

## A PONTRYAGIN MAXIMUM PRINCIPLE IN WASSERSTEIN SPACES FOR CONSTRAINED OPTIMAL CONTROL PROBLEMS\*

BENOÎT BONNET\*\*

**Abstract.** In this paper, we prove a Pontryagin Maximum Principle for constrained optimal control problems in the Wasserstein space of probability measures. The dynamics is described by a transport equation with non-local velocities which are affine in the control, and is subject to end-point and running state constraints. Building on our previous work, we combine the classical method of needle-variations from geometric control theory and the metric differential structure of the Wasserstein spaces to obtain a maximum principle formulated in the so-called Gamkrelidze form.

**Mathematics Subject Classification.** 49K20, 49K27, 58E25.

Received October 30, 2018. Accepted July 13, 2019.

### 1. INTRODUCTION

Transport equations with non-local velocities have drawn a great amount of attention from several scientific communities for almost a century. They were first introduced in statistical physics to describe averaged Coulomb interactions within large assemblies of particles (see *e.g.* [55]), and are still to this day a widely studied topic in mathematical physics. More recently, a growing interest in the mathematical modelling of multi-agent systems has brought to light a whole new panel of problems in which these equations play a central role. Starting from the seminal paper of Cucker and Smale [26] dealing with emergent behaviour in animal flocks, a large literature has been devoted to the fine mathematical analysis of kinetic cooperative systems, *i.e.* systems described by non-local dynamics with attractive velocities, see *e.g.* [4, 12, 32, 39]. Besides, several prominent papers aimed at describing the emergence of patterns which were initially discovered for systems of ODEs in the context of kinetic models described by continuity equations [22, 38]. Simultaneously, Lasry and Lions laid in [43] the foundations of the theory of mean-field games, which is today one of the most active communities working on variational problems involving continuity equations, see *e.g.* [18, 19] and references therein.

---

\*This research is partially supported by the Padua University grant SID 2018 “Controllability, stabilizability and infimum gaps for control systems”, prot. BIRD 187147. The author was supported by the Archimède Labex (ANR-11-LABX-0033), by the A\*MIDEX project (ANR- 11-IDEX-0001-02), funded by the “Investissements d’Avenir” French Government program managed by the French National Research Agency (ANR), and by the SRGI ANR Grant ANR-15-CE40-0018.

The author is indebted to the anonymous reviewers for their careful reading and the many comments they provided that greatly improved the manuscript.

*Keywords and phrases:* Pontryagin Maximum Principle, Wasserstein spaces, metric differential calculus, needle-like variations, state constraints.

Aix Marseille Université, CNRS, ENSAM, Université de Toulon, LIS, Marseille, France.

\*\* Corresponding author: [benoit.bonnet@lis-lab.fr](mailto:benoit.bonnet@lis-lab.fr)

Later on, the focus shifted partly to include control-theoretic problems such as reachability analysis, optimal control, or explicit design of sparse control strategies. For these purposes, the vast majority of the existing contributions have taken advantage of the recent developments in the theory of optimal transport. We refer the reader to [50, 53] for a comprehensive introduction to this ever-expanding topic. In particular, the emergence of powerful tools of analysis in the so-called Wasserstein spaces has allowed for the establishment of a general existence theory for non-local transport equations (see *e.g.* [8, 9]), which incorporates natural Lipschitz and metric estimates in the smooth cases (see [45]).

Apart from a few controllability results as in [30], most of the attention of the community has been devoted to optimal control problems in Wasserstein spaces. The existence of optimal solutions has been investigated with various degrees of generality in [1, 2, 33–35, 48], mostly by means of  $\Gamma$ -convergence arguments. Besides, a few papers have been dealing with numerical methods either in the presence of diffusion terms, which considerably simplify the convergence analysis of the corresponding schemes (see *e.g.* [31]), or in the purely metric setting [5, 17, 47].

