Analysis in the space of measures

P. Cardaliaguet
Paris Dauphine University
April 2019

These notes were written in April 2019 for a course given at the University of Padova where the author was invited. The author wishes to thank the University for its hospitality.

Introduction

The goal of this short course is to introduce some aspects of the calculus in the space of probability measures. These questions have attracted a lot of attention in the last two decades. A first motivation has been the theory of optimal transport and the analysis of the gradient flows in the space of measures [2, 35, 36, 37]. Another subject, the theory of mean field games (MFG), see [9, 16, 29, 30, 31, 32], has shown the necessity to develop calculus and the analysis of partial differential equations in the space of probability measures. Closely related to MFG, the mean field type control problems are nothing but optimal control problem in this space. Let us also note that stochastic control problems in which the cost (or the dynamics) depends on the entire law of the process can be recasted in this framework.

In these short notes we will first briefly survey the various notions related to differentiability in the space of probability measures and compare them. Then we will use these notions to present several computations in this space: optimality conditions, Itô's formulas. We finally discuss a typical example of problem of calculus of variation in the space of probability measures.

Contents

1 The Wasserstein space of probability measures

In this part, we briefly recall without proof classical topics on the space of probability measures and on Wasserstein distance. Standard references on this part are, for instance, the monographs [2, 35, 36, 37].

1.1 Basic definitions and properties

Given $p \geqslant 1$, we denote by $\mathcal{P}_p(\mathbb{R}^d)$ the Wasserstein space of Borel probability measures which satisfy the moment condition

$$\int_{\mathbb{R}^d} |x|^p m(dx) < +\infty.$$

The Wasserstein space on $\mathbb{R}^d$ is endowed with the distance

$$\mathbf{d}_p(m, m') = \left(\inf_{\pi \in \Pi(m,m')} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \pi(dx, dy) \right)^{1/p},$$

where $\Pi(m, m')$ is the set of couplings between m and m', i.e., the Borel measurable measures π on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(A \times \mathbb{R}^d) = m(A)$ and $\pi(\mathbb{R}^d \times A) = m'(A)$ for any Borel set $A \subset \mathbb{R}^d$. We will mostly work on $\mathcal{P}_2(\mathbb{R}^d)$ with the distance $\mathbf{d}_2$.

It is known that the $\mathbf{d}_p$ are distances on $\mathcal{P}_p(\mathbb{R}^d)$ which (almost) metricizes the weak convergence. Namely

Proposition 1.1. *(see [2]) Let (m_n) a sequence in $\mathcal{P}_p(\mathbb{R}^d)$ and $m \in \mathcal{P}_p(\mathbb{R}^d)$. There is an equivalence between:*

(i) $\mathbf{d}_p(m_n, m) \to 0$*,*

(ii) (m_n) *weakly converges to* m *and* $\int_{\mathbb{R}^d} |x|^p m_n(dx) \to \int_{\mathbb{R}^d} |x|^p m(dx)$*.*

(iii) (m_n) *weakly converges to* m *and* $\lim_{R\to+\infty} \sup_n \int_{B_R^c} |x|^p m_n(dx) = 0$*, where* $B_R = \{x \in \mathbb{R}^d,\ |x| \leqslant R\}$*.*

Proposition 1.2 (Existence of optimal plans)**.** *For any $p \geqslant 1$ and any $m, m' \in \mathcal{P}_p(\mathbb{R}^d)$, there exists at least one optimal plan $\pi \in \Pi(m, m')$, i.e., such that*

$$\mathbf{d}_p^p(m, m') = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^p \pi(dx, dy).$$

We denote by $\Pi^{opt}(m, m')$ the set of optimal plans from m to m'.

Proof. Let $\pi_n \in \Pi(m, m')$ be a mininizing sequence. Let us check that it is tight. As $\mathbb{R}^d$ is a Polish space, the measures m and m' are tight: for any $\epsilon > 0$, there exists a compact set $K \subset \mathbb{R}^d$ such $m(\mathbb{R}^d \backslash K) \leqslant \epsilon$ and $m'(\mathbb{R}^d \backslash K) \leqslant \epsilon$. Therefore, for any n,

$$\pi_n(\mathbb{R}^{2d} \backslash K \times K) \leqslant \pi_n((\mathbb{R}^d \backslash K) \times \mathbb{R}^d) + \pi_n(\mathbb{R}^d \times (\mathbb{R}d \backslash K)) \leqslant m(\mathbb{R}^d \backslash K) + m'(\mathbb{R}^d \backslash K) \leqslant 2\epsilon.$$

So (π_n) is tight and, by Prokhorov Theorem, it converges weakly (up to a subsequence denoted in the same way) to a measure π. One easily checks that $\pi \in \Pi(m, m')$. On the other hand, for any $R > 0$,

$$\int_{\mathbb{R}^{2d}} (R \wedge |x-y|^p)\pi(dx,dy) = \lim_n \int_{\mathbb{R}^{2d}} (R \wedge |x-y|^p)\pi_n(dx,dy)$$
$$\leqslant \limsup_n \int_{\mathbb{R}^{2d}} |x-y|^p \pi_n(dx,dy) = \mathbf{d}_p^p(m,m').$$

Letting $R \to +\infty$, we find that

$$\int_{\mathbb{R}^{2d}} |x-y|^p \pi(dx,dy) \leqslant \mathbf{d}_p^p(m,m'),$$

which shows the optimality of π. □

Given $\mathbf{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$, the empirical measure associated with $\mathbf{x}$ is the measure

$$m_{\mathbf{x}}^N := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}.$$

We often use the following result, which is intuitively obvious, but which is not so easy to prove:

Proposition 1.3. *(see [2]) Let $\mathbf{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ and $= (y_1, \dots, y_N) \in (\mathbb{R}^d)^N$. Then*

$$\mathbf{d}_p^p(m_{\mathbf{x}}^N, m_{\mathbf{y}}^N) = \inf_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^N |x_i - y_{\sigma(i)}|^p,$$

where $\mathcal{S}_N$ is the set of permutation over $\{1, \dots, N\}$.

1.2 Formulation in a probability space

Let $(\Omega, \mathbb{F}, \mathbb{P})$ be an atomless probability space. By an atomless probability space, we mean a space such that, for any $A \in \mathbb{F}$ such that $\mathbb{P}[A] > 0$, there exists $B \in \mathbb{F}$ such that $0 < \mathbb{P}[B] \leqslant \mathbb{P}[A]$. Let us recall that, for any Borel measure μ on any Polish space E, there exists a $E-$valued random variable X with law $\mu : \mathcal{L}(X) = \mu$ (See Proposition 9.1.11 in Bogachev [11] or Proposition 9.1.2 and Theorem 13.1.1 in Dudley [20]).

Proposition 1.4. *We have*

$$\mathbf{d}_p(m,m') = \inf_{(X,X')} \mathbb{E}\left[|X-X'|^p\right]^{1/p},$$

where the infimum is computed over the pairs of random variables (X, X') in $\mathbb{R}^d$ such that $\mathcal{L}(X) = m$ and $\mathcal{L}(X') = m'$.

1.3 The Glivenko-Cantelli law of large numbers

Let (X_n) be an i.i.d. sequence of random variable in $L^1(\Omega)$ (where $(\Omega, \mathbb{F}, \mathbb{P})$ is a probability space). Let

$$m^N := \frac{1}{N} \sum_{n=1}^N \delta_{X_n}$$

be the associated sequence of empirical measures. Let m be the common law of the X_n. If $m \in \mathcal{P}_2(\mathbb{R}^d)$, the Glivenko-Cantelli law of large numbers states that the (m^N) converges almost surely to m:

$$\mathbb{P}\left[\lim_{N\to+\infty} \mathbf{d}_2(m^N, m) = 0\right] = 1$$

and one also easily checks that

$$\mathbb{E}\left[\ \mathbf{d}_2^2(m^N, m)\right] = 0.$$

Under additional conditions this convergence can be quantified: let

$$M_q(m) := \int_{\mathbb{R}^d} |x|^q m(dx).$$

Theorem 1.5 (see [18])**.** *If $m \in \mathcal{P}_q(\mathbb{R}^d)$ for some $q > 4$, there there exists a constant $C = C(d, q, M_q(m))$ such that*

$$\mathbb{E}\left[\mathbf{d}_2^2(m^N, m)\right] \leqslant \begin{cases} N^{-1/2} & \text{if } d \leqslant 3 \\ N^{-1/2}\ln(N) & \text{if } d = 3 \\ N^{-2/d} & \text{if } d > 4. \end{cases}$$

2 Differentiability of maps on the Wasserstein space

In this section, we discuss different notions of derivatives in the space of probability measures and explain how they are related. This part is, to a large extend, borrowed from [18].

2.1 The flat derivative

Definition 2.1. *Let $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. We say that U is of class C^1 if there exists a jointly continuous and bounded map $\frac{\delta U}{\delta m} : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ such that*

$$U(m') - U(m) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}((1-h)m + hm', y)(m' - m)(dy)dh \qquad \forall m, m' \in \mathcal{P}_2(\mathbb{R}^d).$$

Moreover we adopt the normalization convention

$$\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m, y)m(dy) = 0 \qquad \forall m \in \mathcal{P}_2(\mathbb{R}^d). \tag{1}$$

Note that, if U is of class C^1, then the following equality holds for any $m \in \mathcal{P}_2(\mathbb{R}^d)$ and $y \in \mathbb{R}^d$

$$\frac{\delta U}{\delta m}(m, y) = \lim_{h\to 0^+} \frac{1}{h}\left(U((1-h)m + h\delta_y) - U(m)\right).$$

Here is a kind of converse.

Proposition 2.2. *Let $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and assume that the limit*

$$V(m, y) := \lim_{h\to 0^+} \frac{1}{h}\left(U((1-h)m + h\delta_y) - U(m)\right)$$

exists and is jointly continuous and bounded on $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$. Then U is C^1 and $\frac{\delta U}{\delta m}(m, y) = V(m, y)$.

Proof. Although the result can be expected, the proof is a little involved and can be found in [17]. □

Let us recall that, if $\phi : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel measurable map and m is a Borel probability measure on $\mathbb{R}^d$, the image of m by ϕ is the Borel probability measure $\phi\sharp m$ defined by

$$\int_{\mathbb{R}^d} f(x)\phi\sharp m(dx) = \int_{\mathbb{R}^d} f(\phi(y))m(dy) \qquad \forall f \in C^0_b(\mathbb{R}^d).$$

Proposition 2.3. *Let U be C^1 and be such that $D_y \frac{\delta U}{\delta m}$ exists and is jointly continuous and bounded on $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$. Then, for any Borel measurable map $\phi : \mathbb{R}^d \to \mathbb{R}^d$ with at most a linear growth, the map $s \to U((id_{\mathbb{R}^d} + s\phi)\sharp m)$ is differentiable at 0 and*

$$\frac{d}{ds}_{|s=0} U((id_{\mathbb{R}^d} + s\phi)\sharp m) = \int_{\mathbb{R}^d} D_y \frac{\delta U}{\delta m}(m, y) \cdot \phi(y)m(dy).$$

Proof. Indeed

$$\begin{aligned} U((id_{\mathbb{R}^d} + s\phi)\sharp m) - U(m) &= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_{h,s}, y)((id_{\mathbb{R}^d} + s\phi)\sharp m) - m)(dy)dh \\ &= \int_0^1 \int_{\mathbb{R}^d} (\frac{\delta U}{\delta m}(m_{h,s}, y + s\phi(y)) - \frac{\delta U}{\delta m}(m_{h,s}, y))m(dy)dh \\ &= s\int_0^1 \int_0^1 \int_{\mathbb{R}^d} D_y \frac{\delta U}{\delta m}(m_{h,s}, y + s\tau\phi(y)) \cdot \phi(y)m(dy)dhd\tau, \end{aligned}$$

where

$$m_{h,s} = (1 - h)m + h(id_{\mathbb{R}^d} + s\phi)\sharp m.$$

Dividing by s and letting $s \to 0^+$ gives the desired result. □

Let us recall that, if $m, m' \in \mathcal{P}_2(\mathbb{R}^d)$, the set $\Pi^{opt}(m, m')$ denotes the set of optimal transport plans between m and m' (see Proposition 1.2).

