is to derive, over the admissible control set $\mathcal{A}_t \times \mathcal{U}_t$, the following value function

$$\begin{aligned} \tilde{V}(t, \xi) &= \inf_{(\alpha, \eta) \in \mathcal{A}_t \times \mathcal{U}_t} \tilde{J}(t, \xi, \alpha, \eta) \\ &= \inf_{(\alpha, \eta) \in \mathcal{A}_t \times \mathcal{U}_t} \mathbb{E} \left[\int_t^T \left(f(X_s, \alpha_s, \mathbb{P}_{X_s}) ds + g(X_T, \mathbb{P}_{X_T}) \right) + \int_{[t, T)} \gamma d\eta_s \right], \end{aligned} \quad (5.2)$$

540 where γ is a fixed constant in \mathbb{R}^d and $\int_{[t, T)} \gamma d\eta_t$ means that a jump at the terminal time T is
541 not allowed.

542 To ensure this McKean–Vlasov control problem (5.1)-(5.2) is well defined, we will make the
543 following assumptions on the coefficients for the coefficients b and σ :

544 **(H3)** There exists a constant $C_{b, \sigma} > 0$ such that for any $x, x' \in \mathbb{R}^d$, $a, a' \in \mathbb{R}^m$, and $\mu, \mu' \in$
545 $\mathcal{P}_2(\mathbb{R}^d)$

$$|b(x, a, \mu) - b(x', a', \mu')| + |\sigma(x, a, \mu) - \sigma(x', a', \mu')| \leq C_{b, \sigma} (|x - x'| + W_2(\mu, \mu') + |a - a'|),$$

546 and

$$|b(x, a, \mu)| + |\sigma(x, a, \mu)| \leq C_{b, \sigma} (1 + |x| + \|\mu\|_2 + |a|).$$

547 Note that under condition **(H3)**, equation (5.1) has a unique strong solution, denoted as
 548 $X^{t,\xi,\alpha,\eta}$, for any admissible control $(\alpha, \eta) \in \mathcal{A}_t \times \mathcal{U}_t$.

549 Furthermore, we will assume that the running functions f and g satisfy the square growth
 550 condition.

551 **(H4)** There exists a constant $C_{f,g} > 0$ such that

$$|f(x, a, \mu)| + |g(x, \mu)| \leq C_{f,g} \left(1 + |x|^2 + |a|^2 + \|\mu\|_2^2 \right),$$

552 for all $x \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$

553 Next, proceeding by the same arguments as [16], we can obtain the DPP for this McKean-
 554 Vlasov regular-singular control problem (5.1)-(5.2).

Under assumptions **(H3)**-**(H4)**, for any $t, s \in [0, T]$ with $t < s$, and $\xi \in L^2(\mathcal{F}_t; \mathbb{R}^d)$ with
 $\mathbb{P}_\xi = \mu$, we have

$$V(t, \mu) \triangleq \tilde{V}(t, \xi) = \inf_{(\alpha, \eta) \in \mathcal{A}_t \times \mathcal{U}_t} \mathbb{E} \left[\int_t^s f(X_r, \alpha_r, \mathbb{P}_{X_r}) dr + \int_{[t,s]} \gamma d\eta_r \right] + V(s, \mathbb{P}_{X_s^{t,\xi}}). \quad (\text{MKV-Regular-Singular-DPP})$$

555 Suppose that $V(t, \mu)$ is sufficiently smooth. By applying Corollary 3.7 to (MKV-Regular-
 556 Singular-DPP), we obtain the dynamic programming equation:

$$\begin{aligned} 0 = & \inf_{(\alpha, \eta) \in \mathcal{A}_t \times \mathcal{U}_t} \int_t^s \left\{ \partial_r V(r, \mathbb{P}_{X_r}) + \mathbb{E} \left[\partial_\mu V(r, \mathbb{P}_{X_r}, X_r) \cdot b(X_r, \alpha_r, \mathbb{P}_{X_r}) \right. \right. \\ & \left. \left. + \frac{1}{2} \text{Tr}(\sigma \sigma^\top(X_r, \alpha_r, \mathbb{P}_{X_r}) \partial_x \partial_\mu V(r, \mathbb{P}_{X_r}, X_r)) + f(X_r, \alpha_r, \mathbb{P}_{X_r}) \right] \right\} dr \\ & + \mathbb{E} \left[\int_t^s (\lambda \partial_\mu V(r, \mathbb{P}_{X_{r-}}, X_{r-}) + \gamma) d\eta_r \right] - \mathbb{E} \left[\sum_{t \leq r \leq s} \partial_\mu V(r, \mathbb{P}_{X_{r-}}, X_{r-}) \cdot \Delta X_r \right] \\ & + \sum_{t \leq r \leq s} \left(V(r, \mathbb{P}_{X_r}) - V(r, \mathbb{P}_{X_{r-}}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} \neq \mathbb{P}_{X_{r-}}\}} \\ & + \mathbb{E} \left[\sum_{t \leq r \leq s} \left(\frac{\delta V}{\delta \mu}(r, \mathbb{P}_{X_r}, X_r) - \frac{\delta V}{\delta \mu}(r, \mathbb{P}_{X_{r-}}, X_{r-}) \right) \mathbf{1}_{\{\mathbb{P}_{X_r} = \mathbb{P}_{X_{r-}}\}} \right], \end{aligned} \quad (5.3)$$

557 for all $0 \leq t < s < T$. Notice that the last two terms of RHS of (5.3) can be written in integral
 558 form

$$\begin{aligned} V(r, \mathbb{P}_{X_r}) - V(r, \mathbb{P}_{X_{r-}}) &= \int_0^1 \mathbb{E}[\partial_\mu V(r, \mathbb{P}_{X_{r-}+h\Delta X_r}, X_{r-} + h\Delta X_r) \cdot \Delta X_r] dh, \\ \frac{\delta V}{\delta \mu}(r, \mathbb{P}_{X_r}, X_r) - \frac{\delta V}{\delta \mu}(r, \mathbb{P}_{X_{r-}}, X_{r-}) &= \int_0^1 \partial_\mu V(r, \mathbb{P}_{X_{r-}+h\Delta X_r}, X_{r-} + h\Delta X_r) \cdot \Delta X_r dh. \end{aligned}$$

559 Hence, by rearranging the last three terms in (5.3), we obtain the following definition of a
 560 classical solution to (5.3).