The derivation of Hamilton–Jacobi and Pontryagin optimality conditions has been an active topic in the community of Wasserstein optimal control in the recent years. Starting from the seminal paper [36] on Hamilton–Jacobi equations in the Wasserstein space, several contributions such as [23, 24] have been aiming at refining a dynamic-programming principle for mean-field optimal control problems. Pontryagin-type optimality conditions, on the other hand, have received less interest. The first result derived in [13] focuses on a multi-scale ODE-PDE system in which the control only acts on the ODE part. In this setting, the Pontryagin Maximum Principle (“PMP” for short) is derived by combining  $\Gamma$ -convergence and mean-field limit arguments. Another approach, introduced in our previous work [14], studies the infinite-dimensional problem by means of the classical technique of needle-variations (see *e.g.* [3, 15]) and makes an extensive use of the theory of Wasserstein subdifferential calculus formalized in [9]. The corresponding maximum principle is formulated as a Hamiltonian flow in the space of measures in the spirit of [8], and is in a sense the most natural generalization to be expected of the usual finite-dimensional Pontryagin-type optimality conditions. We would also like to mention that in [17, 48], the authors derived first-order necessary optimality conditions for special classes of optimal control problems on continuity equations, *via* methods which are quite distinct from that which we already sketched.

It is worth noticing that optimal control problems in Wasserstein spaces bear a lot of similarities with mean-field games. It was highlighted as early as [43] and further detailed *e.g.* in [21] that in the class of so-called potential mean-field games, the self-organization of an ensemble of agents could be equivalently reformulated as an optimal control problem in Wasserstein spaces involving adequately modified functionals. In particular in [20], a PMP was derived for controlled McKean–Vlasov dynamics describing such an optimal control problem from a probabilistic point of view. The analysis therein is carried out by leveraging the formalism of Lions derivatives in Wasserstein spaces (see *e.g.* [19]), which are one of the possible equivalent ways of looking at derivatives in the metric space of probability measures.

In this paper, we further the line of research initiated in [14] by extending our previous result, consisting in a Pontryagin Maximum Principle in Wasserstein spaces, to the setting of constrained optimal control problems. Such problems can be written in the following general form

$$(\mathcal{P}) \quad \begin{cases} \min_{u \in \mathcal{U}} \left[ \int_0^T L(t, \mu(t), u(t)) dt + \varphi(\mu(T)) \right], \\ \text{s.t.} \begin{cases} \partial_t \mu(t) + \nabla \cdot ((v[\mu(t)](t, \cdot) + u(t, \cdot))\mu(t)) = 0, \\ \mu(0) = \mu^0 \in \mathcal{P}_c(\mathbb{R}^d), \end{cases} \\ \text{and} \begin{cases} \Psi^I(\mu(T)) \leq 0, \quad \Psi^E(\mu(T)) = 0, \\ \Lambda(t, \mu(t)) \leq 0 \quad \text{for all } t \in [0, T]. \end{cases} \end{cases}$$

Here, the functions  $(t, \mu, \omega) \mapsto L(t, \mu, \omega)$  and  $\mu \mapsto \varphi(\mu)$  describe running and final costs, while the maps  $(t, \mu) \mapsto \Lambda(t, \mu)$  and  $\mu \mapsto \Psi^I(\mu), \Psi^E(\mu)$  are running and end-point constraints respectively. The velocity field  $(t, x, \mu) \mapsto$

$v[\mu](t, x)$  is a general non-local drift, which can be given *e.g.* in the form of a convolution (see [26, 34, 46]). The control  $(t, x) \mapsto u(t, x)$  is a vector-field which depends on both time and space, as customary in distributed control of partial differential equations (see *e.g.* [52]).