Proposition 2.4. *Under the assumptions of the previous Proposition, let $m, m' \in \mathcal{P}_2(\mathbb{R}^d)$ and $\pi \in \Pi^{opt}(m, m')$. Then*

$$\left| U(m') - U(m) - \int_{\mathbb{R}^{2d}} D_y \frac{\delta U}{\delta m}(m, x) \cdot (y - x)\pi(dx, dy) \right| \leqslant o(\mathbf{d}_2(m, m')).$$

Remark 2.5. The same proof shows that, of π is a transport plan between m and m' (not necessarily optimal), then

$$\left| U(m') - U(m) - \int_{\mathbb{R}^{2d}} D_y \frac{\delta U}{\delta m}(m, x) \cdot (y - x)\pi(dx, dy) \right| \leqslant o\left(\left(\int_{\mathbb{R}^{2d}} |x - y|^2 \pi(dx, dy) \right)^{1/2} \right).$$

Proof. Let $\phi_t(x, y) = (1 - t)x + ty$ and $m_t = \phi_t\sharp\pi$. Then $m_0 = m$ and $m_1 = m'$ and, for any $t \in (0, 1)$ and any s small we have

$$\begin{aligned} U(\phi_{t+s}\sharp\pi) - U(\phi_t\sharp\pi) &= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_{s,h}, y)(\phi_{t+s}\sharp\pi - \phi_t\sharp\pi)(dy)dh \\ &= \int_0^1 \int_{\mathbb{R}^{2d}} \frac{\delta U}{\delta m}(m_{s,h}, (1 - t - s)x + (t + s)y) - \frac{\delta U}{\delta m}(m_{s,h}, (1 - t)x + ty)\ \pi(dx, dy)dh \\ &= s\int_0^1 \int_0^1 \int_{\mathbb{R}^{2d}} D_y \frac{\delta U}{\delta m}(m_{s,h}, (1 - t - \tau s)x + (t + \tau s)y) \cdot (y - x)\ \pi(dx, dy)dhd\tau, \end{aligned}$$

where $m_{s,h} = (1-h)\phi_{t+s}\sharp\pi + h\phi_t\sharp\pi$. So, dividing by s and letting $s \to 0$ we find:

$$\frac{d}{dt}U(\phi_t\sharp\pi) = \int_{\mathbb{R}^{2d}} D_y\frac{\delta U}{\delta m}(\phi_t\sharp\pi, (1-t)x+ty)\cdot(y-x)\,\pi(dx,dy).$$

As $D_y\frac{\delta U}{\delta m}$ is continuous and bounded by C, for any $\epsilon, R > 0$, there exists $r > 0$ such that, if $\mathbf{d}_2(m,m') \leqslant r$ and $|x|, |y| \leqslant R$, then

$$|D_y\frac{\delta U}{\delta m}(\phi_t\sharp\pi, (1-t)x+ty) - D_y\frac{\delta U}{\delta m}(m,x)| \leqslant \epsilon + 2C\mathbf{1}_{|y-x|\geqslant r}.$$

So

$$\begin{aligned}
&\left|\int_{\mathbb{R}^{2d}} D_y\frac{\delta U}{\delta m}(\phi_t\sharp\pi, (1-t)x+ty)\cdot(y-x)\,\pi(dx,dy) - \int_{\mathbb{R}^{2d}} D_y\frac{\delta U}{\delta m}(m,x)\cdot(y-x)\pi(dx,dy)\right| \\
&\leqslant \delta_R + \int_{(B_R)^2}(\epsilon + 2C\mathbf{1}_{|x-y|\geqslant r})|y-x|\pi(dx,dy) \leqslant \delta_R + \epsilon\mathbf{d}_2(m,m') + \frac{2C}{r}\mathbf{d}_2^2(m,m').
\end{aligned}$$

where

$$\begin{aligned}
\delta_R &:= \int_{\mathbb{R}^{2d}\backslash(B_R)^2} |D_y\frac{\delta U}{\delta m}(\phi_t\sharp\pi, (1-t)x+ty)\cdot(y-x)| + |D_y\frac{\delta U}{\delta m}(m,x)\cdot(y-x)|\pi(dx,dy) \\
&\leqslant C\int_{\mathbb{R}^{2d}\backslash(B_R)^2} |y-x|\pi(dx,dy) \leqslant C\mathbf{d}_2(m,m')\pi^{1/2}(\mathbb{R}^{2d}\backslash(B_R)^2) = \mathbf{d}_2(m,m')o_R(1).
\end{aligned}$$

This proves the result. □

2.2 $W-$differentiability

Next we turn to a more geometric definition of derivative in the space of measure. For this, let us introduce the notion of tangent space to $\mathcal{P}_2(\mathbb{R}^d)$.

Definition 2.6 (Tangent space)**.** *The tangent space* $\mathrm{Tan}_m(\mathcal{P}_2(\mathbb{R}^d))$ *of* $\mathcal{P}_2(\mathbb{R}^d)$ *at* $m \in \mathcal{P}_2(\mathbb{R}^d)$ *is the closure in* $L^2_m(\mathbb{R}^d)$ *of* $\{D\phi,\ \phi \in C_c^\infty(\mathbb{R}^d)\}$.

Following [3] we define the super and the subdifferential of a map defined on $\mathcal{P}_2(\mathbb{R}^d)$:

Definition 2.7. *Let* $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, $m \in \mathcal{P}_2(\mathbb{R}^d)$ *and* $\xi \in L^2_m(\mathbb{R}^d, \mathbb{R}^d)$. *We say that* ξ *belongs to the superdifferential* $\partial^+U(m)$ *to* U *at* m *if, for any* $m' \in \mathcal{P}_2(\mathbb{R}^d)$ *and any transport plan* π *from* m *to* m',

$$U(m') \leqslant U(m) + \int_{\mathbb{R}^d\times\mathbb{R}^d} \xi(x)\cdot(y-x)\pi(dx,dy) + o\left(\left(\int_{\mathbb{R}^{2d}} |x-y|^2\pi(dx,dy)\right)^{1/2}\right).$$

We say that ξ *belongs to the subdifferential* $\partial^-U(m)$ *to* U *at* m *if* $-\xi$ *belongs to* $D^+(-U)(m)$. *Finally, we say that the map* U *is* $W-$*differentiable at* m *if* $\partial^+U(m) \cap \partial^-U(m)$ *is not empty.*

One easily checks the following:

Proposition 2.8. *If* U *is* $W-$*differentiable at* m, *then* $\partial^+U(m)$ *and* $\partial^-U(m)$ *are equal and reduce to a singleton, denoted* $\{D_mU(m,\cdot)\}$.

Remark 2.9. On can actually check that $D_mU(m,\cdot)$ belongs to $\mathrm{Tan}_m(\mathcal{P}_2(\mathbb{R}^d))$.

Proof. Let $\xi_1 \in D^+U(m)$ and $\xi_2 \in D^-U(m)$. We have, for any $m' \in \mathcal{P}_2(\mathbb{R}^d)$ and any transport plan π from m to m',

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \xi_2(x) \cdot (y-x)\pi(dx,dy) + o\left(\left(\int_{\mathbb{R}^{2d}} |x-y|^2 \pi(dx,dy)\right)^{1/2}\right)$$
$$\leqslant U(m') - U(m) \leqslant \int_{\mathbb{R}^d \times \mathbb{R}^d} \xi_1(x) \cdot (y-x)\pi(dx,dy) + o\left(\left(\int_{\mathbb{R}^{2d}} |x-y|^2 \pi(dx,dy)\right)^{1/2}\right).$$

In particular, if we choose $m' = (1+h\phi)\sharp m$ and $\pi = (Id, Id + h\phi)\sharp m$ for some $\phi \in L^2_m(\mathbb{R}^d, \mathbb{R}^d)$ and $h > 0$ small, we obtain

$$h \int_{\mathbb{R}^d} \xi_2(x) \cdot \phi(x) m(dx) + o(h) \leqslant U(m') - U(m) \leqslant h \int_{\mathbb{R}^d} \xi_1(x) \cdot \phi(x) m(dx) + o(h),$$

from which we easily infer that $\xi_1 = \xi_2$ in $L^2_m(\mathbb{R}^d)$. □

We have seen in Remark 2.5 that if U is C^1 with $D_y \delta U/\delta m$ continuous and bounded, then U is W-differentiable. In this case it is obvious that $D_m U(m)$ belongs to $Tan_m(\mathcal{P}_2(\mathbb{R}^d))$ by definition.

2.3 Link with the $L-$ derivative

Another possibility for the notion of derivative is to look at the set of Borel probability measures as the law of random variables with values in $\mathbb{R}^d$ and to use the fact that this set has a Hilbertian structure.

Let $(\Omega, \mathbb{F}, \mathbb{P})$ an atomless probability space. Given a map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, we consider its extension $\tilde{U}$ to the set of random variables $L^2(\Omega, \mathbb{R}^d)$:

$$\tilde{U}(X) = U(\mathcal{L}(X)) \qquad \forall X \in L^2(\Omega, \mathbb{R}^d).$$

(recall that $\mathcal{L}(X)$ is the law of X, i.e., $\mathcal{L}(X) := X\sharp\mathbb{P}$. Note that $\mathcal{L}(X)$ belongs to $\mathcal{P}_2(\mathbb{R}^d)$ because $X \in L^2(\Omega)$). The important point is that $L^2(\Omega, \mathbb{F}, \mathbb{P})$ is a Hilbert space, in which the notion of Frechet differentiability makes sense.

For instance, if U is a map of the form

$$U(m) = \int_{\mathbb{R}^d} \phi(x) m(dx) \qquad \forall m \in \mathcal{P}_2(\mathbb{R}^d), \tag{2}$$

where $\phi \in C^0_c(\mathbb{R}^d)$ is given, then

$$\tilde{U}(X) = \mathbb{E}[\phi(X)] \qquad \forall X \in L^2(\Omega, \mathbb{R}^d).$$

Definition 2.10. *The map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is $L-$differentiable at $m \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists $X \in L^2(\Omega, \mathbb{R}^d)$ such that $\mathcal{L}(X) = m$ and the extension $\tilde{U}$ of U is Frechet differentiable at X.*

The following result says that the notion of $L-$differentiability coincides with that of $W-$differentiability and is independent of the probability space and of the representative X. The first statement in that direction goes back to Lions [32], the version given here is can be found in [26] (see also [1], from which the sketch of proof of Lemma 2.13 is largely inspired).

Theorem 2.11. *The map U is $W-$differentiable at $m \in \mathcal{P}_2(\mathbb{R}^d)$ if and only if U is $L-$differentiable at some (or thus any) $X \in L^2(\Omega, \mathbb{R}^d)$ with $\mathcal{L}(X) = m$. In this case*

$$\nabla \tilde{U}(X) = D_m U(m, X).$$

The result can be considered as a structure theorem for the L-derivative.

For instance, if U is as in (2) for some map $\phi \in C^1_c(\mathbb{R}^d)$, then it is almost obvious that

$$\nabla \tilde{U}(X) = D\phi(X)$$

and thus

$$D_m U(m, x) = D\phi(x).$$

The proof of Theorem 2.11 is difficult and we only sketch it briefly. Complete proofs can be found in [26] or [1]. The first step is the fact that, if X and X' have the same law, then so do $\nabla U(X)$ and $\nabla U(X')$:

Lemma 2.12. *Let $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $\tilde{U}$ be its extension. Let X, X' be two random variables in $L^2(\Omega, \mathbb{R}^d)$ with $\mathcal{L}(X) = \mathcal{L}(X')$. If $\tilde{U}$ is Frechet differentiable at X, then $\tilde{U}$ is differentiable at X' and $(X, \nabla U(X))$ has the same law as $(X', \nabla U(X'))$.*

(Sketch of) proof. The idea behind this fact is that, if X and X' have the same law, then one can "almost" find a bi-measurable and measure-preserving transformation $\tau : \Omega \to \Omega$ such that $X' = X \circ \tau$. Admitting this statement for a while, we have, for any $H' \in L^2$ small,

$$\begin{aligned}\tilde{U}(X' + H') &= \tilde{U}((X' + H') \circ \tau) = \tilde{U}(X + H' \circ \tau) = \tilde{U}(X) + \mathbb{E}\left[\nabla \tilde{U}(X) \cdot H' \circ \tau\right] + o(\|H' \circ \tau\|_2) \\ &= \tilde{U}(X') + \mathbb{E}\left[\nabla \tilde{U}(X) \circ \tau^{-1} \cdot H'\right] + o(\|H'\|_2).\end{aligned}$$

This shows that $\tilde{U}$ is differentiable at X' with differential given by $\nabla \tilde{U}(X) \circ \tau^{-1}$. Thus $(X', \nabla \tilde{U}(X')) = (X, \nabla \tilde{U}(X)) \circ \tau^{-1}$, which shows that $(X, \nabla U(X))$ and $(X', \nabla U(X'))$ have the same law.

In fact the existence of τ does not hold in general. However, one can show that, for any $\epsilon > 0$, there exists $\tau : \Omega \to \Omega$ bi-measurable and measure preserving and such that $\|X' - X \circ \tau\|_\infty \leqslant \epsilon$. A (slightly technical) adaptation of the proof above then gives the result (see [14] or [18] for the details). □

Next we show that $\nabla \tilde{U}(X)$ is a function of X:

Lemma 2.13. *Assume that $\tilde{U}$ is differentiable at $X \in L^2(\Omega, \mathbb{R}^d)$. Then there exists a Borel measurable map $g : \mathbb{R}^d \to \mathbb{R}^d$ such that $\nabla \tilde{U}(X) = g(X)$ a.s..*

(Sketch of) proof. To prove the claim, we just need to check that $\nabla \tilde{U}(X)$ is $\sigma(X)-$measurable (see Theorem 20.1 in [10]), which can be recasted into the fact that $\nabla \tilde{U}(X) = \mathbb{E}\left[\nabla \tilde{U}(X) | X\right]$. Let $\mu = \mathcal{L}(X, \nabla \tilde{U}(X))$ and let $\mu(dx, dy) = (\delta_x \otimes \nu_x(dy))\mathbb{P}_X(dx)$ be its disintegration with respect to its first marginal $\mathbb{P}_X$. Let λ be the restriction of the Lebesgue measure to $Q_1 := [0, 1]^d$. Then, as λ has an L^1 density, the optimal transport from λ to ν_x is unique and given by the gradient of a convex map $\psi_x(\cdot)$ (Brenier's Theorem, see [37]). So we can find[1] a measurable map $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ such that, for $\mathbb{P}_X-$a.e. $x \in \mathbb{R}^d$, $\psi_x(\cdot)\sharp\lambda = \nu_x$. Let Z be a random variable with law λ and independent of $(X, \nabla \tilde{U}(X))$.