561 **Definition 5.1** Consider $\widehat{V} \in \mathcal{C}^{1,(1,1)}([0, T] \times \mathcal{P}_2(\mathbb{R}^d))$, define

$$\mathcal{C}(\widehat{V}) := \left\{ (t, \mu, x) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d : \gamma + \lambda \partial_\mu \widehat{V}(t, \mu, x) > 0 \right\}, \quad (5.4)$$

562 denote by $\mathcal{C}_{t,\mu}(\widehat{V})$ the projection of $\mathcal{C}(\widehat{V})$ onto (t, μ) -space, and define

$$H(x, a, \mu, p, M) = f(x, a, \mu) + b(x, a, \mu) \cdot p + \text{Tr}(\sigma(x, a, \mu)^\top M). \quad (5.5)$$

563 We say that \widehat{V} is a classical solution to the dynamic programming equation (5.3) if the following
564 conditions hold

- 565 • for every $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, if $(t, \mu, \xi) \in \mathcal{C}(\widehat{V})$, \mathbb{P} -a.s. with $\xi \sim \mu$.

$$\partial_t \widehat{V}(t, \mu) + \mathbb{E} \left[\inf_{a \in \mathbb{R}^m} H(\xi, a, \mu, \partial_\mu \widehat{V}(t, \mu, \xi), \partial_x \partial_\mu \widehat{V}(t, \mu, \xi) \sigma(\mu, a, \xi)) \right] = 0, \quad (5.6)$$

- 566 • for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\widehat{V}(T, \mu) = \langle g(\cdot, \mu), \mu \rangle$,

- 567 • for every $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ and $\xi \sim \mu$,

$$\gamma + \lambda \partial_\mu \widehat{V}(t, \mu, \xi) \geq 0, \quad \mathbb{P} - \text{a.s.}, \quad (5.7)$$

- 568 • for every $(t, \mu, a) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^m$,

$$\partial_t \widehat{V}(t, \mu) + \mathbb{E} \left[H(\xi, a, \mu, \partial_\mu \widehat{V}(t, \mu, \xi), \partial_x \partial_\mu \widehat{V}(t, \mu, \xi) \sigma(\mu, a, \xi)) \right] \geq 0. \quad (5.8)$$

569 5.2. Verification Theorem

570 Theorem 5.2 (Verification Theorem)

571 (a) Suppose that $\widehat{V} \in \mathcal{C}^{1,(1,1)}([0, T] \times \mathcal{P}_2(\mathbb{R}^d))$ is a classical solution of the dynamic program-
572 ming equation (5.3), then $\widehat{V}(t, \mu) \leq V(t, \mu)$ for any $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, where V is
573 given in (MKV-Regular-Singular-DPP).

574 (b) Furthermore, if there exists $(\alpha^*, \eta^*) \in \mathcal{A}_t \times \mathcal{U}_t$ such that

$$\mathbb{P} \left(\text{Leb a.e. } s \in [t, T], X_s^* := X_s^{t, \xi, \alpha^*, \eta^*} \in \mathcal{C}_{s, \mathbb{P}_{X_s^*}}(\widehat{V}) \right) = 1, \quad (5.9)$$

575 and for every $(s, \mathbb{P}_{X_s^*}, X_s^*) \in \mathcal{C}(\widehat{V})$,

$$\alpha_s^* = \hat{a}(s, X_s^*, \mathbb{P}_{X_s^*}), \quad \text{with } \hat{a}(s, x, \mu) \in \arg \min_{a \in \mathbb{R}^m} H(x, a, \mu, \partial_\mu \widehat{V}(s, \mu), \partial_x \partial_\mu \widehat{V}(s, \mu)), \quad (5.10)$$

576

$$\mathbb{E} \int_{[t, T]} \left(\lambda \partial_\mu \widehat{V}(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*) + \gamma \right) d\eta_s^* = 0, \quad (5.11)$$

577

$$\widehat{V}(s, \mathbb{P}_{X_{s-}^*}) = \widehat{V}(t, \mathbb{P}_{X_s^*}) + \gamma \mathbb{E}[\Delta \eta_s^*] \quad \text{for any } s \in \{t \leq s < T : \mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}, \quad (5.12)$$

and

$$\mathbb{P} \left(\frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_s^*}, X_s^*) = \frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*) + \gamma \Delta \eta_s^*, \text{ for all } t \leq s < T \right. \\ \left. \text{with } \mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*} \right) = 1. \quad (5.13)$$

578 Then \widehat{V} is the value function of problem (5.1)-(5.2). That is, $\widehat{V}(t, \mu) = V(t, \mu)$ for any
579 $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, where V is given in (MKV-Regular-Singular-DPP).

580 **Proof.** (a) Fix $(t, \xi) \in [0, T] \times L^2(\mathcal{F}_t; \mathbb{R}^d)$ with $\mathbb{P}_\xi = \mu$ and consider $X_s = X_s^{t, \xi, \alpha, \eta}$ solution
581 to SDE (5.1) for a given arbitrary admissible control $(\alpha, \eta) \in \mathcal{A}_t \times \mathcal{U}_t$. Under conditions **(H3)**-
582 **(H4)**,

$$\mathbb{E} \left[\sup_{t \leq s \leq T} |X_s^{t, \xi, \alpha, \eta}|^2 \right] \leq C_T (1 + \mathbb{E}[|\xi|^2]) < \infty.$$

583 This implies assumption **(H)**. Applying Corollary 3.7 to $\widehat{V}(s, \mathbb{P}_{X_s^{t, \xi}})$ between t and T ,

$$\begin{aligned} \widehat{V}(t, \mu) &= \widehat{V}(T, \mathbb{P}_{X_T}) - \int_t^T \left\{ \partial_s \widehat{V}(s, \mathbb{P}_{X_s}) + \mathbb{E} \left[\partial_\mu \widehat{V}(s, \mathbb{P}_{X_s}, X_s) \cdot b(X_s, \alpha_s, \mathbb{P}_{X_s}) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \text{Tr}(\sigma \sigma^\top(X_s, \alpha_s, \mathbb{P}_{X_s}) \partial_x \partial_\mu \widehat{V}(s, \mathbb{P}_{X_s}, X_s)) \right] ds \right\} \\ &\quad - \mathbb{E} \left[\int_t^T \lambda \partial_\mu \widehat{V}(s, \mathbb{P}_{X_{s-}}, X_{s-}) \cdot d\eta_s^c \right] \\ &\quad - \sum_{t \leq s < T} \left(\widehat{V}(s, \mathbb{P}_{X_s}) - \widehat{V}(s, \mathbb{P}_{X_{s-}}) \right) \mathbf{1}_{\{\mathbb{P}_{X_s} \neq \mathbb{P}_{X_{s-}}\}} \\ &\quad - \mathbb{E} \left[\sum_{t \leq s < T} \left(\frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_s}, X_s) - \frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_s}, X_{s-}) \right) \mathbf{1}_{\{\mathbb{P}_{X_s} = \mathbb{P}_{X_{s-}}\}} \right]. \end{aligned} \quad (5.14)$$

584 Since \widehat{V} is a classical solution of (5.3), by (5.8),

$$\begin{aligned} &\partial_s \widehat{V}(s, \mathbb{P}_{X_s}) + \mathbb{E} \left[\partial_\mu \widehat{V}(s, \mathbb{P}_{X_s}, X_s) \cdot b(X_s, \alpha_s, \mathbb{P}_{X_s}) + \frac{1}{2} \text{Tr}(\sigma \sigma^\top(X_s, \alpha_s, \mathbb{P}_{X_s}) \partial_x \partial_\mu \widehat{V}(s, \mathbb{P}_{X_s}, X_s)) \right] \\ &\geq -f(X_s, \alpha_s, \mathbb{P}_{X_s}). \end{aligned}$$