The methodology that we follow relies on the technique of *packages of needle-variations*, combined with a *Lagrange multiplier rule*. In essence, this method allows to recover the maximum principle from a family of finite-dimensional first-order optimality conditions by means of the introduction of a suitable costate. Even though classical in the unconstrained case, this direct approach does require some care to be translated to constrained problem. Indeed, the presence of constraints induces an unwanted dependency between the Lagrange multipliers and the needle-parameters. This extra difficulty can be circumvented by considering  $N$ -dimensional perturbations of the optimal trajectory instead of a single one, and by performing a limiting procedure as  $N$  goes to infinity. Originally introduced in [11] for smooth optimal control problems with end-point constraints, this approach was extended in [51] to the case of non-smooth and state-constrained problems. When trying to further adapt this method to the setting of Wasserstein spaces, one is faced with an extra structural difficulty. In the classical statement of the maximum principle, the presence of state constraints implies a mere  $BV$  regularity in time for the covectors. However, a deep result of optimal transport theory states that solutions of continuity equations in Wasserstein spaces coincide exactly with absolutely continuous curves (see *e.g.* [9], Thm. 8.3.1). Whence, in order to write a well-defined Wasserstein Hamiltonian flow in the spirit of [8, 14], we choose to formulate a maximum principle in the so-called Gamkrelidze form (see *e.g.* [10]), which allows to recover a stronger absolute continuity in time of the costates at the price of an extra regularity assumption on the state constraints.

This article is structured as follows. In Section 2, we recall general results of analysis in measure spaces along with elements of subdifferential calculus in Wasserstein spaces and existence results for continuity equations. We also introduce several notions of non-smooth analysis, including a general Lagrange multiplier rule formulated in terms of *Michel-Penot* subdifferentials. In Section 3, we state and prove our main result, that is Theorem 3.1. The argument is split into four steps which loosely follow the methodology already introduced in [14]. We exhibit in Appendix A a series of examples of functionals satisfying the structural assumptions **(H)** of Theorem 3.1, and we provide in Appendix B the analytical expression of the Wasserstein gradient of a functional involved in the statement of Theorem 3.1.

## 2. PRELIMINARY RESULTS

In this section, we recall several notions about analysis in the space of measures, optimal transport theory, Wasserstein spaces, subdifferential calculus in the space  $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ , and continuity equations with non-local velocities. We also introduce some elementary notions of non-smooth calculus in Banach spaces. For a complete introduction to these topics, see [9, 50, 53] and [25, 44] respectively.

### 2.1. Analysis in measure spaces and the optimal transport problem

In this section, we introduce some classical notations and results of measure theory, optimal transport and analysis in Wasserstein spaces. We denote by  $(\mathcal{M}_+(\mathbb{R}^d), \|\cdot\|_{TV})$  the set of real-valued non-negative Borel measures defined over  $\mathbb{R}^d$  endowed with the total variation norm, and by  $\mathcal{L}^d$  the standard Lebesgue measure on  $\mathbb{R}^d$ . It is known by Riesz Theorem (see *e.g.* [7], Thm. 1.54) that this space can be identified with the topological dual of the Banach space  $(C_0^0(\mathbb{R}^d), \|\cdot\|_{C^0})$ , which is the completion of the space of continuous and compactly supported functions  $C_c^0(\mathbb{R}^d)$  endowed with the  $C^0$ -norm.

We denote by  $\mathcal{P}(\mathbb{R}^d) \subset \mathcal{M}_+(\mathbb{R}^d)$  the set of Borel probability measures, and for  $p \geq 1$ , we denote by  $\mathcal{P}_p(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$  the set of measures with finite  $p$ -th moment, *i.e.*

$$\mathcal{P}_p(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) \text{ s.t. } \int_{\mathbb{R}^d} |x|^p d\mu(x) < +\infty \right\}.$$

The *support* of a Borel measure  $\mu \in \mathcal{M}_+(\mathbb{R}^d)$  is defined as the closed set  $\text{supp}(\mu) = \{x \in \mathbb{R}^d \text{ s.t. } \mu(\mathcal{N}) > 0 \text{ for any neighbourhood } \mathcal{N} \text{ of } x\}$ . We denote by  $\mathcal{P}_c(\mathbb{R}^d) \subset \mathcal{P}(\mathbb{R}^d)$  the set of probability measures with compact support.