[1]Warning: here the proof is sloppy and the possibility of a measurable selection should be justified.

Note that $\mu = \mathcal{L}(X, \nabla\tilde{U}(X)) = \mathcal{L}(X, \psi_X(Z))$ because, for any $f \in C_b^0(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\begin{aligned}\mathbb{E}\left[f(X, \psi_X(Z))\right] &= \int_{\mathbb{R}^d}\int_{Q_1} f(x, \psi_x(z))\lambda(dz)\mathbb{P}_X(dx) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x,y)(\psi_x\sharp\lambda)(dy)\mathbb{P}_X(dx) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x,y))\nu_x(dy)\mathbb{P}_X(dx) = \int_{\mathbb{R}^{2d}} f(x,y)\mu(dx,dy).\end{aligned}$$

So, for any ϵ,

$$\tilde{U}(X + \epsilon\nabla\tilde{U}(X)) = \tilde{U}(X + \epsilon\psi_X(Z)),$$

from which we infer, taking the derivative with respect to ϵ at $\epsilon = 0$:

$$\mathbb{E}\left[\left|\nabla\tilde{U}(X)\right|^2\right] = \mathbb{E}\left[\nabla\tilde{U}(X)\cdot\psi_X(Z)\right].$$

Note that, as Z is independent of $(X, \nabla\tilde{U}(X))$, we have

$$\mathbb{E}\left[\nabla\tilde{U}(X)\cdot\psi_X(Z)\right] = \mathbb{E}\left[\nabla\tilde{U}(X)\cdot\mathbb{E}\left[\psi_x(Z)\right]_{x=X}\right],$$

where, for $\mathbb{P}_X-$a.e. x,

$$\mathbb{E}\left[\psi_x(Z)\right] = \int_{Q_1}\psi_x(z)\lambda(dz) = \int_{Q_1} y\ (\psi_x\sharp\lambda)(dy) = \int_{\mathbb{R}^d} y\ \nu_x(dy) = \mathbb{E}\left[\nabla\tilde{U}(X)|X = x\right].$$

So, by the tower property of the condition expectation, we have

$$\mathbb{E}\left[\left|\nabla\tilde{U}(X)\right|^2\right] = \mathbb{E}\left[\nabla\tilde{U}(X)\cdot\mathbb{E}\left[\nabla\tilde{U}(X)|X\right]\right] = \mathbb{E}\left[\left|\mathbb{E}\left[\nabla\tilde{U}(X)|X\right]\right|^2\right].$$

Using again standard properties of the conditional expectation we infer the equality $\nabla\tilde{U}(X) = \mathbb{E}\left[\nabla\tilde{U}(X)|X\right]$, which shows the result. □

Proof of Theorem 2.11. Let us first assume that U is $W-$differentiable at some $m \in \mathcal{P}_2(\mathbb{R}^d)$. Then there exists $\xi := D_m U(m,\cdot) \in L^2_m(\mathbb{R}^d)$ such that, for any $m' \in \mathcal{P}_2(\mathbb{R}^d)$ and any transport plan π between m and m' we have

$$\left|U(m') - U(m) - \int_{\mathbb{R}^d\times\mathbb{R}^d}\xi(x)\cdot(y-x)\pi(dx,dy)\right| \leqslant o\left(\left(\int_{\mathbb{R}^{2d}}|x-y|^2\pi(dx,dy)\right)^{1/2}\right).$$

Therefore, for any $X \in L^2$ such that $\mathcal{L}(X) = m$, for any $H \in L^2$, if we denote by m' the law of $X + H$ and by π the law of $(X, X+H)$, we have

$$\begin{aligned}\left|\ \tilde{U}(X+H) - \tilde{U}(X) - \mathbb{E}\left[\xi(X)\cdot H\right]\right| &= \left|\ U(m') - \tilde{U}(m) - \int_{\mathbb{R}^{2d}}\xi(x)\cdot(y-x)\pi(x,y)\right| \\ &\leqslant o\left(\left(\int_{\mathbb{R}^{2d}}|x-y|^2\pi(dx,dy)\right)^{1/2}\right)) \\ &= o\left(\mathbb{E}\left[|X-Y|^2\right]^{1/2}\right).\end{aligned}$$

This shows that U is L-differentiable.

Conversely, let us assume that U is L-differentiable at m. We know from Lemma 2.13 that, for any $X \in L^2$ such that $\mathcal{L}(X) = m$, $\tilde{U}$ is differentiable at X and $\nabla\tilde{U}(X) = \xi(X)$ for some Borel measurable map $\xi : \mathbb{R}^d \to \mathbb{R}^d$. In view of Lemma 2.12, the map ξ does not depend on the choice of X. So, for any $\epsilon > 0$, there exists $r > 0$ such that, for any X with $\mathcal{L}(X) = m$ and any $H \in L^2$ with $\|H\| \leqslant r$, one has

$$\left| \tilde{U}(X+H) - \tilde{U}(X) - \mathbb{E}\left[\xi(X)\cdot H\right]\right| \leqslant \epsilon.$$

Let now $m' \in \mathcal{P}_2(\mathbb{R}^d)$ and π be a transport plan between m and m' such that $\int_{\mathbb{R}^{2d}} |x-y|^2\pi(dx,dy) \leqslant r^2$. Let (X,Y) with law π. We set $H = Y - X$ and note that $\|H\|_2 \leqslant r$. So we have

$$\left| U(m') - \tilde{U}(m) - \int_{\mathbb{R}^{2d}} \xi(x)\cdot(y-x)\pi(x,y)\right| = \left| \tilde{U}(X+H) - \tilde{U}(X) - \mathbb{E}\left[\xi(X)\cdot H\right]\right| \leqslant \epsilon.$$

This proves the W-differentiability of U. □

2.4 Higher order derivatives

We say that U is partially C^2 if U is C^1 and if $D_y\delta U/\delta m$ and $D^2_{yy}\delta U/\delta m$ exist and are continuous and bounded on $\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d$.

We say that U is C^2 if $\frac{\delta U}{\delta m}$ is C^1 in m with a continuous and bounded derivative: namely $\frac{\delta^2 U}{\delta m^2} = \frac{\delta}{\delta m}(\frac{\delta U}{\delta m}) : \mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$ is continuous in all variables and bounded. We say that U is twice $L-$differentiable if the map D_mU is $L-$differentiable with respect to m with a second order derivative $D^2_{mm}U = D^2_{mm}U(m,y,y')$ which is continuous and bounded on $\mathcal{P}_2(\mathbb{R}^d)\times\mathbb{R}^d\times\mathbb{R}^d$ with values in $\mathbb{R}^{d\times d}$. One can check that this second order derivative enjoys standard properties of derivatives, such as the symmetry:

$$D^2_{mm}U(m,y,y') = D^2_{mm}U(m,y',y).$$

See [16].

2.5 To go further

For a general description of the notion of derivatives and the historical background, we refer to [18], Chap V. The notion of flat derivative is very natural and has been introduced in several contexts and under various assumptions. We follow here [16].

The initial definition of sub and super differential in the space $\mathcal{P}_2(\mathbb{R}^d)$, introduced in [2], is the following: ξ belongs to $\partial^+U(m)$ if $\xi \in Tan_m(\mathcal{P}_2(\mathbb{R}^d))$ and

$$U(m') \leqslant U(m) + \inf_{\pi\in\Pi^{opt}(m,m')} \int_{\mathbb{R}^d\times\mathbb{R}^d} \xi(x)\cdot(y-x)\pi(dx,dy) + o(\mathbf{d}_2(m,m')).$$

It is proved in [26] that this definition coincides with the one introduced in Definition 2.7.

The notion of L-derivative and the structure of this derivative has been first discussed by Lions in [32] (see also [14] for a proof in which the function is supposed to be continuously differentiable). The proof, without the extra continuity condition, of this structure property is due to Gangbo and Tudorascu [26] (see also [1] for simpler arguments, revisited here in a loose way).

3 Calculus on the Wasserstein space

In this part we collect several useful tools in relation with the calculus in $\mathcal{P}_2(\mathbb{R}^d)$: optimality conditions, projection over empirical measures and various Itô's formulas.

3.1 Optimality conditions

This part is borrowed from [17].

Proposition 3.1 (First order conditions)**.** *Assume that the map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ has a maximum at some measure $m \in \mathcal{P}_2(\mathbb{R}^d)$. If U has a linear derivative, then*

$$\frac{\delta U}{\delta m}(\bar{m}, y) \leqslant 0 \ \forall y \in \mathbb{R}^d \qquad \text{and} \qquad \frac{\delta U}{\delta m}(\bar{m}, y) = 0 \text{ for } \bar{m}\text{−a.e. } y \in \mathbb{R}^d. \tag{3}$$

In particular, if U is partially C^2, then

$$D_m U(m, y) = 0 \text{ and } D^2 D_m U(m, y) \geqslant 0 \qquad \text{for } \bar{m}\text{−a.e. } y \in \mathbb{R}^d \tag{4}$$

An example: Assume that $U(m) = \displaystyle\int_{\mathbb{R}^d} \phi(y) dm(y)$ where $\phi : \mathbb{R}^d \to \mathbb{R}$ is a smooth map such that $\phi(0) = 0$, $\phi < 0$ outside 0 and $D^2\phi(0) < 0$. Then U has a maximum at $\hat{m} = \delta_0$ (the Dirac mass at 0). As

$$\frac{\delta U}{\delta m}(m, y) = \phi(y) - \int_{\mathbb{R}^d} \phi(y') dm(y'),$$

we have

$$\frac{\delta U}{\delta m}(\hat{m}, y) = \phi(y), \ D_m U(\hat{m}, y) = D\phi(y), \ D_{ym} U(\hat{m}, y) = D^2\phi(y).$$

In particular, $\frac{\delta U}{\delta m}(\hat{m}, y)$ and $D_m U(\hat{m}, y)$ vanish only on the support of $\hat{m} = \delta_0$ and $D_{ym} U(\delta_0, y)$ is negative on the support of $\hat{m}$.

Proof. For any $y_0 \in \mathbb{R}^d$ and $s \in (0, 1]$, we have, by then optimality of $\bar{m}$,

$$U((1-s)\bar{m} + s\delta_{y_0}) - U(\bar{m}) = \int_0^s \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}((1-t)\bar{m} + t\delta_{y_0}, y)(\delta_{y_0} - \bar{m})(dy) \leqslant 0.$$

We divide by s and let $s \to 0^+$. By the continuity of $\delta U / \delta m$ we obtain

$$\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\bar{m}, y)(\delta_{y_0} - \bar{m})(dy) \leqslant 0,$$

which implies that, for any $y_0 \in \mathbb{R}^d$ and because of the convention in (1),

$$\frac{\delta U}{\delta m}(\bar{m}, y_0) \leqslant 0. \tag{5}$$

If we integrate this inequality in y_0 against $\bar{m}$, we find an equality (again because of the convention in (1)). So

$$\frac{\delta U}{\delta m}(\bar{m}, y_0) = 0 \qquad \text{for } \bar{m}\text{−a.e. } y_0 \in \mathbb{R}^d. \tag{6}$$

Then (5) and (6) imply that the map $y \to \frac{\delta U}{\delta m}(\bar{m}, y)$ has a maximum at $\bar{m}$−a.e. $y_0 \in \mathbb{R}^d$, and thus, using the standard optimality conditions for a map defined on $\mathbb{R}^d$, we obtain (4).

□

Proposition 3.2 (Second order condition)**.** *Assume that the map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ has a maximum at some measure $\bar m \in \mathcal{P}_2(\mathbb{R}^d)$ and that U is fully C^2. Then, for any bounded Borel map $\phi : \mathbb{R}^d \to \mathbb{R}^d$, we have*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} D^2_{mm}U(\bar m, y, y')\phi(y)\cdot\phi(y')\bar m(dy)\bar m(dy') \leqslant 0. \tag{7}$$

Proof. We fix a Borel measurable, bounded vector field $\phi : \mathbb{R}^d \to \mathbb{R}^d$ and note that, for any $s \in \mathbb{R}$ and by optimality of $\bar m$,

$$U((id + s\phi)\sharp\bar m) \leqslant U(\bar m).$$

Therefore

$$\frac{d}{ds}_{|s=0} U((id + s\phi)\sharp\bar m) = 0 \text{ and } \frac{d^2}{ds^2}_{|s=0} U((id + s\phi)\sharp\bar m) \leqslant 0.$$

As

$$\frac{d}{ds}U((id + s\phi)\sharp\bar m) = \int_{\mathbb{R}^d} D_m U((id + s\phi)\sharp\bar m, y)\cdot\phi(y)\bar m(dy),$$

we have

$$\frac{d^2}{ds^2}_{|s=0} U((id + s\phi)\sharp\bar m) = \int_{\mathbb{R}^{2d}} D^2_{mm}U(y, y')\phi(y)\cdot\phi(y')\bar m(dy)\bar m(dy'),$$

where U and its derivatives are computed at $\bar m$. This gives (7). $\square$

3.2 Projection over finite dimensional spaces

It is often very convenient to translate results known on a map defined on the Wasserstein space to its restriction to the set of empirical measures. The computations here come from [16].

Given $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and $N \in \mathbb{N}$ be a large integer, we set

$$U^N(x_1, \dots, x_N) = U(m^N_{\mathbf{x}}), \text{ where } \mathbf{x} = (x_1, \dots, x_N),\ m^N_{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^N \delta_{x_n}.$$

Note that $U^N : (\mathbb{R}^d)^N \to \mathbb{R}$.