585 Moreover, by (5.7), we have

$$\mathbb{E} \left[\int_t^T \lambda \partial_\mu \widehat{V}(s, \mathbb{P}_{X_{s-}}, X_{s-}) \cdot d\eta_s^c \right] \geq -\gamma E \left[\int_t^T d\eta_s^c \right].$$

586 According to the Fundamental Theorem of Calculus and (5.7),

$$\begin{aligned} \widehat{V}(s, \mathbb{P}_{X_s}) - \widehat{V}(s, \mathbb{P}_{X_{s-}}) &= \int_0^1 \mathbb{E} \left[\lambda \partial_\mu \widehat{V}(s, \mathbb{P}_{X_{s-} + h\Delta X_s}, X_{s-} + h\Delta X_s) \cdot \Delta \eta_s \right] dh \geq -\gamma \mathbb{E}[\Delta \eta_s], \\ \frac{\delta \widehat{V}}{\delta \mu}(r, \mathbb{P}_{X_r}, X_r) - \frac{\delta \widehat{V}}{\delta \mu}(r, \mathbb{P}_{X_r}, X_{r-}) &= \int_0^1 \partial_\mu \widehat{V}(r, \mathbb{P}_{X_r}, X_{r-} + h\Delta X_r) \cdot \Delta X_r dh \geq -\gamma \Delta \eta_s. \end{aligned}$$

587 From the above three inequalities, with the terminal condition for \widehat{V} plugged into (5.14), we
588 obtain

$$\widehat{V}(t, \mu) \leq \mathbb{E} \left[\int_t^T \left(f(X_s, \alpha_s, \mathbb{P}_{X_s}) dt + g(X_T, \mathbb{P}_{X_T}) \right) + \int_{[t, T)} \gamma d\eta_s \right].$$

589

590 (b) Applying (5.14) with $\mathbb{P}_{X_s^*}$, we see from (5.4), (5.9), and (5.10) that the second term on the
591 right hand side of (5.14) becomes

$$\mathbb{E} \left[\int_t^T f(X_s^*, \alpha_s^*, \mathbb{P}_{X_s^*}) ds \right].$$

592 Furthermore, by (5.11)-(5.13),

$$\begin{aligned}
& -\mathbb{E}\left[\int_t^T \lambda \partial_\mu \widehat{V}(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*) d\eta_s^{*,c}\right] - \sum_{t \leq s < T} \left(\widehat{V}(s, \mathbb{P}_{X_s^*}) - \widehat{V}(s, \mathbb{P}_{X_{s-}^*})\right) 1_{\{\mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}} \\
& -\mathbb{E}\left[\sum_{t \leq s < T} \left(\frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_s^*}, X_s^*) - \frac{\delta \widehat{V}}{\delta \mu}(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*)\right) 1_{\{\mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}}\right] \\
& = \mathbb{E}\left[\int_t^T \gamma d\eta_s^{*,c}\right] + \gamma \sum_{t \leq s < T} \mathbb{E}[\Delta \eta_s^*] 1_{\{\mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}} + \gamma \mathbb{E}\left[\sum_{t \leq s < T} \Delta \eta_s^* 1_{\{\mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}}\right] \\
& = \mathbb{E}\left[\int_t^T \gamma d\eta_s^*\right].
\end{aligned}$$

593 Now by applying (5.14) to (α^*, η^*) , we conclude that

$$\widehat{V}(t, \mu) = \mathbb{E}\left[\int_t^T f(X_s^*, \alpha_s^*, \mathbb{P}_{X_s^*}) ds + g(X_T^*, \mathbb{P}_{X_T^*}) + \int_{[t, T)} \gamma d\eta_s^*\right].$$

594

□

595 **Remark 5.3** According to Verification Theorem 5.2, if there exists $(\alpha^*, \eta^*) \in \mathcal{A}_t$ satisfying
596 (5.10)-(5.12), then $\widehat{V} = V$ and (α^*, η^*) is optimal. Let us denote $\mathcal{C}(V)$ and $\mathcal{D}(V)$ for the
597 η^* -continuation region and the η^* -action region, respectively, with

$$\mathcal{C}(V) = \{(t, \mu, x) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d : \gamma + \lambda \partial_\mu V(t, \mu, x) > 0\}, \quad (5.15)$$

$$\mathcal{D}(V) = \cup_{i=1}^d \mathcal{D}_i(V), \quad (5.16)$$

where

$$\mathcal{D}_i(V) = \{(t, \mu, x) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d : \gamma_i + \lambda_i (\partial_\mu V)_i(t, \mu, x) = 0\}, \quad 1 \leq i \leq d.$$

598 Moreover, if $(t, \mathbb{P}_{X_{t-}^*}, X_{t-}^*)$ starts from η^* -action region $\mathcal{D}(V)$ with positive probability, then
599 $(t, \mathbb{P}_{X_{t-}^*}, X_{t-}^*)$ jumps immediately to a point on the boundary of $\mathcal{C}(V)$, denoted as $\partial \mathcal{C}(V)$. In
600 fact, (5.9) suggests that if $(t, \mathbb{P}_{X_{t-}^*}, X_{t-}^*)$ is in the η^* -action region with a positive probability,
601 then $(t, \mathbb{P}_{X_t^*}, X_t^*)$ must be either in the interior of $\mathcal{C}(V)$ or in its boundary \mathbb{P} -a.s.. To check that
602 indeed it can not be in the interior of $\mathcal{C}(V)$, let us define

$$\tau := \inf \left\{ h \in [0, 1] : (t, \mathbb{P}_{X_{t-}^* + h \Delta X_t^*}, X_{t-}^* + h \Delta X_t^*) \in \partial \mathcal{C}(V) \right\}.$$

603 If $\mathbb{P}_{X_{t-}^*} \neq \mathbb{P}_{X_t^*}$, according to (5.12), together with the Fundamental Theorem of Calculus

$$\begin{aligned}
0 & = V(t, \mathbb{P}_{X_t^*}) - V(t, \mathbb{P}_{X_{t-}^*}) + \gamma \mathbb{E}[\Delta \eta_t^*] \\
& = \int_0^1 \mathbb{E}\left[\left(\lambda \partial_\mu V(t, \mathbb{P}_{X_{t-}^* + h \Delta X_t^*}, X_{t-}^* + h \Delta X_t^*) + \gamma\right) \Delta \eta_t^*\right] dh \\
& = \int_0^1 \mathbb{E}\left[\left(\lambda \partial_\mu V(t, \mathbb{P}_{X_{t-}^* + h \Delta X_t^*}, X_{t-}^* + h \Delta X_t^*) + \gamma\right) 1_{\{h < \tau\}} \Delta \eta_t^*\right] dh \\
& \quad + \int_0^1 \mathbb{E}\left[\left(\lambda \partial_\mu V(t, \mathbb{P}_{X_{t-}^* + h \Delta X_t^*}, X_{t-}^* + h \Delta X_t^*) + \gamma\right) 1_{\{h > \tau\}} \Delta \eta_t^*\right] dh.
\end{aligned}$$

604 From (5.9), the integral over $[0, \tau]$ is nonnegative and the integral over $(\tau, 1]$ is positive, thus a
605 contradiction unless $\tau = 1$ \mathbb{P} -a.s., which means that $(t, \mathbb{P}_{X_t^*}, X_t^*)$ is on $\partial\mathcal{C}(V)$ \mathbb{P} -a.s. Similarly,
606 if $\mathbb{P}_{X_t^*} = \mathbb{P}_{X_{t-}^*}$, according to (5.13), together with the Fundamental Theorem of Calculus, $\tau = 1$
607 \mathbb{P} -a.s. In either case, once $(t, \mathbb{P}_{X_t^*}, X_t^*)$ reaches the boundary, η_t^* acts only to prevent it from
608 entering the interior of $\mathcal{D}(V)$.