We say that a sequence  $(\mu_n) \subset \mathcal{P}(\mathbb{R}^d)$  of Borel probability measures *converges narrowly* towards  $\mu \in \mathcal{P}(\mathbb{R}^d)$  – denoted by  $\mu_n \xrightarrow[n \rightarrow +\infty]{*} \mu$  – provided that

$$\int_{\mathbb{R}^d} \phi(x) d\mu_n(x) \xrightarrow[n \rightarrow +\infty]{} \int_{\mathbb{R}^d} \phi(x) d\mu(x), \quad (2.1)$$

for all  $\phi \in C_b^0(\mathbb{R}^d)$ . Here  $C_b^0(\mathbb{R}^d)$  denotes the set of continuous and bounded functions over  $\mathbb{R}^d$ . Remark that the corresponding *narrow topology* coincides with the restriction to  $\mathcal{P}(\mathbb{R}^d)$  of the weak-\* topology in  $\mathcal{M}_+(\mathbb{R}^d)$ . In the following proposition, we recall a useful convergence property on the measures of Borel sets, usually referred to as the *Portmanteau Theorem* (see *e.g.* [7], Prop. 1.62).

**Proposition 2.1** (Portmanteau Theorem). *Let  $(\mu_n) \subset \mathcal{M}_+(\mathbb{R}^d)$  be a sequence of measures converging in the weak-\* topology towards  $\mu \in \mathcal{M}_+(\mathbb{R}^d)$ . Then for any Borel set  $A \subset \mathbb{R}^d$  such that  $\mu(\partial A) = 0$ , it holds that  $\mu_n(A) \xrightarrow[n \rightarrow +\infty]{} \mu(A)$ .*

We recall in the following definitions the notion of *pushforward* of a Borel probability measure through a Borel map, along with that of *transport plan*.

**Definition 2.2** (Pushforward of a measure through a Borel map). Given  $\mu \in \mathcal{P}(\mathbb{R}^d)$  and a Borel map  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , the *pushforward*  $f_{\#}\mu$  of  $\mu$  through  $f(\cdot)$  is the unique Borel probability measure which satisfies  $f_{\#}\mu(B) = \mu(f^{-1}(B))$  for any Borel set  $B \subset \mathbb{R}^d$ .

**Definition 2.3** (Transport plan). Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ . We say that  $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$  is a *transport plan* between  $\mu$  and  $\nu$  – denoted by  $\gamma \in \Gamma(\mu, \nu)$  – provided that

$$\gamma(A \times \mathbb{R}^d) = \mu(A) \quad \text{and} \quad \gamma(\mathbb{R}^d \times B) = \nu(B),$$

for any pair of Borel sets  $A, B \subset \mathbb{R}^d$ . This property can be equivalently formulated in terms of pushforwards by  $\pi_{\#}^1 \gamma = \mu$  and  $\pi_{\#}^2 \gamma = \nu$ , where the maps  $\pi^1, \pi^2 : \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$  denote the projection operators on the first and second factor respectively.

In 1942, the Russian mathematician Leonid Kantorovich introduced in [42] the *optimal mass transportation problem* in its modern mathematical formulation. Given two probability measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  and a cost function  $c : \mathbb{R}^{2d} \rightarrow (-\infty, +\infty]$ , one aims at finding a *transport plan*  $\gamma \in \Gamma(\mu, \nu)$  such that

$$\int_{\mathbb{R}^{2d}} c(x, y) d\gamma(x, y) = \min_{\gamma'} \left\{ \int_{\mathbb{R}^{2d}} c(x, y) d\gamma'(x, y) \text{ s.t. } \gamma' \in \Gamma(\mu, \nu) \right\}.$$

This problem has been extensively studied in very broad contexts (see *e.g.* [9, 53]) with high levels of generality on the underlying spaces and cost functions. In the particular case where  $c(x, y) = |x - y|^p$  for some real number  $p \geq 1$ , the optimal transport problem can be used to define a distance over the subset  $\mathcal{P}_p(\mathbb{R}^d)$  of  $\mathcal{P}(\mathbb{R}^d)$ .