Proposition 3.3. *Assume that $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is L-continuous differentiable with bounded derivative. Then U^N is of class C^1 on $(\mathbb{R}^d)^N$ and*

$$D_{x_i}U^N(\mathbf{x}) = \frac{1}{N}D_m U(m^N_{\mathbf{x}}, x_i) \qquad \forall i \in \{1, \dots, N\}.$$

If in addition U is fully C^2, then

$$D_{x_i x_j}U^N(\mathbf{x}) = \frac{1}{N^2}D^2_{mm}U(m, x_i, x_j) \qquad \forall i \neq j \in \{1, \dots, N\}$$

and

$$D_{x_i x_i}U^N(\mathbf{x}) = \frac{1}{N^2}D^2_{mm}U(m, x_i, x_i) + \frac{1}{N}D_y D_m U(m^N_{\mathbf{x}}, x_i) \qquad \forall i \in \{1, \dots, N\}.$$

Proof. We first assume that the (x_i) are all distinct. Then, for any direction $\mathbf{v} = (v_1, \dots, v_N) \in (\mathbb{R}^d)^N$, we can find a smooth map $\phi : \mathbb{R}^d \to \mathbb{R}^d$ such that $\phi(x_i) = v_i$ for any i. Note that

$$(Id + h\phi)\sharp m^N_{\mathbf{x}} = m^N_{\mathbf{x}+h\mathbf{v}}$$

for any $h \in \mathbb{R}$, so that, by Proposition 2.3 we have, at $h = 0$,

$$\frac{d}{dh}U^N(\mathbf{x} + h\mathbf{v}) = \frac{d}{dh}\ U((Id + h\phi)\sharp m^N_{\mathbf{x}}) = \int_{\mathbb{R}^d} D_mU(m^N_{\mathbf{x}}, y) \cdot \phi(y) m^N_{\mathbf{x}}(dy).$$

So

$$\frac{d}{dh}U^N(\mathbf{x} + h\mathbf{v}) = \frac{1}{N}\sum_{i=1}^N D_mU(m^N_{\mathbf{x}}, x_i) \cdot \phi(x_i) = \frac{1}{N}\sum_{i=1}^N D_mU(m^N_{\mathbf{x}}, x_i) \cdot v_i.$$

This shows that U^N has continuous directional derivatives in the dense open set $\{\mathbf{x} \in (\mathbb{R}^d)^N,\ x_i \neq x_j \text{ if } i \neq j\}$. Therefore U^N is C^1 and the above formula holds. The second order differentiability can be proved by the same arguments. □

3.3 Itô's formula

Given a flow $(m(t))$ of probability measures satisfying a differential equation (typically a continuity equation, a Fokker-Planck equation, etc...) we consider the derivative of a map U along this flow.

3.3.1 First order Itô's formula

We start with a simple continuity equation:

$$\partial_t m + \mathrm{div}(mb) = 0.$$

Here we assume that the continuous drift $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is such that $b(t, \cdot)$ is uniformly (in t) Lipschitz continuous on $\mathbb{R}^d$. Then the Cauchy-Lipschitz Theorem states that there exists a unique solution $(X_t^{t_0,x_0})$ to

$$x'(t) = b(t, x(t)),\ t \in [0, T], \qquad x(t_0) = x_0.$$

We denote by $X_t^{x_0}$ the solution to this equation.

Given an initial measure $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$, the flow of measures $m(t) := X_t^{\cdot}\sharp m_0$ is a solution in the sense of distribution of

$$\partial_t m + \mathrm{div}(mb) = 0 \qquad \text{in } (0, T) \times \mathbb{R}^d, \qquad m(0) = m_0 \text{ in } \mathbb{R}^d.$$

One can show (but we admit this fact here, see [2]) that $(m(t))$ is the unique solution to this equation.

Proposition 3.4. *Let $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $m(t) = X_t^{t_0,\cdot}\sharp m_0$. If U is $L - C^1$, then $t \to U(m(t))$ is of class C^1 and*

$$\frac{d}{dt}U(m(t)) = \int_{\mathbb{R}^d} D_mU(m(t), y) \cdot b(t, y) m(t)(dy).$$

Proof. It is only a small variant of the proof of Proposition 2.3 and we omit it. □

3.3.2 Itô's formula for the law of SDEs (or the Fokker-Planck equation)

When U is partially C^2, Itô's formula holds for the law of a diffusion of the form

$$dX_t = b_t dt + \sigma_t dB_t.$$

The following result is [18, Theorem 5.98].

Theorem 3.5. *Assume that where (b_t) and (σ_t) are progressively measurable with values in $\mathbb{R}^d$ and $\mathbb{R}^{n\times d}$ respectively and satisfy, for some $T>0$,*

$$\mathbb{E}\left[\int_0^T (|b_s|^2 + |\sigma_s|^4)ds\right] < +\infty.$$

Assume in addition that U is partially C^2 with

$$\sup_{\mu\in\mathcal{P}_2(\mathbb{R}^d)} \int_{\mathbb{R}^d} \Big| D_y D_m U(\mu, y)\Big|^2 \mu(dy) < +\infty. \tag{8}$$

Then, for any $t\in[0,T]$,

$$U(\mathcal{L}(X_t)) = \mathcal{L}(X_0)) + \int_0^t \mathbb{E}\left[D_m U(\mathcal{L}(X_s), X_s)\cdot b_s\right] ds + \frac{1}{2}\int_0^t \mathbb{E}\left[\mathrm{Tr}\left(a_s D_y D_m U(\mathcal{L}(X_s), X_s)\right)\right] ds, \tag{9}$$

where $a_s = \sigma_s\sigma_s^*$.

Remark: if $b_s = b(s, X_s)$ and $\sigma_s = \sigma(s, X_s)$, the above expression becomes, if we set $m(t) =: \mathcal{L}(X_t)$,

$$\begin{aligned} U(m(t)) = U(m_0) &+ \int_0^t \int_{\mathbb{R}^d} D_m U(m(s), y)\cdot b(s,y) m(s,dy)ds \\ &+ \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} \mathrm{Tr}\left(a(s,y) D_y D_m U(m(s), y)\right) m(s,dy)ds. \end{aligned}$$

Sketch of proof in a particular case. For simplicity, we do the proof only in the case $a = \sqrt{2}Id$ and $b = b(s,x)$ is a smooth and globally Lipschitz continuous vector field. In this case it is known that the law $m(t)$ of X_t has smooth density which satisfies

$$\partial_t m - \Delta m + \mathrm{div}(mb) = 0.$$

Therefore

$$\begin{aligned} \frac{d}{dt}U(m(t)) &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m(t), y)\partial_t m(t,y)dy \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m(t), y)(\Delta m(t,y) + \mathrm{div}(m(t,y)b(t,y))dy \\ &= \int_{\mathbb{R}^d} D_y D_m U(m(t), y) m(t,y)dy - \int_{\mathbb{R}^d} D_m U(m(t), y)\cdot b(t,y)m(t,y)dy. \end{aligned}$$

$\square$

An alternative proof is to use the law of large numbers as described in Section 1.3. Let (X_t^i) be i.i.d. copies of the process X and let m_t^N be the associated empirical measure:

$$m_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}.$$

We know that m_t^N converges to the law of X_t in $\mathcal{P}_2(\mathbb{R}^d)$ (Theorem 1.5). On the other hand, if we consider the finite dimensional projection U^N of U:

$$U^N(\mathbf{x}) := U(m_{\mathbf{x}}^N), \forall \mathbf{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N,$$

then we can use Itô's calculus and the expressions of the derivatives of U^N in Proposition 3.3 to obtain the result: see [18], Chap. 5 for the details.

3.3.3 Itô's formula for the conditional law

A similar Itô's formula holds for the *conditional law* of an Itô's process of the form,

$$dX_t = b_t dt + \sigma_t^0 dB_t + \sigma_t^1 dW_t,$$

where B and W are independent $d-$dimensional Brownian motions living on different probability spaces $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ and $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ and where b, σ^0 and σ^1 are progressively measurable with respect to the filtration generated by W and B, with

$$\mathbb{E}\left[\int_0^T (|b_s|^2 + |\sigma_s^0|^4 + |\sigma_s^1|^4) ds\right] < +\infty.$$

We assume that U is globally C^2 with

$$\sup_{\mu\in\mathcal{P}_2(\mathbb{R}^d)} \int_{\mathbb{R}^d} |D_m U(\mu, x)|^2 \mu(dx) + \int_{\mathbb{R}^d} |D_x D_m U(\mu, x)|^2 \mu(dx) + \int_{\mathbb{R}^d\times\mathbb{R}^d} \left|D^2_{\mu\mu} U(\mu, x, y)\right|^2 \mu(dx)\mu(dy) < +\infty.$$

Then, letting $\mu_t(\omega^1) = [X_t|W](\omega^1)$, we have, $\mathbb{P}^1-$a.s.,

$$\begin{aligned} U(\mu_t) = U(\mu_0) &+ \int_0^t \mathbb{E}^0 \left[D_m U(\mu_s, X_s)\cdot b_s\right] ds + \int_0^t \mathbb{E}^0 \left[(\sigma_s^1)^* D_m U(\mu_s, X_s)\right]\cdot dW_s \\ &+ \frac{1}{2}\int_0^t \mathbb{E}^0 \left[\mathrm{Tr}\left(a_s D_y D_m U(\mu_s, X_s)\right))\right] ds + \frac{1}{2}\int_0^t \mathbb{E}^0\tilde{\mathbb{E}}^0 \left[\mathrm{Tr}\left(D^2_{\mu\mu} U(\mu_s, X_s, \tilde{X}_s)\sigma_s^1(\tilde{\sigma}_s^1)^*\right)\right] ds \end{aligned} \tag{10}$$

where $\tilde{X}$ and $\tilde{\sigma}^1$ are independent copies of X and σ^1 is defined on the space $(\tilde{\Omega}^0\times\tilde{\Omega}^1, \tilde{\mathbb{P}}^0\otimes\tilde{\mathbb{P}}^1)$, while $a_s := (\sigma_s^0(\sigma_s^0)^* + \sigma_s^1(\sigma_s^1)^*)$. See [18, Theorem 11.13].

4 Calculus of variation in the Wasserstein space

In this section we study the optimal control of the continuity equation, formally given by

$$\inf_\alpha \int_0^T \int_{\mathbb{R}^d} L(x, \alpha(t,x)) m(t,dx) dt + \int_0^T F(m(t)) dt + G(m(T))$$

where $\alpha = \alpha(t,x)$ is a distributed control and m solves the continuity equation

$$\partial_t m + \mathrm{div}(m\alpha) = 0 \text{ in } (0,T)\times\mathbb{R}^d, \qquad m(0) = m_0.$$

In the cost, T is a finite horizon, $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a continuous Lagrangian with a quadratic growth:

$$C_0^{-1}|\alpha|^2 - C_0 \leqslant L(x,\alpha) \leqslant C_0(|\alpha|^2+1)$$

and the maps $F, G : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ are assumed to be (at least) continuous and bounded on $\mathcal{P}_2(\mathbb{R}^d)$. To simplify the discussion, we also assume that L is (at least) of class C^1 with respect to p, so that

$$L(x,\alpha) = \sup_{b\in\mathbb{R}^d} D_\alpha L(x,b)\cdot(\alpha - b) + L(x,b) \tag{11}$$

4.1 Existence of a solution

In order to find a minimum to the above problem, we need to relax a little the formulation. Let $\mathcal{W}$ be the set of Borel vector measures on $(0,T)\times\mathbb{R}^d$ with values in $\mathbb{R}^d$. We say that a pair (m,w) is admissible if $m \in C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$, $w \in \mathcal{W}$, w is absolutely continuous with respect to the measure $dt \otimes m(t,dx)$ and the following equality holds in the sense of distributions:

$$\partial_t m + \mathrm{div}(w) = 0 \text{ in } (0,T)\times\mathbb{R}^d, \qquad m(0) = m_0.$$

If (m,w) is admissible, we denote by dw/dm the Radon-Nykodim derivative.

We consider the functional

$$J(m,w) := \int_0^T \int_{\mathbb{R}^d} L(x, \frac{dw}{dm}(t,x))m(t,dx)dt + \int_0^T F(m(t))dt + G(m(T)) \tag{12}$$

if (m,w) is admissible and $J(m,w) = +\infty$ otherwise.

Theorem 4.1. *We assume that $F, G : \mathcal{P}_{2-\delta}(\mathbb{R}^d) \to \mathbb{R}$ are continuous for some $\delta \in (0,2)$. Then there exists at least a pair (m,w) which minimizes J.*

Remark: The above result extends to more general Lagrangians of the form $L(x,\alpha,m)$ with a non-linear dependence with respect to m. If F and G are convex on $\mathcal{P}_2(\mathbb{R}^d)$, the proof shows also applies even if F and G are only continuous (or lsc) on $\mathcal{P}_2(\mathbb{R}^d)$. The above result is a (small) generalization of the famous Benamou-Brenier formulation of optimal transport [7].

Let us start with a remark on the regularity of the solution (m,w).