609 5.3. Relation to the Maximum Principle

610 As discussed in the introduction, the maximum principle and the dynamic programming
611 principle are the two most common approaches in solving stochastic optimal control problems.
612 Under certain differentiability conditions, these two principles are related in the sense that the
613 derivatives of the value function and the solution to the adjoint equation along the optimal state
614 process are connected. This relationship has been discussed in [43], [35] for classical diffusion
615 processes, in [24], [38], [44] for diffusion with jumps, and in [3] for singular stochastic optimal
616 control problems.

617 In this section, we will build a similar relationship for McKean–Vlasov (MKV) regular-
618 singular control case. In order to put it in the framework suitable for the maximum principle,
619 we will make the following assumptions.

620 **(H5)** Measurable functions $b(\cdot, a, \cdot)$, $\sigma(\cdot, a, \cdot)$ and $f(\cdot, a, \cdot)$ are differentiable with respect to x
621 and μ and all derivatives are bounded and Lipschitz continuous, where Lipschitz constants
622 are independent of $a \in \mathbb{R}^m$. Moreover, functions b, σ and f are continuously differentiable
623 with respect to the control a , and all their derivatives are continuous and bounded. Similarly,
624 $g(\cdot, \cdot)$ are differentiable with respect to x and μ and all derivatives are bounded and Lipschitz
625 continuous.

626 **(H6)** The Hamiltonian $H(\cdot, \cdot, \cdot, p, M)$ defined in (5.5) is convex with respect to (x, μ, a) , i.e., for
627 every $x, x' \in \mathbb{R}^d$, $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, and $a, a' \in \mathbb{R}^m$,

$$\begin{aligned} H(x', a', \mu', p, M) &\geq H(x, a, \mu, p, M) + \partial_x H(x, a, \mu, p, M) \cdot (x' - x) \\ &\quad + \partial_a H(x, a, \mu, p, M) \cdot (a' - a) + \mathbb{E}[\partial_\mu H(x, a, \mu, p, M, \xi) \cdot (\xi' - \xi)], \end{aligned}$$

628 where ξ and ξ' are square integrable random variables with distributions μ and μ' (respectively).

629 The \mathbb{F} -adapted processes $(p_t, M_t)_t$ are given by the adjoint equation

$$\begin{cases} dp_t &= -\left\{ \partial_x H(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}, p_t, M_t) + \bar{\mathbb{E}}[\partial_\mu H(\bar{X}_t^*, \bar{\alpha}_t^*, \mathbb{P}_{X_t^*}, \bar{p}_t, \bar{M}_t, X_t^*)] \right\} dt \\ &\quad + M_t dW_t, \\ p_T &= \partial_x g(X_T^*, \mathbb{P}_{X_T^*}) + \bar{\mathbb{E}}[\partial_\mu g(\bar{X}_T^*, \mathbb{P}_{X_T^*}, X_T^*)]. \end{cases} \quad (5.17)$$

630 Now recall from [29] the following maximum principle for the MKV mixed regular-singular
631 control problem.

632 **Proposition 5.4** (1) Let X^* be the optimal solution associated with the optimal strategy (α^*, η^*)
633 of McKean–Vlasov mixed regular-singular control problem (5.1)-(5.2). Assume condition **(H5)**.
634 Then there exists a unique pair of \mathbb{F} -adapted processes $(p_t, M_t)_t$ for the adjoint equation (5.17)
635 such that for all $(\alpha, \eta) \in \mathcal{A}_0 \times \mathcal{U}_0$, the following inequality holds \mathbb{P} -a.s., dt -a.e.

$$\partial_a H(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}, p_t, M_t)(\alpha_t - \alpha_t^*) + \mathbb{E} \left[\int_{[0, T)} (\gamma + \lambda p_t) d(\eta_t - \eta_t^*) \right] \geq 0.$$

636 (2) Let $(\hat{\alpha}, \hat{\eta}) \in \mathcal{A}_0 \times \mathcal{U}_0$ be an admissible control, and $X_t^{\hat{\alpha}, \hat{\eta}}$, $(p_t^{\hat{\alpha}}, M_t^{\hat{\alpha}})$ be the solution of (5.1)
637 and (5.17) respectively associated with $(\hat{\alpha}, \hat{\eta})$. Assume conditions **(H5)**-**(H6)**. And suppose that
638 the singular control $(\hat{\alpha}, \hat{\eta})$ satisfies that

$$\mathbb{E} \left[\int_0^T \partial_a H(X_t^{\hat{\alpha}, \hat{\eta}}, \hat{\alpha}_t, \mathbb{P}_{X_t^{\hat{\alpha}, \hat{\eta}}}, p_t^{\hat{\alpha}}, M_t^{\hat{\alpha}})(\alpha_t - \hat{\alpha}_t) dt \right] + \mathbb{E} \left[\int_{[0, T)} (\gamma + \lambda p_t^{\hat{\alpha}}) d(\eta_t - \hat{\eta}_t) \right] \geq 0,$$

639 for any $(\alpha, \eta) \in \mathcal{A}_0 \times \mathcal{U}_0$. Then $(\hat{\alpha}, \hat{\eta})$ is an optimal control of the McKean–Vlasov singular
640 control problem (5.1)-(5.2) starting at time $t = 0$.

641 We now show that the adjoint process can be expressed in terms of the Lions derivative of
642 the value function for the MKV mixed regular-singular control problem.

643 **Theorem 5.5** Assume that $V \in \mathcal{C}^{1, (2,1)}([0, T] \times \mathcal{P}_2(\mathbb{R}^d))$. If $(\alpha^*, \eta^*) \in \mathcal{A}_0 \times \mathcal{U}_0$ satisfies condi-
644 tions (5.9)-(5.12), then the solution of the adjoint equation (5.17) is given by

$$p_t = \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*), \quad M_t = \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) \sigma(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}), \quad 0 \leq t \leq T. \quad (5.18)$$

645 **Proof.** Observe that

$$\min_{(\mu, a) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^m} F(t, \mu, a) = F(t, \mathbb{P}_{X_t^*}, \hat{a}(t, X_t^*, \mathbb{P}_{X_t^*})),$$

646 where $F(t, \mu, a)$ is given by

$$F(t, \mu, a) := \partial_t V(t, \mu) + \mathbb{E} [H(\xi, a, \mu, \partial_\mu V(t, \mu, \xi), \partial_x \partial_\mu V(t, \mu, \xi))].$$