**Definition 2.4** (Wasserstein distance and Wasserstein spaces). Given two probability measures  $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ , the *p-Wasserstein distance* between  $\mu$  and  $\nu$  is defined by

$$W_p(\mu, \nu) = \min_{\gamma} \left\{ \left( \int_{\mathbb{R}^{2d}} |x - y|^p d\gamma(x, y) \right)^{1/p} \text{ s.t. } \gamma \in \Gamma(\mu, \nu) \right\}.$$

The set of plans  $\gamma \in \Gamma(\mu, \nu)$  achieving this optimal value is denoted by  $\Gamma_o(\mu, \nu)$  and is referred to as the set of *optimal transport plans* between  $\mu$  and  $\nu$ . The space  $(\mathcal{P}_p(\mathbb{R}^d), W_p)$  of probability measures with finite  $p$ -th moment endowed with the  $p$ -Wasserstein metric is called the Wasserstein space of order  $p$ .

We recall some of the interesting properties of these spaces in the following proposition (see *e.g.* [9], Chap. 7 or [53], Chap. 6).

**Proposition 2.5** (Properties of the Wasserstein spaces). *The Wasserstein spaces  $(\mathcal{P}_p(\mathbb{R}^d), W_p)$  are separable geodesic spaces. The topology generated by the  $p$ -Wasserstein metric metrizes the weak-\* topology of probability measures induced by the narrow convergence (2.1). More precisely, it holds that*

$$W_p(\mu_n, \mu) \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{if and only if} \quad \begin{cases} \mu_n \xrightarrow[n \rightarrow +\infty]{}^* \mu, \\ \int_{\mathbb{R}^d} |x|^p d\mu_n(x) \xrightarrow[n \rightarrow +\infty]{} \int_{\mathbb{R}^d} |x|^p d\mu(x). \end{cases}$$

Given two measures  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ , the Wasserstein distances are ordered, *i.e.*  $W_{p_1}(\mu, \nu) \leq W_{p_2}(\mu, \nu)$  whenever  $p_1 \leq p_2$ . Moreover, when  $p = 1$ , the following Kantorovich–Rubinstein duality formula holds

$$W_1(\mu, \nu) = \sup_{\phi} \left\{ \int_{\mathbb{R}^d} \phi(x) d(\mu - \nu)(x) \text{ s.t. } \text{Lip}(\phi; \mathbb{R}^d) \leq 1 \right\}, \quad (2.2)$$

where  $\text{Lip}(\phi; \mathbb{R}^d)$  denotes the Lipschitz constant of the map  $\phi(\cdot)$  on  $\mathbb{R}^d$ .

In what follows, we shall mainly restrict our considerations to the Wasserstein spaces of order 1 and 2 built over  $\mathcal{P}_c(\mathbb{R}^d)$ . We end these introductory paragraphs by recalling the concepts of *disintegration* and *barycenter* in the context of optimal transport.

**Definition 2.6** (Disintegration and barycenter). Let  $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$  and  $\gamma \in \Gamma(\mu, \nu)$  be a transport plan between  $\mu$  and  $\nu$ . We define the *disintegration*  $\{\gamma_x\}_{x \in \mathbb{R}^d} \subset \mathcal{P}_p(\mathbb{R}^d)$  of  $\gamma$  on its first marginal  $\mu$ , usually denoted by  $\gamma = \int \gamma_x d\mu(x)$ , as the  $\mu$ -almost uniquely determined Borel family of probability measures such that

$$\int_{\mathbb{R}^{2d}} \phi(x, y) d\gamma(x, y) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x, y) d\gamma_x(y) d\mu(x),$$

for any Borel map  $\phi \in L^1(\mathbb{R}^{2d}, \mathbb{R}; \gamma)$ . The barycenter  $\bar{\gamma} \in L^p(\mathbb{R}^d, \mathbb{R}^d; \mu)$  of the plan  $\gamma$  is then defined by

$$\bar{\gamma}(x) = \int_{\mathbb{R}^d} y d\gamma_x(y).$$

for  $\mu$ -almost every  $x \in \mathbb{R}^d$ .