Lemma 4.2. *Assume that (m,w) is admissible. Then*

$$\mathbf{d}_2(m(t),m(s)) \leqslant (t-s)^{1/2}\left(\int_0^T\int_{\mathbb{R}^d}\left|\frac{dw}{dm}(t,x)\right|^2 m(t,dx)dt\right)^{1/2}$$

Proof. Let $\xi : \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}$ be a smooth positive kernel on $(-1,1)\times\mathbb{R}^d$, vanishing outside of this set and with integral 1. We set $\xi_\epsilon(t,x) = \epsilon^{-d-1}\xi(t/\epsilon, x/\epsilon)$. We also extend (m,w) by $(m(t),w) = (m_0,0)$ for $t \leqslant 0$ and $(m(t),w) = (m(T),0)$ for $t \geqslant T$. Then (m,w) still satisfies the continuity equation

$$\partial_t m + \mathrm{div}(w) = 0 \text{ in } \mathbb{R}\times\mathbb{R}^d$$

and therefore so does $(m^\epsilon, w^\epsilon) = \xi^\epsilon \star (m,w)$ because the equation is linear. Note that $m^\epsilon > 0$ on $(0,T)\times\mathbb{R}^d$ and we set $\alpha^\epsilon := w^\epsilon/m^\epsilon$. Then α^ϵ is a smooth vector field: let us denote $X_t^{\epsilon,s,x}$ the associated flow, i.e., the solution to

$$\left\{\begin{array}{l} \dot{X}_t^{\epsilon,s,x} = \alpha^\epsilon(t, X_t^{\epsilon,s,x}), \\ X_s^{\epsilon,s,x} = x \end{array}\right.$$

Let us fix $\epsilon \leqslant s \leqslant t \leqslant T-\epsilon$. Then $m^\epsilon(t) = (x \to X_t^{\epsilon,s,x})\sharp m(s)$, so that

$$\mathbf{d}_2^2(m^\epsilon(t), m^\epsilon(s)) \leqslant \int_{\mathbb{R}^{2d}} |x-y|^2 \pi(x,y),$$

where $\pi \in \Pi(m(t), m(s))$ is defined by $\pi = (x \to (X_t^{\epsilon,s,x}, x))\sharp m^\epsilon(s)$. Then

$$\begin{aligned}\mathbf{d}_2^2(m(t), m(s)) &\leqslant \int_{\mathbb{R}^{2d}} |X_t^{\epsilon,s,x} - x|^2 m^\epsilon(s,dx) \leqslant (t-s)\int_{\mathbb{R}^{2d}}\int_s^t |\alpha^\epsilon(\tau, X_\tau^{\epsilon,s,x})|^2 d\tau \; m^\epsilon(s,x)\\ &= (t-s)\int_\epsilon^{T-\epsilon}\int_{\mathbb{R}^{2d}} |\alpha^\epsilon(\tau,y)|^2 m^\epsilon(\tau,x)d\tau.\end{aligned}$$

By convexity of the map $(z,r) \to |z|^2/(2r)$ on $\mathbb{R}^d \times (0,+\infty)$ (since it is equal to $\sup_{y\in\mathbb{R}^d} y\cdot z - r|y|^2/2$) and since ξ^ϵ has a support in $(-\epsilon,\epsilon)\times\mathbb{R}^d$, we have

$$\begin{aligned}\int_\epsilon^{T-\epsilon}\int_{\mathbb{R}^{2d}} |\alpha^\epsilon(\tau,y)|^2 m^\epsilon(\tau,x)d\tau &= \int_\epsilon^{T-\epsilon}\int_{\mathbb{R}^{2d}} \left|\frac{w^\epsilon}{m^\epsilon}(\tau,y)\right|^2 m^\epsilon(\tau,x)d\tau\\ &\leqslant \int_0^T\int_{\mathbb{R}^{2d}}\left|\frac{dw}{dm}(\tau,x)\right|^2 m(\tau,dx)d\tau.\end{aligned}$$

So we have proved that

$$\mathbf{d}_2^2(m^\epsilon(t), m^\epsilon(s)) \leqslant (t-s)\int_0^T\int_{\mathbb{R}^{2d}}\left|\frac{dw}{dm}(\tau,x)\right|^2 m(\tau,dx)d\tau.$$

Then we let $\epsilon \to 0$ and obtain the result. □

Proof of Theorem 4.1. Choosing $w=0$ and $m(t)=m_0$, we see that the infimum is not $+\infty$. Let (m_n, w_n) be a minimizing sequence. From our growth assumption on L and the fact that F and G are bounded, we infer that

$$\int_0^T\int_{\mathbb{R}^d}\left|\frac{dw_n}{dt\otimes dm_n}\right|^2 m_n(t,dx)dt \leqslant C. \tag{13}$$

By Lemma 4.2 we have there that (m_n) is uniformly continuous in $\mathcal{P}_2(\mathbb{R}^d)$. Note that this also implies that

$$\sup_{n\in\mathbb{N}}\sup_{t\in[0,T]} M_2(m_n(t)) \leqslant C.$$

Hence we know that there exists a subsequence, still denoted by (m_n), such that (m_n) converge to some $m \in C^0([0,T],\mathcal{P}_2(\mathbb{R}^d))$ in $C^0([0,T],\mathcal{P}_{2-\delta}(\mathbb{R}^d))$.

On the other hand, the total variation $|w_n|$ of w_n is uniformly bounded because

$$|w_n|([0,T]\times\mathbb{R}^d) \leqslant \left(\int_0^T\int_{\mathbb{R}^d} m_n(t,dx)dt\right)^{1/2}\left(\int_0^T\int_{\mathbb{R}^d}\left|\frac{dw_n}{dm_n}(t,x)\right|^2 m_n(t,dx)dt\right)^{1/2} \leqslant T^{1/2}C^{1/2}.$$

So, extracting a further subsequence, we can assume that (w_n) converges in distribution to some vector measure w on $[0,T]\times\mathbb{R}^d$. It remains to check that (m,w) is admissible and a minimizer.

By (13) we have that, for any test function $\phi \in C_c^\infty([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$,

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} \phi(t,x) \cdot w(dt,dx) - \int_0^T \int_{\mathbb{R}^d} \frac{1}{2}|\phi(t,x)|^2 m(t,dx)dt \\
&\qquad = \lim_n \int_0^T \int_{\mathbb{R}^d} \phi(t,x) \cdot w_n(dt,dx) - \int_0^T \int_{\mathbb{R}^d} \frac{1}{2}|\phi(t,x)|^2 m_n(t,dx)dt \\
&\qquad \leqslant \limsup_n \int_0^T \int_{\mathbb{R}^d} \left| \frac{dw_n}{dt \otimes dm_n}(t,x) \right|^2 m_n(t,dx)dt \leqslant C.
\end{aligned}$$

This proves that w is absolutely continuous with respect to $dt \otimes dm$. Moreover, one easily checks that (m,w) satisfies the continuity equation. So (m,w) is admissible.

We finally prove that (m,w) is a minimizer. As F and G are continuous on $\mathcal{P}_{2-\delta}(\mathbb{R}^d)$, we have

$$\lim_n \int_0^T F(m_n(t))dt + G(m_n(T)) = \int_0^T F(m(t))dt + G(m(T)).$$

For any $a \in C_c^\infty(\mathbb{R}^d)$ and $b \in C_b^0(\mathbb{R})$ such that $a(x) \cdot p + b(x) \leqslant L(x,p)$ for any $x,p \in \mathbb{R}^d$, we have

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} a(x) \cdot w(dt,dx) + \int_0^T \int_{\mathbb{R}^d} b(x) m(t,dx)dt \\
&\qquad = \lim_n \int_0^T \int_{\mathbb{R}^d} a(x) \cdot w_n(dt,dx) + \int_0^T \int_{\mathbb{R}^d} b(x) m_n(t,dx)dt \\
&\qquad \leqslant \limsup_n \int_0^T \int_{\mathbb{R}^d} L(x, \frac{dw_n}{dt \otimes dm_n}(t,x)) m_n(t,dx)dt.
\end{aligned}$$

Taking the supremum with respect to a, b in the above inequality (recall (11)) proves that

$$\begin{aligned}
&\int_0^T \int_{\mathbb{R}^d} L(x, \frac{dw}{dt \otimes dm}(t,x)) m(t,dx)dt \\
&\qquad \leqslant \limsup_n \int_0^T \int_{\mathbb{R}^d} L(x, \frac{dw_n}{dt \otimes dm_n}(t,x)) m_n(t,dx)dt.
\end{aligned}$$

This shows that

$$J(m,w) \leqslant \limsup_n J(m_n, w_n) = \inf J$$

and therefore that (m,w) is a minimum of the problem. □

4.2 Necessary conditions

In this section we write some necessary conditions for our optimal control problem. For this, in addition to the assumptions of the previous part, we assume here that F and G are of class C^1 (with respect to m) and that the derivatives

$$f(x,m) = \frac{\delta F}{\delta m}(m,x), \qquad g(x,m) = \frac{\delta G}{\delta m}(m,x)$$

and the Hamiltonian

$$H(x,p) = \sup_{\alpha \in \mathbb{R}^d} -\alpha \cdot p - L(x,\alpha)$$

are such that, for any $m \in C^0([0,T],\mathcal{P}_2(\mathbb{R}^d))$, the solution u to the Hamilton-Jacobi

$$\left\{\begin{array}{l} -\partial_t u + H(x,Du) = f(x,m(t)) \qquad \text{in } (0,T)\times\mathbb{R}^d,\\ u(T,x) = g(x,m(T)) \qquad \text{in } \mathbb{R}^d. \end{array}\right.$$

is uniformly Lipschitz continuous and semi-concave. Several structure condition are known for this to holds: the simplest one being that $f(\cdot,m)$ and $g(\cdot,m)$ are bounded in C^2 independently of m and $H(x,p)=|p|^2$. See also [12].

Theorem 4.3. *Assume that $(\bar m,\bar w)$ is a minimum of J defined by (12). Let $\bar u:[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be the viscosity solution to the (backward in time) Hamilton-Jacobi equation*

$$\left\{\begin{array}{l} -\partial_t \bar u + H(x,D\bar u) = f(x,\bar m(t)) \qquad \text{in } (0,T)\times\mathbb{R}^d,\\ \bar u(T,x) = g(x,\bar m(T)) \qquad \text{in } \mathbb{R}^d. \end{array}\right. \tag{14}$$

Then $dt\otimes d\bar m-$a.e.,

$$-D_\alpha L(x,\frac{d\bar w}{dt\otimes d\bar m}(t,x)) \in D^+\bar u(t,x)),$$

where $D^+\bar u(t,x)$ is the superdifferential of $\bar u$ at (t,x).

In fact, it is known that $\bar u$ is differentiable on the support of $\bar m$, so that the above inclusion can be rewritten as

$$\frac{d\bar w}{dt\otimes d\bar m}(t,x) = -D_pH(x,D\bar u(t,x))$$

$dt\otimes d\bar m-$a.e. This shows that the pair $(\bar u,\bar m)$ is a solution to the mean field game system

$$\left\{\begin{array}{l} -\partial_t \bar u + H(x,D\bar u) = f(x,\bar m(t)) \qquad \text{in } (0,T)\times\mathbb{R}^d,\\ \partial_t\bar m - \operatorname{div}(\bar m D_pH(x,D\bar u)) = 0 \qquad \text{in } (0,T)\times\mathbb{R}^d,\\ \bar m(0)=m_0,\ \bar u(T,x) = g(x,\bar m(T)) \qquad \text{in } \mathbb{R}^d. \end{array}\right.$$

The proof requires some intermediate steps. Let J^l be the linearized energy defined by

$$\begin{aligned} J^l(m,w) := &\int_0^T\int_{\mathbb{R}^d} L(x,\frac{dw}{dm}(t,x))m(t,dx)dt + \int_0^T\int_{\mathbb{R}^d} f(x,\bar m(t))m(t,dx)dt\\ &+\int_{\mathbb{R}^d} g(x,\bar m(T))m(T,dx) \end{aligned} \tag{15}$$

if (m,w) is admissible and $J(m,w)=+\infty$ otherwise. One easily checks that the following holds:

Lemma 4.4. *$(\bar m,\bar w)$ is a minimum of J^l.*

The main step of the proof is the lack of duality gap explained in Lemma 4.5 below. This statement relies on Von Neumann min-max Theorem that we recall now: Let A and B two convex subsets of some vector spaces, B being compact and let $\mathcal{L}:A\times B\to\mathbb{R}$ be such that

(i) $a\to\mathcal{L}(a,b)$ is concave for any $b\in B$,

(ii) $b\to\mathcal{L}(a,b)$ is convex and lsc for any $a\in A$.

Then

$$\min_{b\in B}\sup_{a\in A}\mathcal{L}(a,b) = \sup_{a\in A}\min_{b\in B}\mathcal{L}(a,b). \tag{16}$$

We use below a version (see Theorem A.1 [33]) in which the compactness of B is replaced by a coercivity of $\mathcal{L}$ with respect to the b variable. More precisely, let us assume (i) and, instead of B compact and (ii), that there exists $a^*\in A$ and $C^*>0$ such that (ii') $C^* > \sup_{a\in A}\inf_{b\in B}\mathcal{L}(a,b)$,

(ii") the set $B^* := \{b \in B,\ \mathcal{L}(a^*,b) \leqslant C^*\}$ is non-empty and $\mathcal{L}(a,\cdot)$ is lsc on B_* for any $a \in A$.

Then (16) holds.