647 Differentiating $F(t, \mu, a)$ with respect to μ and then evaluating at $\mathbb{P}_{X_t^*}$ and \hat{a} yields

$$\partial_\mu F(t, \mathbb{P}_{X_t^*}, \hat{a}(t, X_t^*, \mathbb{P}_{X_t^*})) = 0. \quad (5.19)$$

648 To simplify notations, set

$$b_t^* := b(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}), \quad \bar{b}_t^* := b(\bar{X}_t^*, \bar{\alpha}_t^*, \mathbb{P}_{X_t^*}), \quad \sigma_t^* := \sigma(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}), \quad \bar{\sigma}_t^* := \sigma(\bar{X}_t^*, \bar{\alpha}_t^*, \mathbb{P}_{X_t^*}).$$

649 Then straightforward calculation of (5.19) suggests

$$\begin{aligned} & \partial_t \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) + \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) b_t^* + \bar{\mathbb{E}} [\partial_\mu^2 V(t, \mathbb{P}_{X_t^*}, \bar{X}_t^*, X_t^*) \bar{b}_t^*] \\ & + \frac{1}{2} \left\{ \text{Tr}(\sigma_t^* (\sigma_t^*)^\top \partial_{xx} \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*)) + \bar{\mathbb{E}} [\text{Tr}(\bar{\sigma}_t^* (\bar{\sigma}_t^*)^\top \partial_x \partial_\mu^2 V(t, \mathbb{P}_{X_t^*}, \bar{X}_t^*, X_t^*))] \right\} \\ = & - \left\{ \partial_x H(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}, \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*), \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) \sigma_t^*) \right. \\ & \left. + \bar{\mathbb{E}} [\partial_\mu H(\bar{X}_t^*, \bar{\alpha}_t^*, \mathbb{P}_{X_t^*}, \partial_\mu V(t, \mathbb{P}_{X_t^*}, \bar{X}_t^*), \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, \bar{X}_t^*), X_t^* \bar{\sigma}_t^*)] \right\}. \quad (5.20) \end{aligned}$$

650 Applying Corollary 3.5 to $\partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*)$ between t and T yields

$$\begin{aligned}
& \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) - \partial_\mu V(T, \mathbb{P}_{X_T^*}, X_T^*) \tag{5.21} \\
= & - \int_t^T \left\{ \partial_s \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) + \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) b_s^* + \frac{1}{2} \text{Tr}(\sigma_s^* (\sigma_s^*)^\top \partial_{xx} \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*)) \right\} ds \\
& - \int_t^T \mathbb{E} \left[\partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*) \bar{b}_s^* + \frac{1}{2} \text{Tr}(\bar{\sigma}_s^* (\bar{\sigma}_s^*)^\top \partial_{\bar{x}} \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*)) \right] ds \\
& - \int_t^T \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) \sigma_s^* dW_s \\
& - \lambda \underbrace{\left(\int_t^T \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) d\eta_s^* - \sum_{t \leq s < T} \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) \Delta \eta_s^* \right)}_{I_1} \\
& - \underbrace{\sum_{t \leq s < T} \left(\partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) - \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_{s-}^*) \right) 1_{\{\mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}}}_{I_1^2} \\
& - \underbrace{\sum_{t \leq s < T} \left(\partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) - \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_{s-}^*) \right) 1_{\{\mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}}}_{I_2^2} \\
& - \lambda \mathbb{E} \left[\underbrace{\int_t^T \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*) d\bar{\eta}_s^* - \sum_{t \leq s < T} \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*) \Delta \bar{\eta}_s^*}_{I_3} \right] \\
& - \underbrace{\mathbb{E} \left[\sum_{t \leq s < T} \left(\frac{\delta(\partial_\mu V)}{\delta \mu}(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*) - \frac{\delta(\partial_\mu V)}{\delta \mu}(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_{s-}^*) \right) 1_{\{\mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}} \right]}_{I_4} \\
& - \underbrace{\sum_{t \leq s < T} \left(\partial_\mu V(s, \mathbb{P}_{X_s^*}, X_{s-}^*) - \partial_\mu V(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*) \right) 1_{\{\mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}}}_{I_5}.
\end{aligned}$$

651 Now, we calculate I_1 and I_3 . On one hand, $\mathcal{D}(V) \subset \arg \min\{\gamma + \lambda \partial_\mu V(\mu, x)\}$, then for every
652 $(t, \mu, x) \in \mathcal{D}(V)$ in (5.16),

$$\partial_\mu \left(\gamma + \lambda \partial_\mu V(\mu, x) \right) = \partial_{\mu\mu} V(\mu, x, x') = 0, \quad \partial_x \left(\gamma + \lambda \partial_\mu V(\mu, x) \right) = \partial_x \partial_\mu V(\mu, x) = 0.$$

653 On the other hand, (5.11) implies that $\mathbb{E} \left[\int_t^T 1_{(s, \mathbb{P}_{X_s^*}, X_s^*) \in \mathcal{C}(V)} d\eta_s^{*,c} \right] = 0$, therefore

$$\begin{aligned}
I_1 : & = \lambda \int_t^T \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) d\eta_s^{*,c} \\
& = \lambda \int_t^T 1_{\{(s, \mathbb{P}_{X_s^*}, X_s^*) \in \mathcal{C}(V)\}} \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) d\eta_s^{*,c} \\
& \quad + \lambda \int_t^T 1_{\{(s, \mathbb{P}_{X_s^*}, X_s^*) \in \mathcal{D}(V)\}} \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) d\eta_s^{*,c} \\
& = 0.
\end{aligned}$$

654 Similarly, we have

$$I_3 := \mathbb{E} \left[\int_t^T \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*) d\bar{\eta}_s^* - \sum_{t \leq s < T} \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_s^*) \Delta \bar{\eta}_s^* \right] = 0.$$

655 Next, we calculate $I_2^2 + I_5$. By the Fundamental Theorem of Calculus,

$$\begin{aligned}
I_2^2 + I_5 &: = \sum_{t \leq s < T} \left(\partial_\mu V(s, \mathbb{P}_{X_s^*}, X_s^*) - \partial_\mu V(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*) \right) 1_{\{\mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}} \\
&= \lambda \sum_{t \leq s < T} \left(\int_0^1 \partial_x \partial_\mu V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*) \Delta \eta_s^* dh \right. \\
&\quad \left. + \int_0^1 \bar{\mathbb{E}}[\partial_\mu^2 V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*, \bar{X}_{s-}^* + h\Delta \bar{X}_s^*) \Delta \eta_s^*] dh \right) 1_{\{\mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}}.
\end{aligned}$$

656 To prove that the right hand side in the above equation vanishes, it suffices to check that if
657 $\Delta \eta_s^* > 0$, then for every $h \in [0, 1]$, \mathbb{P} -a.s., $\bar{\mathbb{P}}$ -a.s.

$$\partial_x \partial_\mu V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*) = 0, \quad (5.22)$$

$$\partial_\mu^2 V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*, \bar{X}_{s-}^* + h\Delta \bar{X}_s^*) = 0. \quad (5.23)$$

658 From (5.12), we see that when $s \in \{t \leq s < T : \mathbb{P}_{X_s^*} \neq \mathbb{P}_{X_{s-}^*}\}$,

$$\begin{aligned}
0 &= V(s, \mathbb{P}_{X_s^*}) - V(s, \mathbb{P}_{X_{s-}^*}) + \gamma \mathbb{E}[\Delta \eta_s^*] \\
&= \int_0^1 \mathbb{E} \left[\left(\lambda \partial_\mu V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*) + \gamma \right) \Delta \eta_s^* \right] dh.
\end{aligned}$$

659 The right hand side is nonnegative, $\Delta \eta_s^* > 0$ implies that $(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*) \in \mathcal{D}(V)$,
660 \mathbb{P} -a.s., for every $h \in [0, 1]$. This implies (5.22)-(5.23). Hence, $I_2^2 + I_5 = 0$.