**Proposition 2.7** (Wasserstein estimate between disintegrations). *Let  $\mu \in \mathcal{P}_c(\mathbb{R}^d)$ ,  $\gamma^1 = \int \gamma_x^1 d\mu(x) \in \mathcal{P}_c(\mathbb{R}^{2d})$  and  $\gamma^2 = \int \gamma_x^2 d\mu(x) \in \mathcal{P}_c(\mathbb{R}^{2d})$ . Then, it holds that*

$$W_1(\gamma^1, \gamma^2) \leq \int_{\mathbb{R}^d} W_1(\gamma_x^1, \gamma_x^2) d\mu(x). \quad (2.3)$$

*Proof.* Take  $\xi \in \text{Lip}(\mathbb{R}^{2d})$  with  $\text{Lip}(\xi; \mathbb{R}^{2d}) \leq 1$ . One has that

$$\int_{\mathbb{R}^{2d}} \xi(x, r) d(\gamma^1 - \gamma^2)(x, r) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \xi(x, r) d(\gamma_x^1 - \gamma_x^2)(r) d\mu(x) \leq \int_{\mathbb{R}^d} W_1(\gamma_x^1, \gamma_x^2) d\mu(x)$$

by Kantorovich duality (2.2) since the maps  $r \mapsto \xi(x, r)$  are 1-Lipschitz for all  $x \in \mathbb{R}^d$ . Taking now the supremum over  $\xi \in \text{Lip}(\mathbb{R}^{2d})$  with  $\text{Lip}(\xi; \mathbb{R}^{2d}) \leq 1$  yields the desired estimate, again as a consequence of (2.2).  $\square$

## 2.2. Subdifferential calculus in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$

In this section, we recall some key notions of subdifferential calculus in Wasserstein spaces. We also prove in Proposition 2.14 a general chainrule formula along multi-dimensional families of perturbations for sufficiently regular functional defined over  $\mathcal{P}_c(\mathbb{R}^d)$ . We refer the reader to ([9], Chaps. 9–11) for a thorough introduction to the theory of subdifferential calculus in Wasserstein spaces, as well as to [37] and ([53], Chap. 15) for complementary material.

Let  $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$  be a lower semicontinuous and proper functional with effective domain  $D(\phi) = \{\mu \in \mathcal{P}_2(\mathbb{R}^d) \text{ s.t. } \phi(\mu) < +\infty\}$ . We introduce in the following definition the concept of *extended Fréchet subdifferential* in  $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ , following the terminology of ([9], Chap. 10).

**Definition 2.8** (Extended Wasserstein subdifferential). Let  $\mu \in D(\phi)$ . We say that a transport plan  $\gamma \in \mathcal{P}_2(\mathbb{R}^{2d})$  belongs to the *extended subdifferential*  $\partial\phi(\mu)$  of  $\phi(\cdot)$  at  $\mu$  provided that

- (i)  $\pi_{\#}^1 \gamma = \mu$ .
- (ii) For any  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ , it holds that

$$\phi(\nu) - \phi(\mu) \geq \inf_{\mu \in \Gamma_o^{1,3}(\gamma, \nu)} \int_{\mathbb{R}^{3d}} \langle x_2, x_3 - x_1 \rangle d\mu(x_1, x_2, x_3) + o(W_2(\mu, \nu)),$$

where  $\Gamma_o^{1,3}(\gamma, \nu) = \{\mu \in \Gamma(\gamma, \nu) \text{ s.t. } \pi_{\#}^{1,3} \mu \in \Gamma_o(\mu, \nu)\}$ .