Lemma 4.5 (No duality gap). *Under the assumption of Theorem 4.3 and if J^l is defined by* (15), *we have*

$$\inf_{(m,w)} J^l(m,w) = \sup_{\phi} \int_{\mathbb{R}^d} \phi(0,x)m_0(dx),$$

where the supremum in the right-hand side is taken over the maps $\phi \in C^1([0,T]\times\mathbb{R}^d,\mathbb{R})$ such that

$$\|\partial_t\phi\|_\infty + \|D\phi\|_\infty < +\infty \tag{17}$$

and ϕ is a subsolution of the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t\phi + H(x,D\phi) \leqslant f(x,\bar{m}(t)) & \text{in } (0,T)\times\mathbb{R}^d, \\ \phi(T,x) \leqslant g(x,\bar{m}(T)) & \text{in } \mathbb{R}^d. \end{cases} \tag{18}$$

Proof. We split the equality in two parts. First we check that given a finite Borell mesure m_T on $\mathbb{R}^d$,

$$\inf_{(m,w),\, m(T)=m_T} J(m,w) = \sup_{\phi} \int_{\mathbb{R}^d} (g(x,\bar{m}(T)) - \phi(T,x))m_T(dx) + \int_{\mathbb{R}^d} \phi(0,x)m_0(dx), \tag{19}$$

where the infimum in the LHD is taken over all admissible pairs (m,w) satisfying the continuity equation and the boundary constraints $m(0) = m_0$ and $m(T) = m_T$ while the supremum in the RHS is taken over all $\phi \in C^1$ such that (17) holds and

$$-\partial_t\phi + H(x,D\phi) \leqslant f(x,\bar{m}(t)) \qquad \text{in } (0,T)\times\mathbb{R}^d.$$

Then we relax the constraint on m_T. Note that the LHS of 19 is $+\infty$ if m_T is not of mass 1 because in this case there is no admissible pair (m,w) with $m(T) = m_T$.

To prove (19), let us introduce some notations. Let A be the set of C^1 maps $\phi : [0,T]\times\mathbb{R}^d \to \mathbb{R}$ for which (17) holds and let B be the set of pairs (m,w), $m = m(t,dx)dt$ being a Borel mesure on $[0,T]\times\mathbb{R}^d$ and w being a Borel vector measure on $[0,T]\times\mathbb{R}^d$ with values in $\mathbb{R}^d$, which is absolutely continuous with respect to m. Note that a pair (m,w) is admissible in the LHS of (19) if and only if, for any $\phi \in A$ we have

$$\int_0^T\int_{\mathbb{R}^d} \partial_t\phi m + \int_0^T\int_{\mathbb{R}^d} D\phi\cdot w - \int_{\mathbb{R}^d}\phi(T,x)m_T(dx) + \int_{\mathbb{R}^d}\phi(0,x)m_0(dx) = 0.$$

Therefore the LHS of (19) equals $\inf_{(m,w)\in B}\sup_{\phi\in A}\mathcal{L}(\phi,(m,w))$, where $\mathcal{L} : A\times B \to \mathbb{R}$ is defined by

$$\begin{aligned}\mathcal{L}(\phi,(m,w)) = \int_0^T\int_{\mathbb{R}^d} (L(x,\frac{dw}{dt\otimes dm}(t,x)) + f(x,\bar{m}(t)))m(t,dx)dt + \int_{\mathbb{R}^d} g(x,\bar{m}(T))m_T(dx) \\ + \int_0^T\int_{\mathbb{R}^d}\partial_t\phi m + \int_0^T\int_{\mathbb{R}^d} D\phi\cdot w - \int_{\mathbb{R}^d}\phi(T,x)m_T(dx) + \int_{\mathbb{R}^d}\phi(0,x)m_0(dx).\end{aligned}$$

Note that $\mathcal{L}$ is concave in ϕ and convex in (m, w). Let us set $\phi^*(t,x) = at(1+b|x|^2)^{1/2}$ (for $a, b > 0$) and note that the set $B^* := \{(m,w) \in B,\ \mathcal{L}(\phi^*, (m,w)) \leqslant C^*\}$ is nonempty and compact for any $a, C^* > 0$ large enough and b small enough because

$$\mathcal{L}(\phi^*, (m,w)) \geqslant C_0^{-1} \int_0^T \int_{\mathbb{R}^d} \left| \frac{dw}{dt \otimes dm} \right|^2 m(t,dx)dt - C_0 T - \|f\|_\infty m([0,T] \times \mathbb{R}^d)$$
$$+ a \int_0^T \int_{\mathbb{R}^d} (1+b|x|^2)^{1/2} m(t,dx)dt - ab^{1/2} T \int_0^T \int_{\mathbb{R}^d} |w|(dt,dx) - C(a,b).$$

Moreover, $\mathcal{L}(\phi, \cdot)$ is lsc on B_* for any $\phi \in B$. By the min-max Theorem stated above, the LHS of (19) equals $\sup\limits_{\phi \in A} \inf\limits_{(m,w) \in B} \mathcal{L}(\phi, (m,w))$. Note that, for any $\phi \in A$,

$$\begin{aligned}
&\inf_{(m,w)\in B} \mathcal{L}(\phi, (m,w)) \\
&= \inf_{(m,w)} \int_0^T \int_{\mathbb{R}^d} (L(x, \frac{dw}{dt \otimes dm}(t,x)) + f(x, \bar m(t))) m(t,dx)dt + \int_{\mathbb{R}^d} g(x, \bar m(T)) m_T(dx) \\
&\qquad + \int_0^T \int_{\mathbb{R}^d} \partial_t \phi m + \int_0^T \int_{\mathbb{R}^d} D\phi \cdot w - \int_{\mathbb{R}^d} \phi(T,x) m_T(dx) + \int_{\mathbb{R}^d} \phi(0,x) m_0(dx) \\
&= \inf_m \int_0^T \int_{\mathbb{R}^d} (-H(x, D\phi) + f(x, \bar m(t)) + \partial_t \phi) m(t,dx)dt + \int_{\mathbb{R}^d} g(x, \bar m(T)) m_T(dx) \\
&\qquad - \int_{\mathbb{R}^d} \phi(T,x) m_T(dx) + \int_{\mathbb{R}^d} \phi(0,x) m_0(dx) \\
&= \begin{cases} \displaystyle\int_{\mathbb{R}^d} (g(x, \bar m(T)) - \phi(T,x)) m_T(dx) + \int_{\mathbb{R}^d} \phi(0,x) m_0(dx) & \text{if } -H(x, D\phi) + f(x, \bar m(t)) + \partial_t \phi \geqslant 0 \\ -\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

(Here we are a little sloppy: see [Chap. IX, Prop. 2.1][21] for details). We conclude that (19) holds.

Let us now relax the constraint on m_T. By (19) we have

$$J(\bar m, \bar w) = \inf_{m_T} \sup_{\phi \in C^1,\ \phi \text{ subsolution}} \int_{\mathbb{R}^d} (g(x, \bar m(T)) - \phi(T,x)) m_T(dx) + \int_{\mathbb{R}^d} \phi(0,x) m_0(dx).$$

One can check again (using similar arguments) that the min-max Theorem applies and we obtain

$$\begin{aligned}
J(\bar m, \bar w) &= \sup_{\phi \in C^1,\ \phi \text{ subsolution}} \inf_{m_T} \int_{\mathbb{R}^d} (g(x, \bar m(T)) - \phi(T,x)) m_T(dx) + \int_{\mathbb{R}^d} \phi(0,x) m_0(dx) \\
&= \sup_{\phi \in C^1,\ \phi \text{ subsolution}} \begin{cases} \displaystyle\int_{\mathbb{R}^d} \phi(0,x) m_0(dx) & \text{if } g(x, \bar m(T)) - \phi(T,x) \geqslant 0 \\ -\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

This completes the proof. □

Note that, by the comparison principle for Hamilton-Jacobi equation, if $\bar u$ is the viscosity solution to (14) and if ϕ is as in Lemma (4.5), then $\phi \leqslant \bar u$. Hence

$$\sup_\phi \int_{\mathbb{R}^d} \phi(0,x) m_0(dx) \leqslant \int_{\mathbb{R}^d} \bar u(0,x) m_0(dx).$$

Conversely, by the convexity of $H = H(x, p)$ with respect to p, one can build by convoluting $\bar{u}$ with a smooth kernel a sequence of maps ϕ_n as in Lemma 4.5 such that ϕ_n converges uniformly to $\bar{u}$. This shows the following:

Lemma 4.6. *There exists a sequence (ϕ_n) as in Lemma 4.5 such that ϕ_n converges uniformly to $\bar{u}$ and*

$$\sup_{\phi} \int_{\mathbb{R}^d} \phi(0,x)m_0(dx) = \lim_n \int_{\mathbb{R}^d} \phi_n(0,x)m_0(dx) = \int_{\mathbb{R}^d} \bar{u}(0,x)m_0(dx).$$

Proof of Theorem 4.3. By the absence of duality gap (Lemma 4.5), we have

$$\begin{aligned} o_n(1) &= J^l(\bar{m}, \bar{w}) - \int_{\mathbb{R}^d} \phi_n m_0 \\ &= \int_0^T \int_{\mathbb{R}^d} L(x, \frac{d\bar{w}}{dt \otimes d\bar{m}}(t,x))\bar{m}(t,dx)dt + \int_0^T \int_{\mathbb{R}^d} f(x, \bar{m}(t))\bar{m}(t,dx)dt \\ &\quad + \int_{\mathbb{R}^d} g(x, \bar{m}(T))\bar{m}(T,dx) - \int_{\mathbb{R}^d} \phi_n m_0 \end{aligned}$$

Testing the continuity equation satisfied by $(\bar{m}, \bar{w})$ with ϕ_n, we have

$$\int_0^T \int_{\mathbb{R}^d} \partial_t \phi_n \bar{m} + \int_0^T \int_{\mathbb{R}^d} D\phi_n \cdot \bar{w} - \int_{\mathbb{R}^d} \phi_n(T,x)\bar{m}(T,dx) + \int_{\mathbb{R}^d} \phi_n(0,x)m_0(dx) = 0.$$

Using the (in)equality satisfied by ϕ_n, this implies that

$$\int_0^T \int_{\mathbb{R}^d} (H(x, D\phi_n) - f(x, \bar{m}) + \frac{d\bar{w}}{dt \otimes d\bar{m}} \cdot D\phi_n)\bar{m} - \int_{\mathbb{R}^d} \phi_n(T,x)\bar{m}(T,dx) + \int_{\mathbb{R}^d} \phi_n(0,x)m_0(dx) \leqslant 0.$$

Therefore,

$$\begin{aligned} o_n(1) \geqslant &\int_0^T \int_{\mathbb{R}^d} (L(x, \frac{d\bar{w}}{dt \otimes d\bar{m}}) + H(x, D\phi_n) + \frac{d\bar{w}}{dt \otimes d\bar{m}})\bar{m}(t,dx)dt \\ &+ \int_{\mathbb{R}^d} (g(x, \bar{m}(T)) - \phi_n(T,x))\bar{m}(T,dx). \end{aligned}$$

We use the uniform convexity of H to infer that

$$o_n(1) \geqslant \int_0^T \int_{\mathbb{R}^d} C^{-1} \left| D_\alpha L(x, \frac{d\bar{w}}{dt \otimes d\bar{m}}) + D\phi_n \right|^2 \bar{m}(t,dx)dt.$$

As u is semi-concave, for any $(t,x) \in [0,T] \times \mathbb{R}^d$,

$$\limsup_n D\phi_n(t,x) \subset D^+u(t,x),$$

where the $\limsup$ is understood in the Kuratowski sense. This shows that

$$-D_\alpha L(x, \frac{d\bar{w}}{dt \otimes d\bar{m}})(t,x) \in D^+u(t,x)$$

for $dt \otimes d\bar{m}$–a.e. (t,x). □

4.3 The mean field limit

Here we study the relation between the continuous model studied in the previous sections and some finite dimensional models with a finite (but large) number of controlled particles. Let N be the number of particles. Let $\bar{\mathbf{x}}_0^N = (\bar{x}_0^1, \dots, \bar{x}_0^N) \in (\mathbb{R}^d)^N$ be a given initial condition. The problem of calculus of variation consists in minimizing the cost

$$J^N(x_0^1, \dots, x_0^N, \alpha^1, \dots, \alpha^N) = \frac{1}{N}\sum_{i=1}^N \int_0^T L(x_t^i, \alpha_t^i)dt + \int_0^T F(m_{\mathbf{x}_t}^N)dt + G(m_{\mathbf{x}_T}^N)$$

where $\alpha^1, \dots, \alpha^N$ are the controls (in $L^2((0,T), \mathbb{R}^d)$) and $\mathbf{x}_t = (x_t^1, \dots, x_t^N)$ is the solution to

$$\dot{x}_t^i = \alpha_t^i, \; t \in [0,T], \qquad x_0^i = \bar{x}_0^i.$$

and where, as usual,

$$m_{\mathbf{x}}^N = \frac{1}{N}\sum_{i=1}^N \delta_{x^i}, \qquad \mathbf{x} = (x^1, \dots, x^N).$$

Theorem 4.7. *Assume that L, F and G are as in Theorem 4.1. Let $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$. There exists a sequence of initial conditions $(\bar{\mathbf{x}}_0^N)$ with*

$$m_{\bar{\mathbf{x}}_0}^N \to \bar{m}_0 \qquad \text{in } \mathcal{P}_2(\mathbb{R}^d)$$

and, if $(\bar{\alpha}^i)$ is a minimum for $J^N(\bar{\mathbf{x}}_0^N, \cdot)$ with associated trajectory $\bar{\mathbf{x}}_t = (\bar{x}_t^1, \cdots, \bar{x}_t^N)$, and if we denote by $\bar{m}^N(t) := m_{\bar{\mathbf{x}}_t}^N$ the associated empirical measure, then the sequence $(\bar{m}^N)$ is tight in $C^0([0,T], \mathcal{P}_{2-\epsilon}(\mathbb{R}^d))$ (for any $\epsilon \in (0,2)$) and, for any cluster point $m \in C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$, there exists w such that (m, w) is admissible and minimizes the cost J defined in (12).