661 Now we compute $I_2^1 + I_4$. By the Fundamental Theorem of Calculus,

$$\begin{aligned}
I_2^1 &= \sum_{t \leq s < T} \left(\int_0^1 \partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_{s-}^* + h\Delta X_s^*) \Delta \eta_s^* dh \right) 1_{\{\mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}}, \\
I_4 &= \bar{\mathbb{E}} \left[\sum_{t \leq s < T} \left(\int_0^1 \lambda \partial_\mu^2 V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_s^*, \bar{X}_{s-}^* + h\Delta \bar{X}_s^*) \Delta \eta_s^* dh \right) 1_{\{\mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}} \right].
\end{aligned}$$

662 To prove the right hand side in the above equation vanishes, it suffices to show if $\Delta \eta_s^* > 0$, then
663 for every $h \in [0, 1]$, \mathbb{P} -a.s., $\bar{\mathbb{P}}$ -a.s.

$$\partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_{s-}^* + h\Delta X_s^*) = 0, \quad \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^*, \bar{X}_{s-}^* + h\Delta \bar{X}_s^*) = 0. \quad (5.24)$$

664 From (5.13), when $s \in \{t \leq s < T : \mathbb{P}_{X_s^*} = \mathbb{P}_{X_{s-}^*}\}$

$$\begin{aligned}
0 &= \frac{\delta V}{\delta \mu}(s, \mathbb{P}_{X_s^*}, X_s^*) - \frac{\delta V}{\delta \mu}(s, \mathbb{P}_{X_{s-}^*}, X_{s-}^*) + \gamma \Delta \eta_s^* \\
&= \int_0^1 \left(\lambda \partial_\mu V(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*) + \gamma \right) \Delta \eta_s^* dh.
\end{aligned}$$

665 The right hand side is nonnegative, $\Delta \eta_s^* > 0$ implies that $(s, \mathbb{P}_{X_{s-}^* + h\Delta X_s^*}, X_{s-}^* + h\Delta X_s^*) \in \mathcal{D}(V)$, \mathbb{P} -a.s.,
666 for every $h \in [0, 1]$. This implies (5.24)

$$\partial_x \partial_\mu V(s, \mathbb{P}_{X_s^*}, X_{s-}^* + h\Delta X_s^*) = 0, \quad \partial_\mu^2 V(s, \mathbb{P}_{X_s^*}, X_s^* + h\Delta X_s^*, \bar{X}_s^*) = 0.$$

667 Hence $I_2^1 = 0$ and $I_4 = 0$.

668 Note that all terms $I_1 - I_5$ associated with η_s^* vanish. By (5.20)-(5.21), we obtain

$$\left\{ \begin{array}{l} d\partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) = -\left\{ \partial_x H(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}, \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*), \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) \sigma_t^*) \right. \\ \quad \left. + \bar{\mathbb{E}}[\partial_\mu H(\bar{X}_t^*, \bar{\alpha}_t^*, \mathbb{P}_{X_t^*}, \partial_\mu V(t, \mathbb{P}_{X_t^*}, \bar{X}_t^*), \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, \bar{X}_t^*) \bar{\sigma}_t^*)] \right\} dt \\ \quad + \partial_x \partial_\mu V(t, \mathbb{P}_{X_t^*}, X_t^*) \sigma(X_t^*, \alpha_t^*, \mathbb{P}_{X_t^*}) dW_t, \\ \partial_\mu V(T, \mathbb{P}_{X_T^*}, X_T^*) = \partial_x g(X_T^*, \mathbb{P}_{X_T^*}) + \bar{\mathbb{E}}[\partial_\mu g(\bar{X}_T^*, \mathbb{P}_{X_T^*}, X_T^*)]. \end{array} \right.$$

669 Thus, the pair (p, M) given by (5.18) is the solution of the adjoint equation (5.17). \square

Remark 5.6 (Mean-field games with singular controls) *When the controls are regular, mean-field game (MFG) and McKean–Vlasov control are related ([13], chapter 6). In the case of singular controls, [25] established the existence of an optimal control to McKean–Vlasov singular control using a similar method for MFGs with singular controls. In fact, one may mimic the case of regular controls and show that MFG with singular control and McKean–Vlasov singular control are also connected. To see this, assume for simplicity that σ in (5.1) does not depend upon on the regular control α , then the FBSDE system given by the maximum principle (5.1)-(5.17) for the McKean–Vlasov control problem may be also identified with (at least formally) the FBSDE system given by applying the maximum principle to the following auxiliary MFG problem, where the dynamics follow*

$$\left\{ \begin{array}{l} dX_s = b(X_s, \alpha_s, \mu_s) ds + \sigma(X_s, \mu_s) dW_s + \lambda d\eta_s, \quad t \leq s \leq T, \\ X_{t-} = \xi \in L^2(\mathcal{F}_t; \mathbb{R}^d), \end{array} \right.$$

670 and the cost functional is given by

$$\begin{aligned} I(\alpha, \eta) &= \mathbb{E} \left[g(X_T, \mu_T) + \int_{\mathbb{R}^d} \frac{\delta g}{\delta \mu}(y, \mu_T, X_T) \mu_T(dy) + \int_t^T f(X_s, \mu_s, \alpha_s) ds \right. \\ &\quad + \int_t^T \int_{\mathbb{R}^d} \frac{\delta H}{\delta \mu} \left(y, \hat{\alpha}(\mu_s, \partial_y u(s, y), \partial_{yy}^2 u(s, y) \sigma(y, \mu_s)), \mu_s, \partial_y u(s, y), \partial_{yy}^2 u(s, y) \sigma(y, \mu_s) \right) \mu_s(dy) \\ &\quad \left. + \int_t^T \gamma d\eta_s \right]. \end{aligned}$$

671 Here $\hat{\alpha}(x, \mu, p, M)$ is the minimizer of $H(x, a, \mu, p, M)$ in (5.5), and the function u is the linear
672 derivative of the value function $V(t, \mu)$ in (MKV-Regular-Singular-DPP) along the optimal path.

673 5.4. Example: Mean-Variance Singular Control

674 We now analyze a class of one-dimensional mean-variance singular control problem, where
675 coefficients of the dynamics in (5.1) are specialized with

$$\begin{aligned} b(x, a, \mu) &= rx + \rho a, \\ \sigma(x, a, \mu) &= \sigma a, \end{aligned}$$

676 for $(x, \mu, a) \in \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \times \mathbb{R}$, with r, ρ and $\sigma > 0$ constants in \mathbb{R} . And in (5.2), the running
677 cost $f \equiv 0$, and the mean-variance terminal cost function is

$$g(x, \mu) = \frac{\beta}{2} (x - \bar{\mu})^2 - x,$$

678 where the constant $\beta > 0$, and $\bar{\mu} := \int_{\mathbb{R}} x\mu(dx)$.

679 We will search for a classical solution to the dynamic programming equation (5.3). In the
680 waiting region $\mathcal{C}(V)$ in (5.15), $V(t, \mu)$ would satisfy the HJB equation (5.6), which corresponds
681 to the classical linear quadratic McKean–Vlasov control problem. Now take $V(t, \mu)$ of the
682 following form

$$V(t, \mu) = A(t)\text{Var}(\mu) + B(t)\bar{\mu}^2 + C(t)\bar{\mu} + D(t),$$

683 for some time-dependent functions A, B, C , and $D : [0, T] \rightarrow \mathbb{R}$. Solving the corresponding
684 HJB equation (5.6) in the similar way as in Section 4.3, we see that $A(t), B(t), C(t)$ and $D(t)$
685 satisfy the following ODEs

$$\begin{cases} \dot{A}(t) - \left(\frac{\rho^2}{\sigma^2} - 2r\right)A(t) = 0, & A(T) = \frac{\beta}{2}, \\ \dot{B}(t) - \frac{\rho^2}{\sigma^2} \frac{B^2(t)}{A(t)} + 2rB(t) = 0, & B(T) = 0, \\ \dot{C}(t) + rC(t) - \frac{\rho^2}{\sigma^2} \frac{B(t)}{A(t)} = 0, & C(T) = -1, \\ \dot{D}(t) - \frac{\rho^2}{\sigma^2} \frac{C(t)^2}{4A(t)} = 0, & D(T) = 0, \end{cases}$$

686 which can be explicitly solved such that

$$\begin{aligned} A(t) &= \frac{\beta}{2} \exp\left((2r - \frac{\rho^2}{\sigma^2})(T - t)\right), \quad B(t) = 0, \\ C(t) &= -\exp\left(r(T - t)\right), \quad D(t) = \frac{1}{4} \exp\left(\frac{\rho^2}{\sigma^2}(T - t)\right) - \frac{1}{4}. \end{aligned}$$

687 Therefore, $\mathcal{C}(V)$ in (5.15) and $\mathcal{D}(V)$ in (5.16) are now given by

$$\begin{aligned} \mathcal{C}(V) &= \left\{ (t, \bar{\mu}, x) \in [0, T] \times \mathbb{R} \times \mathbb{R} : \lambda\beta \exp\left((2r - \frac{\rho^2}{\sigma^2})(T - t)\right)(x - \bar{\mu}) \right. \\ &\quad \left. - \lambda \exp(-r(T - t)) + \gamma > 0 \right\}, \\ \mathcal{D}(V) &= \left\{ (t, \bar{\mu}, x) \in [0, T] \times \mathbb{R} \times \mathbb{R} : \lambda\beta \exp\left((2r - \frac{\rho^2}{\sigma^2})(T - t)\right)(x - \bar{\mu}) \right. \\ &\quad \left. - \lambda \exp(-r(T - t)) + \gamma = 0 \right\}. \end{aligned}$$

688 From (5.9)-(5.12), $\mathbb{P}(\text{Leb a.e. } t \in [0, T], (t, \mathbb{P}_{X_t^*}, X_t^*) \in \mathcal{C}(V)) = 1$, where X_t^* is the controlled
689 process associated with the optimal control (α^*, η^*) , where

$$\alpha_t^* = -\frac{\rho}{\sigma^2}(X_t^* - \mathbb{E}[X_t^*]) + \frac{\rho}{\beta\sigma^2} \exp\left(\left(\frac{\rho^2}{\sigma^2} - r\right)(T - t)\right),$$

690 and η^*

$$\mathbb{E} \int_0^T 1_{\{(t, \mathbb{P}_{X_t^*}, X_t^*) \in \mathcal{D}(V)\}} d\eta_t^* = 0. \quad (5.25)$$

691 The controlled process X_t^* then follows

$$\begin{aligned} dX_t^* &= \left(\left(r - \frac{\rho^2}{\sigma^2}\right)X_t^* + \frac{\sigma^2}{\rho^2} \mathbb{E}[X_t^*] + \frac{\rho^2}{\beta\sigma^2} \exp\left(\frac{\rho^2}{\sigma^2} - r\right)(T - t) \right) dt \\ &\quad + \left(-\frac{\rho}{\sigma}(X_t^* - \mathbb{E}[X_t^*]) + \frac{\rho}{\beta\sigma} \exp\left(\frac{\rho^2}{\sigma^2} - r\right)(T - t) \right) dW_t + \lambda d\eta_t^*. \end{aligned} \quad (5.26)$$

692 Solving (5.25)-(5.26) is equivalent to solving a one-dimensional Skorokhod problem, see [30]. By
693 Theorem 4.1 in [30], such an η^* exists uniquely.

694 **Acknowledgments.** We thank the reviewers for bringing [39] to our attention and for their
695 feedback.

696 References

- 697 [1] Ambrosio, L., Gigli, N., Savaré, G., 2008. Gradient flows: In metric spaces and in the space
698 of probability measures. Springer Science & Business Media.
- 699 [2] Andersson, D., Djehiche, B., 2010. A maximum principle for SDEs of mean-field type.
700 Applied Mathematics and Optimization 63, 341–356.
- 701 [3] Bahlali, K., Chighoub, F., Mezerdi, B., 2012. On the relationship between the stochastic
702 maximum principle and dynamic programming in singular stochastic control. An Interna-
703 tional Journal of Probability and Stochastic Processes 84, 233–249.
- 704 [4] Barreiro-Gomez, J., Duncan, T.E., Tembine, H., 2019. Linear–quadratic mean–field–type
705 games: Jump–diffusion process with regime switching. IEEE Transactions on Automatic
706 Control 64, 4329–4336.
- 707 [5] Bayraktar, E., Cosso, A., Pham, H., 2018. Randomized dynamic programming principle
708 and Feynman-Kac representation for optimal control of McKean-Vlasov dynamics. Trans-
709 actions of the American Mathematical Society 370, 2115–2160.
- 710 [6] Bensoussan, A., Frehse, J., Yam, S.C.P., 2017. On the interpretation of the master equation.
711 Stochastic Processes and their Applications 127, 2093–2137.
- 712 [7] Billingsley, P., 2013. Convergence of probability measures. John Wiley & Sons.
- 713 [8] Buckdahn, R., Li, J., Peng, S., Rainer, C., 2017. Mean-field stochastic differential equations
714 and associated PDEs. The Annals of Probability 45, 824–878.
- 715 [9] Burzoni, M., Ignazio, V., Reppen, A.M., Soner, H.M., 2020. Viscosity solutions for con-
716 trolled McKean–Vlasov jump–diffusions. SIAM Journal on Control and Optimization 58,
717 1676–1699.
- 718 [10] Cardaliaguet, P., 2010. Notes on mean field games. Technical Report. Paris Dauphine.
- 719 [11] Cardaliaguet, P., Delarue, F., Lasry, J.M., Lions, P.L., 2019. The master equation and
720 the convergence problem in mean field games. Series: Annals of Mathematics Studies,
721 Princeton University Press.