For the proof we need two intermediate results. The first one states that the discrete problem is embedded into the continuous one. The second one provides the tightness property claimed in the Theorem.

Lemma 4.8. *Let $\alpha := (\alpha^i)$ a family of controls and $\mathbf{x} := (x^i)$ be the associated solutions. Let $w^N(t) := \frac{1}{N}\sum_{i=1}^N \alpha_t^i \delta_{x_t^i}$. Then the pair $(m_{\mathbf{x}}^N, w^N)$ is admissible for the initial condition $m_{\mathbf{x}_0}^N$ and*

$$J^N(x_0^1, \dots, x_0^N, \alpha^1, \dots, \alpha^N) = J(m^N, w^N).$$

Remark 4.9. The fact that the discrete problem is embedded into the continuous implies that

$$\inf_{\alpha} J^N(x_0^1, \dots, x_0^N, \alpha^1, \dots, \alpha^N) \geqslant \inf_{(m,w)} J(m^N, w^N),$$

where the infimum in the right-hand side is taken over the pairs (m, w) which are admissible for the initial condition $m_{\mathbf{x}_0}^N$. Let us warn the reader that there is no equality in general, since in the continuous problem the mass can split.

Here is an example: Assume that $d = 1$, $L(x, \alpha) = |\alpha|^2/2$, $F = 0$ and $G_0 : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is equal to 0 at the all the Dirac masses and is negative elsewhere. Then, for $N = 1$ and for the terminal cost $G = \lambda G_0$ (where $\lambda > 0$ is to be chosen below),

$$J^1(x_0^1, \alpha^1) = \int_0^T \frac{1}{2}|\alpha_t^1|^2 dt + \lambda G_0(\delta_{x_T^1}) \geqslant 0,$$

while choosing the admissible pair (m^1, w^1) given by $m^1(t) = (\delta_{-t} + \delta_t)/2$, $w^1(t) = (-\delta_{-t} + \delta_t)/2$, one has

$$J(m^1, w^1) = \int_0^T \frac{1}{2} \left| \frac{dw^1}{dm^1} \right|^2 dt + \lambda G_0((\delta_{-T} + \delta_T)/2) = T + \lambda G_0((\delta_{-T} + \delta_T)/2),$$

which is negative if $\lambda > T/|G_0((\delta_{-T}+\delta_T)/2)|$ (this choice being possible since $G_0((\delta_{-T}+\delta_T)/2) < 0$).

Proof. By definition, $w^N(t)$ is a.c. with respect to $m^N(t)$ and its density is given by

$$\frac{dw^N(t)}{dm^N(t)}(x) = \sum_{i=1}^N \alpha_t^i \mathbf{1}_{x=x_t^i}.$$

So $(m^N_{\mathbf{x}}, w^N)$ is admissible (for the initial condition $m^N_{\mathbf{x}_0}$) and

$$\begin{aligned} J(m^N, w^N) &= \int_0^T \int_{\mathbb{R}^d} L(x, \frac{dw^N}{dm^N}(t,x)) m^N(t,dx)dt + \int_0^T F(m^N(t))dt + G(m^N(T)) \\ &= \frac{1}{N} \sum_{i=1}^N L(x_t^i, \alpha_t^i)dt + \int_0^T F(m^N(t))dt + G(m^N(T)) = J^N(x_0^1, \dots, x_0^N, \alpha^1, \dots, \alpha^N). \end{aligned}$$

□

We now state the tightness property:

Lemma 4.10. *Assume that $(\bar{\mathbf{x}}_0^N)$ is a sequence of initial conditions with*

$$m^N_{\bar{\mathbf{x}}_0} \to \bar{m}_0 \qquad \text{in } \mathcal{P}_2(\mathbb{R}^d).$$

Let $(\bar{\alpha}^i)$ be a minimum for $J^N(\bar{\mathbf{x}}_0^N, \cdot)$ with associated trajectory $\bar{\mathbf{x}}_t = (\bar{x}_t^1, \cdots, \bar{x}_t^N)$ and let $\bar{m}^N(t) := m^N_{\bar{\mathbf{x}}_t}$ be the associated empirical measure. Then the sequence $(\bar{m}^N)$ is tight in $C^0([0,T], \mathcal{P}_{2-\epsilon}(\mathbb{R}^d))$ (for any $\epsilon \in (0,2)$) and, for any cluster point $m \in C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$, there exists w such that (m,w) is admissible and

$$J(m,w) \leqslant \limsup_N J^N(\mathbf{x}_0^{N,\epsilon_N}, \bar{\alpha}^N).$$

Proof. We already know by Lemma 4.8 that, if we set $\bar{w}^N(t) := \frac{1}{N}\sum_{i=1}^N \bar{\alpha}_t^i \delta_{x_t^i}$, then

$$J^N(x_0^1, \dots, x_0^N, \bar{\alpha}^1, \dots, \bar{\alpha}^N) = J(\bar{m}^N, \bar{w}^N).$$

It is easily to check that $J^N(\mathbf{x}_0^{N,\epsilon_N}, \bar{\alpha}^N)$ is bounded. Then arguing exactly as in the proof of the existence result (Theorem 4.1), the claim follows. □

Proof of Theorem 4.7. Let $(\bar{m}, \bar{w})$ be a minimum for J with initial condition (m_0, w_0). Arguing as in the proof of Lemma 4.2, we can mollify $(\bar{m}, \bar{w})$ in such a way that (m^ϵ, w^ϵ) is smooth, admissible (for a possibly different initial condition m_0^ϵ), with $m^\epsilon > 0$ and

$$m^\epsilon \to m \text{ in } C^0([0,T], \mathcal{P}_2(\mathbb{R}^d)), \qquad J(m^\epsilon, w^\epsilon) \leqslant J(\bar{m}, \bar{w}) + \epsilon.$$

Let us set $\alpha^\epsilon(t,x) := w^\epsilon(t,x)/m^\epsilon(t,x)$. Note that, by the smoothness and the positive property of m^ϵ and w^ϵ, $m^\epsilon(t) = X^{\epsilon,\cdot}(t)\sharp m_0^\epsilon$, where $X^{\epsilon,x}(t)$ is the solution of the differential equation

$$\begin{cases} \frac{d}{dt}X^{\epsilon,x}(t) = \alpha^\epsilon(t, X_t^{\epsilon,x}) \\ X_0^{\epsilon,x} = x \end{cases}$$

Let $\mathbf{x}_0^{N,\epsilon} = (x_0^{N,\epsilon,1}, \dots, x_0^{N,\epsilon,N}) \in (\mathbb{R}^d)^N$ be a sequence of initial conditions for the discrete system such that $m^N_{\mathbf{x}_0^{N,\epsilon}} \to m_0^\epsilon$ in $\mathcal{P}_2(\mathbb{R}^d)$. We define $\alpha_t^{\epsilon,N,i} = \alpha(t, X_t^{\epsilon,x_0^{\epsilon,N,i}})$ and note that the associated solution with initial condition $x_0^{N,\epsilon,i}$ is $X_t^{\epsilon,x_0^{\epsilon,N,i}}$. The associated empirical measure $m^N(t)$ associated with the particles $X_t^{\epsilon,x_0^{\epsilon,N,i}}$ is, by definition, given by

$$m^N(t) = X^{\epsilon,\cdot}(t)\sharp m^N_{\mathbf{x}_0^{N,\epsilon}}.$$

As $m^N_{\mathbf{x}_0^{N,\epsilon}}$ converges to m_0^ϵ in $\mathcal{P}_2(\mathbb{R}^d)$, $(m^N(t))$ converges to $(m^\epsilon(t))$ in $C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$. The same argument shows that

$$\frac{1}{N}\sum_{i=1}^N \int_0^T L(X_t^{\epsilon,x_0^{\epsilon,N,i}}, \alpha_t^{\epsilon,N,i})dt = \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} L(x,\alpha)(X^{\epsilon,\cdot}(t), \alpha(t, X^{\epsilon,\cdot}(t)))\sharp m^N_{\mathbf{x}_0^{N,\epsilon}} dt$$

converges to

$$\int_0^T \int_{\mathbb{R}^d} L(x,\alpha)(X^{\epsilon,\cdot}(t), \alpha(t, X^{\epsilon,\cdot}(t)))\sharp m_0(dx)dt = \int_0^T \int_{\mathbb{R}^d} L(x, \alpha^\epsilon(t,x)) m_0^\epsilon(dx)dt.$$

This proves that

$$\lim_N J^N(\mathbf{x}_0^{N,\epsilon}, (\alpha_t^{\epsilon,N,i})) = J(m^\epsilon, w^\epsilon).$$

In particular,

$$\limsup_{\epsilon\to 0} \limsup_{N\to+\infty} \inf_\alpha J^N(\mathbf{x}_0^{N,\epsilon}, \alpha) \leqslant J(\bar m, \bar w).$$

Let now $\epsilon_N \to 0$ be such that $m^N_{\mathbf{x}_0^{N,\epsilon_N}}$ converges to m_0 in $\mathcal{P}_2(\mathbb{R}^d)$ and

$$\limsup_{N\to+\infty} \inf_\alpha J^N(\mathbf{x}_0^{N,\epsilon_N}, \alpha) \leqslant J(\bar m, \bar w).$$

Let also $\bar\alpha^N := (\bar\alpha^i)$ be a minimum for $J^N(\mathbf{x}_0^{N,\epsilon_N}, \cdot)$. We denote by $\bar{\mathbf{x}}_t = (\bar x_t^1, \cdots, \bar x_t^N)$, the associated trajectory and by $\bar m^N(t) := m^N_{\bar{\mathbf{x}}_t}$ the associated empirical measure. We know for Lemma 4.10 that the sequence $(\bar m^N)$ is tight in $C^0([0,T], \mathcal{P}_{2-\epsilon}(\mathbb{R}^d))$ (for any $\epsilon \in (0,2)$) and that, if $m \in C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$ is a cluster point, then there exists w such that (m,w) is admissible with

$$J(m,w) \leqslant \limsup_N J^N(\mathbf{x}_0^{N,\epsilon_N}, \bar\alpha^N) \leqslant J(\bar m, \bar w).$$

As $m(0) = m_0$, we conclude that (m,w) is a minimum of J. □

4.4 The associated Hamilton-Jacobi equation

In this section we consider the value function

$$\mathcal{V}(t_0, m_0) = \inf_{(m,w)} \int_{t_0}^{T} \int_{\mathbb{R}^d} L(x, \frac{dw}{dm}(t,x))m(t,dx)dt + \int_{t_0}^{T} F(m(t))dt + G(m(T)),$$

where the infimum is taken over the admissible pairs (m, w) defined as previously on $[t_0, T] \times \mathbb{R}^d$ instead in $[0, T] \times \mathbb{R}^d$ and with initial condition $m(t_0) = m_0$.

Proposition 4.11. *Assume that F and G are Lipschitz continuous and that*

$$|D_x L(x, \alpha)| \leqslant C(1 + |\alpha|) \tag{20}$$

The map $\mathcal{V}$ is locally Lipschitz continuous on $[0, T] \times \mathcal{P}_2(\mathbb{R}^d)$.

Proof. In view of Theorem 4.7 we just need to prove the Lipschitz continuity of the value function associated with J^N and then pass to the limit. Let $\mathbf{x}_0, \mathbf{y}_0 \in (\mathbb{R}^d)^N$ and α be optimal for $J^N(\mathbf{x}_0, \cdot)$ (where J^N is define in the previous section). Then

$$J^N(\mathbf{y}_0, \alpha) = \frac{1}{N} \sum_{i=1}^{N} \int_0^T L(y_t^i, \alpha_t^i)dt + \int_0^T F(m_{\mathbf{y}_t}^N)dt + G(m_{\mathbf{y}_T}^N).$$

Note that $y_t^i = y_0^i - x_0^i + x_t^i$, so that by Proposition 1.3,

$$\mathbf{d}_2^2(m_{\mathbf{x}_t}^N, m_{\mathbf{y}_t}^N) = \inf_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^{N} |y_t^i - x_t^{\sigma(i)}|^2 = \inf_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^{N} |y_0^i - x_0^{\sigma(i)}|^2 = \mathbf{d}_2^2(m_{\mathbf{x}_0}^N, m_{\mathbf{y}_0}^N).$$

On the other hand, using assumption (20),

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^{N} \int_0^T L(y_t^i, \alpha_t^i)dt &\leqslant \frac{1}{N} \sum_{i=1}^{N} \int_0^T (L(x_t^i, \alpha_t^i) + C(1 + |\alpha_t^i|)|x_0^i - y_0^i|)dt \\
&\leqslant \frac{1}{N} \sum_{i=1}^{N} \int_0^T L(x_t^i, \alpha_t^i) + C \left(\frac{1}{N} \sum_{i=1}^{N} \int_0^T (1 + |\alpha_t^i|)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^{N} |x_0^i - y_0^i|^2 \right)^{1/2}.
\end{aligned}$$

By the growth assumption on L, the quantity $\frac{1}{N}\sum_{i=1}^{N} \int_0^T |\alpha_t^i|^2 dt$ remains bounded for bounded (in $\mathcal{P}_2(\mathbb{R}^d)$) measures $m_{\mathbf{x}_0}^N$. Therefore

$$J^N(\mathbf{y}_0, \alpha) = J^N(\mathbf{x}_0, \alpha) + C\mathbf{d}_2(m_{\mathbf{x}_0}^N, m_{\mathbf{y}_0}^N),$$

where C depends on the bound of $m_{\mathbf{x}_0}^N$ in $\mathcal{P}_2(\mathbb{R}^d)$. □