- 722 [12] Carmona, R., Delarue, F., 2015. Forward-backward stochastic differential equations and
723 controlled McKean Vlasov dynamics. *The Annals of Probability* 43, 2647–2700.
- 724 [13] Carmona, R., Delarue, F., 2018. *Probabilistic Theory of Mean Field Games with Applica-*
725 *tions*, vol. I. Springer.
- 726 [14] Chassagneux, J.F., Crisan, D., Delarue, F., 2014. A probabilistic approach to clas-
727 sical solutions of the master equation for large population equilibria. ArXiv preprint,
728 arXiv:1411.3009, to appear in *Memoirs of the AMS* .
- 729 [15] Cosso, A., Gozzi, F., Kharroubi, I., Pham, H., Rosestolato, M., 2021a. Master Bellman
730 equation in the Wasserstein space: Uniqueness of viscosity solutions. ArXiv preprint,
731 arXiv:2107.10535 .
- 732 [16] Cosso, A., Gozzi, F., Kharroubi, I., Pham, H., Rosestolato, M., 2021b. Optimal con-
733 trol of path-dependent McKean-Vlasov SDEs in infinite dimension. ArXiv preprint
734 arXiv:2012.14772 .
- 735 [17] Cosso, A., Pham, H., 2019. Zero-sum stochastic differential games of generalized McKean-
736 Vlasov type. *Journal de Mathématiques Pures et Appliquées* 129, 180–212.
- 737 [18] Cox, A.M., Källblad, S., Larsson, M., Svaluto-Ferro, S., 2021. Controlled measure-valued
738 martingales: a viscosity solution approach. arXiv preprint arXiv:2109.00064 .
- 739 [19] Cuchiero, C., Larsson, M., Svaluto-Ferro, S., 2019. Probability measure-valued polynomial
740 diffusions. *Electronic Journal of Probability* 24.
- 741 [20] Djete, M.F., Possamai, D., Tan, X., 2022. McKean–Vlasov optimal control: the dynamic
742 programming principle. *The Annals of Probability* 50, 791–833.
- 743 [21] Dos Reis, G., Platonov, V., 2021. On the relation between stratonovich and Itô integrals
744 with functional integrands of conditional measure flows. arXiv:2111.03523 [.
- 745 [22] Dos Reis, G., Platonov, V., 2022. Itô-Wentzell-Lions formula for measure dependent ran-
746 dom fields under full and conditional measure flows. *Potential Analysis* , 1–32.
- 747 [23] Fleming, W.H., Viot, M., 1979. Some measure-valued Markov processes in population
748 genetics theory. *Indiana University Mathematics Journal* 28, 817–843.
- 749 [24] Framstad, N.C., Øksendal, B., Sulem, A., 2004. Sufficient stochastic maximum principle for
750 the optimal control of jump diffusions and applications to finance. *Journal of optimization*
751 *theory and applications* 121, 77–98.
- 752 [25] Fu, G., Horst, U., 2017. Mean field games with singular controls. *SIAM Journal on Control*
753 *and Optimization* 55, 3833–3868.

- 754 [26] Gangbo, W., Tudorascu, A., 2019. On differentiability in the Wasserstein space and well-
755 posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*
756 125, 119–174.
- 757 [27] Gaviraghi, B., 2017. Theoretical and numerical analysis of Fokker-Planck optimal control
758 problems for jump-diffusion processes. Ph.D. thesis. Universität Würzburg.
- 759 [28] Hafayed, M., Abba, A., Abbas, S., 2014. On mean-field stochastic maximum principle for
760 near-optimal controls for Poisson jump diffusion with applications. *International Journal*
761 *of Dynamics and Control* 2, 262–284.
- 762 [29] Hafayed, M., Meherrem, S., Eren, Ş., Guçoglu, D.H., 2018. On optimal singular control
763 problem for general McKean-Vlasov differential equations: Necessary and sufficient opti-
764 mality conditions. *Optimal Control Applications and Methods* 39, 1202–1219.
- 765 [30] Hu, Y., Øksendal, B., Sulem, A., 2017. Singular mean-field control games. *Stochastic*
766 *Analysis and Applications* 35, 823–851.
- 767 [31] Li, J., 2018. Mean-field forward and backward SDEs with jumps and associated nonlocal
768 quasi-linear integral-PDEs. *Stochastic Processes and their Applications* 128, 3118–3180.
- 769 [32] Lions, P.L., Audio Conference, 2006-2012. Cours au collège de france: Théorie des jeux à
770 champ moyens .
- 771 [33] McKean, H., 1969. Propagation of chaos for a class of non-linear parabolic equations. Lec-
772 ture Series in Differential Equations, Session 7, Catholic Univ, 1967. *Stochastic Differential*
773 *Equations* , 41–57.
- 774 [34] Medvedev, P., 2007. *Stochastic Integration Theory*. (No. 14). Oxford University Press on
775 Demand.
- 776 [35] Pham, H., 2009. *Continuous-time stochastic control and optimization with financial appli-*
777 *cations*. (Vol 61). Springer Science & Business Media.
- 778 [36] Pham, H., Wei, X., 2018. Bellman equation and viscosity solutions for mean-field stochastic
779 control problem. *ESAIM: Control, Optimisation and Calculus of Variations* 24, 437–461.
- 780 [37] Protter, P., 2005. *Stochastic Integration and Differential Equations*. (No. 21). Springer,
781 *Stochastic Modeling and Applied Probability*.
- 782 [38] Shi, J.T., Wu, Z., 2011. Relationship between MP and DPP for the stochastic optimal
783 control problem of jump diffusions. *Applied Mathematics & Optimization* 63, 151–189.
- 784 [39] Talbi, M., Touzi, N., Zhang, J., 2021. Dynamic programming equation for the mean field
785 optimal stopping problem. ArXiv preprint, arXiv:2103.05736 .

- 786 [40] Tang, M., Meng, Q., 2019. Linear-quadratic optimal control problems for mean-field
787 stochastic differential equations with jumps. *Asian Journal of Control* 21, 809–823.
- 788 [41] Wu, C., Zhang, J., 2020. Viscosity solutions to parabolic master equations and McKean–
789 Vlasov SDEs with closed-loop controls. *The Annals of Applied Probability* 30, 936–986.
- 790 [42] Yong, J., 2013. Linear-quadratic optimal control problems for mean-field stochastic differ-
791 ential equations. *SIAM journal on Control and Optimization* 51, 2809–2838.
- 792 [43] Yong, J., Zhou, X.Y., 1999. *Hamiltonian systems and HJB equations*. (Vol 43). Springer.
- 793 [44] Zhang, X., Elliott, R.J., Siu, T.K., 2012. A stochastic maximum principle for a Markov
794 regime-switching jump-diffusion model and its application to finance. *SIAM Journal on*
795 *Control and Optimization* 50, 964–990.