From standard arguments in optimal control, $\mathcal{V}$ satisfies the dynamic programming principle

$$\mathcal{V}(t_0, m_0) = \inf_{(m,w)} \int_{t_0}^{t_1} \int_{\mathbb{R}^d} L(x, \frac{dw}{dm}(t,x))m(t,dx)dt + \int_{t_0}^{t_1} F(m(t))dt + \mathcal{V}(t_1, m(t_1))$$

for any $0 \leqslant t_0 \leqslant t_1 \leqslant T$. Let us give heuristic arguments explaining that, at least at a formal level, one expects $\mathcal{V}$ to solve the following Hamilton-Jacobi equation:

$$\begin{cases} -\partial_t \mathcal{V}(t,m) + \displaystyle\int_{\mathbb{R}^d} H(x, D_m \mathcal{V}(t,m,x))m(dx) = F(m) & \text{in } (0,T) \times \mathcal{P}_2(\mathbb{R}^d) \\ \mathcal{V}(T,m) = G(m) \qquad \text{in } \mathcal{P}_2(\mathbb{R}^d), \end{cases} \tag{21}$$

where

$$H(x,p) = \sup_{\alpha\in\mathbb{R}^d} -p\cdot\alpha - L(x,\alpha).$$

For this we suppose that $\mathcal{V}$ is of class C^1. Then, by dynamic programming principle with $t_1 = t_0 + h$ (where $h > 0$ is small),

$$\begin{aligned}
\mathcal{V}(t_0,m_0) &= \inf_{(m,w)} \int_{t_0}^{t_0+h}\int_{\mathbb{R}^d} L(x,\frac{dw}{dm}(t,x))m(t,dx)dt + hF(m_0) + o(h) \\
&\quad + \mathcal{V}(t_0,m_0) + \int_{t_0}^{t_0+h}\left(\partial_t\mathcal{V}(t,m(t)) + \int_{\mathbb{R}^d}\frac{\delta\mathcal{V}}{\delta m}(t,m(t),x)\partial_t m(t,dx)\right)dt \\
&= \inf_{(m,w)} \int_{t_0}^{t_0+h}\int_{\mathbb{R}^d} L(x,\frac{dw}{dm}(t,x))m(t,dx)dt + hF(m_0) + o(h) \\
&\quad + \mathcal{V}(t_0,m_0) + h\partial_t\mathcal{V}(t_0,m_0) + \int_{t_0}^{t_0+h}\partial_t\int_{\mathbb{R}^d} D_m\mathcal{V}(t,m(t),x)\cdot\frac{dw}{dm}m(t,dx)dt \\
&= -h\int_{\mathbb{R}^d} H(x,D_m\mathcal{V}(t_0,m_0,x)m_0(dx) + hF(m_0) + o(h) \\
&\quad + \mathcal{V}(t_0,m_0) + h\partial_t\mathcal{V}(t_0,m_0).
\end{aligned}$$

Simplifying by $\mathcal{V}(t_0,m_0)$, dividing by h and letting $h\to 0$ gives (21).

The rigorous justification of the above computation requires the notion of viscosity solution in $\mathcal{P}_2(\mathbb{R}^d)$. Although it is not too difficult to show that $\mathcal{V}$ solves, in the sense of viscosity solutions against smooth test functions, equation (21), I do not know if this is enough to characterize the value function $\mathcal{V}$. See however [26] where this characterization is achieved in terms of sub- and super-differential.

A way to overcome this issue is to write the equation in the space of random variables $L^2(\Omega,\mathbb{R}^d)$. Let us recall that one lifts the solution $\mathcal{V}$ to $\tilde{\mathcal{V}}$ defined by $\tilde{\mathcal{V}}(t,X) = \mathcal{V}(t,\mathcal{L}(X))$. Then equation (21) becomes

$$\begin{cases} -\partial_t\tilde{\mathcal{V}}(t,X) + \mathbb{E}\left[H(X,\nabla\tilde{\mathcal{V}}(t,X)\right] = F(\mathcal{L}(X)) \qquad \text{in } (0,T)\times L^2(\Omega,\mathbb{R}^d) \\ \tilde{\mathcal{V}}(T,X) = G(\mathcal{L}(X)) \qquad \text{in } L^2(\Omega,\mathbb{R}^d). \end{cases}$$

Following Crandall-Lions [19] it is not specially difficult to check, under suitable assumption on H, that the above equation has a unique viscosity solution (see also [22] for a monograph on HJ in infinite dimension and [5] for a discussion of this approach in the framework of mean field control).

4.5 Further reading

The problem described in this section is often called mean field-type control or mean-field optimal control in the literature. This subject has known an impressive development in the recent years and a general list of references is out of the scope of these short notes. We present below a (somewhat arbitrary) selection of results, in the hope that they are a little representative of the topic.

The existence of a solution described here is classical and it strongly related with the Benamou-Brenier's approach of optimal transport [7]. Note that in some works, a restriction is made on the regularity in space of the distributed control α [24]. As explained above, this technical restriction does not seem necessary.

The optimality conditions as stated here are, again, familiar in the context of optimal transport, where the solution u of the Hamilton-Jacobi equation is the so-called Kantorovich potential. The generalization to the framework of this section is formally described in [31] in the context of mean field games and then used to build solution for problems with local interactions (see for instance, among many other references, [8, 15, 33]).

The limit of the $N-$particle system has been recently studied by several authors and we refer to the nice paper by Lacker [28] for an extension to the stochastic control setting and further references. The mean field limit is also studied in the context of first order mean fields games in [23].

Concerning the analysis of the value function, if there are by now many references on the subject, the picture is not completely clear yet. Several works study the HJ in metric spaces (see, e.g. among many others, [4, 25, 27] and the references therein). There the Hamiltonian depends on the metric slope of the unknown, which, in the case of $\mathcal{P}_2(\mathbb{R}^d)$, means that it depends on $\|D_m \mathcal{V}\|_{L^2_m}$ and so that $L(x,v) = l(|p|)$ for some $l : \mathbb{R} \to \mathbb{R}$. In that case the characterization of the value function quite well understood and intrinsic, in the sense that it only depends on the metric space. Another possibility is to embed the problem into the space $L^2(\Omega, \mathbb{R}^d)$ of random variables, as we explained in Section 2.3. There is a complete characterization of the value function in terms of viscosity solutions in that setting: this approach has been first presented by Lions in [32] and the reader can find a detailed analysis in [5]. Note however that the test functions are C^1 (or C^2) functions on $L^2(\Omega, \mathbb{R}^d)$ and therefore cannot be expressed as test functions on $\mathcal{P}_2(\mathbb{R}^d)$. In [25], the authors discuss the equivalence between the viscosity solutions, in terms of sub- and super-differential, in $\mathcal{P}_2(\mathbb{R}^d)$ and in $L^2(\Omega, \mathbb{R}^d)$. Although this result is a major step towards a better understand, it is not completely clear how it applies to equation (21). In addition, it certainly does not apply to the more difficult problem with a diffusion. Let us finally note a different and direct approach, for first order problems and bounded controls, developped [34].

We have chosen in this chapter to describe problems in which the control $\alpha = \alpha(t,x)$ is distributed, which corresponds to the situation in which one controls the behavior of all the particles. If this approach is well adapted for some models, it is not always realistic: indeed it is not always possible to control in an optimal way *all* the particles. An alternative is to use sparse controls, as discussed for instance in [13]. On the other hand, if one considers the evolution of the density m as an evolution of (small) controllers, the approach presented does not take into account the possible rationality behavior of these controllers. The correct concept in this setting is the notion of mean field games: see [29, 30, 31, 32] and the monographs [9, 18]. As already pointed out, the optimization problem described in Section 4 can be used to build solution of the so-called MFG system by variational methods: see for instance [8, 15, 33].

References

[1] Alfonsi, A., & Jourdain, B. (2018). Lifted and geometric differentiability of the squared quadratic Wasserstein distance. arXiv preprint arXiv:1811.07787.

[2] Ambrosio, L., Gigli, N., Savaré, G. Gradient flows in metric spaces and in the space of probability measures. Second edition. Lectures in Mathematics

ETH Zürich. Birkhäuser Verlag, Basel, 2008.

[3] Ambrosio, L., & Gangbo, W. (2008). Hamiltonian ODEs in the Wasserstein space of probability measures. Communications on Pure and Applied Mathematics, 61(1), 18-53.

[4] Ambrosio, L., & Feng, J. (2014). On a class of first order Hamilton-Jacobi equations in metric spaces. Journal of Differential Equations, 256(7), 2194-2245.

[5] Bandini, E., Cosso, A., Fuhrman, M., & Pham, H. (2019). Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem. Stochastic Processes and their Applications, 129(2), 674-711.

[6] Bardi, M., & Capuzzo-Dolcetta, I. (2008). Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations. Springer Science & Business Media.

[7] Benamou, J. D., & Brenier, Y. (2000). A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numerische Mathematik, 84(3), 375-393.

[8] Benamou, J. D., Carlier, G., & Santambrogio, F. (2017). Variational mean field games. In Active Particles, Volume 1 (pp. 141-171). Birkhäuser, Cham.

[9] A. Bensoussan, J. Frehse, and P. Yam, Mean field games and mean field type control theory, vol. 101, Springer, 2013.

[10] Billingsley, P. (2008). Probability and measure. John Wiley & Sons.

[11] V.I. Bogachev. Measure Theory, Volume 2. Springer-Verlag Berlin Heidelberg, 2007.

[12] Cannarsa, P., & Sinestrari, C. (2004). Semiconcave functions, Hamilton-Jacobi equations, and optimal control (Vol. 58). Springer Science & Business Media.

[13] Caponigro, M., Fornasier, M., Piccoli, B., & Trélat, E. (2015). Sparse stabilization and control of alignment models. Mathematical Models and Methods in Applied Sciences, 25(03), 521-564.

[14] Cardaliaguet, P. (2010). Notes on mean field games. Technical report.

[15] Cardaliaguet, P. (2015). Weak solutions for first order mean field games with local coupling. In Analysis and geometry in control theory and its applications (pp. 111-158). Springer, Cham.

[16] Cardaliaguet P., Delarue F., Lasry J.-M., Lions P.-L. *The master equation and the convergence problem in mean field games.* To appear in Annals of Mathematics Studies.

[17] Cardaliaguet P., Cirant M., Porretta A. In preparation.

[18] Carmona, R., Delarue, F. (2018). Probabilistic Theory of Mean Field Games with Applications I-II. Springer Nature.

[19] M. G. Crandall, P.-L. Lions, Hamilton-Jacobi equations in infinite dimensions, I, J. Funct. Anal. 62 (1985) 379-396.

[20] R.M. Dudley. Real Analysis and Probability. Wadsworth & Brooks/Cole, 1989.

[21] Ekeland, I., & Temam, R. (1999). Convex analysis and variational problems (Vol. 28). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[22] Fabbri, G., Gozzi, F., & Swiech, A. (2017). Stochastic optimal control in infinite dimension. Probability and Stochastic Modelling. Springer.

[23] Fischer, M., & Silva, F. J. (2019). On the asymptotic nature of first order mean field games. arXiv preprint arXiv:1903.03602.

[24] Fornasier, M., & Solombrino, F. (2014). Mean-field optimal control. ESAIM: Control, Optimisation and Calculus of Variations, 20(4), 1123-1152.

[25] Gangbo, W., & Swiech, A. (2015). Metric viscosity solutions of Hamilton-Jacobi equations depending on local slopes. Calculus of Variations and Partial Differential Equations, 54(1), 1183-1218.

[26] Gangbo, W., & Tudorascu, A. (2019). On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. Journal de Mathématiques Pures et Appliquées, 125, 119-174.

[27] Hynd, R., & Kim, H. K. (2015). Value functions in the Wasserstein spaces: finite time horizons. Journal of Functional Analysis, 269(4), 968-997.

[28] Lacker, D. (2017). Limit theory for controlled McKean–Vlasov dynamics. SIAM Journal on Control and Optimization, 55(3), 1641-1672.

[29] J.-M. Lasry and P.-L. Lions, Jeux à champ moyen. i–le cas stationnaire, Comptes Rendus Mathématique, 343 (2006), pp. 619–625.

[30] J.-M. Lasry and P.-L. Lions Jeux à champ moyen. ii–horizon fini et contrôle optimal, Comptes Rendus Mathématique, 343 (2006), pp. 679–684.

[31] J.-M. Lasry and P.-L. Lions, Mean field games, Japanese journal of mathematics, 2 (2007), pp. 229–260.

[32] P.-L. Lions, Cours au college de france, Available at www. college-de-france. fr, (2007).

[33] Orrieri, C., Porretta, A., & Savaré, G. (2018). A variational approach to the mean field planning problem. arXiv preprint arXiv:1807.09874.

[34] Marigonda, A., & Quincampoix, M. (2018). Mayer control problem with probabilistic uncertainty on initial positions. Journal of Differential Equations, 264(5), 3212-3252.

[35] Santambrogio, F. (2015). Optimal transport for applied mathematicians. Birkäuser, NY, 55, 58-63.

[36] Villani, C. (2003). Topics in optimal transportation (No. 58). American Mathematical Soc..

[37] Villani, C. (2008). Optimal transport: old and new (Vol. 338). Springer Science & Business Media.