We furthermore say that a transport plan  $\gamma \in \mathcal{P}_2(\mathbb{R}^{2d})$  belongs to the *strong extended subdifferential*  $\partial_S\phi(\mu)$  of  $\phi(\cdot)$  at  $\mu$  if the stronger condition

$$\phi(\nu) - \phi(\mu) \geq \int_{\mathbb{R}^{3d}} \langle x_2, x_3 - x_1 \rangle d\mu(x_1, x_2, x_3) + o(W_{2,\mu}(\mu, \nu)), \quad (2.4)$$

holds for any  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$  and  $\mu \in \Gamma(\gamma, \nu)$ , where the quantity  $W_{2,\mu}(\mu, \nu)$  is defined by

$$W_{2,\mu}(\mu, \nu) = \left( \int_{\mathbb{R}^{2d}} |x_1 - x_3|^2 d\mu(x_1, x_2, x_3) \right)^{1/2}.$$

We proceed by recalling the notions of *regularity* and *metric slope* that are instrumental in deriving a sufficient condition for the extended subdifferential of a functional to be non-empty. This result is stated in Theorem 2.10 and its proof can be found in Theorem 10.3.10 from [9].

**Definition 2.9** (Regular functionals over  $(\mathcal{P}_2(\mathbb{R}^d), W_2)$  and metric slope). A proper and lower semicontinuous functional  $\phi(\cdot)$  is said to be *regular* provided that whenever  $(\mu_n) \subset \mathcal{P}_2(\mathbb{R}^d)$  and  $(\gamma_n) \subset \mathcal{P}_2(\mathbb{R}^{2d})$  are taken such that

$$\begin{cases} \mu_n \xrightarrow[n \rightarrow +\infty]{W_2} \mu, & \phi(\mu_n) \xrightarrow[n \rightarrow +\infty]{} \tilde{\phi}, \\ \gamma_n \in \partial_S\phi(\mu_n), & \gamma_n \xrightarrow[n \rightarrow +\infty]{W_2} \gamma, \end{cases}$$

it implies that  $\gamma \in \partial\phi(\mu)$  and  $\tilde{\phi} = \phi(\mu)$ . Furthermore, we define the *metric slope*  $|\partial\phi|(\mu)$  of the functional  $\phi(\cdot)$  at  $\mu \in D(\phi)$  as

$$|\partial\phi|(\mu) = \limsup_{\nu \rightarrow \mu} \left[ \frac{(\phi(\mu) - \phi(\nu))^+}{W_2(\mu, \nu)} \right],$$

where  $(\bullet)^+$  denotes the positive part.

**Theorem 2.10** (Link between extended subdifferentials and metric slopes). *Let  $\phi(\cdot)$  be a proper, bounded from below, lower semicontinuous and regular functional over  $\mathcal{P}_2(\mathbb{R}^d)$ . Then, the extended subdifferential  $\partial\phi(\mu)$  of  $\phi(\cdot)$  at some  $\mu \in D(\phi)$  is non-empty if and only if  $|\partial\phi|(\mu) < +\infty$ .*

*In this case, there exists a unique minimal selection in  $\partial\phi(\mu)$  – denoted by  $\partial^\circ\phi(\mu)$  – which satisfies*

$$\left( \int_{\mathbb{R}^{2d}} |r|^2 d(\partial^\circ\phi(\mu))(x, r) \right)^{1/2} = \min_{\gamma} \left\{ \left( \int_{\mathbb{R}^{2d}} |r|^2 d\gamma(x, r) \right)^{1/2} \text{ s.t. } \gamma \in \partial\phi(\mu) \right\} = |\partial\phi|(\mu).$$

*This minimal selection can moreover be explicitly characterized as follows. Let  $\mu_\tau$  be the minimizer of the Moreau-Yosida functional*

$$\Phi_M(\mu, \tau; \cdot) : \nu \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \frac{1}{2\tau} W_2^2(\mu, \nu) + \phi(\nu) \quad (2.5)$$

*at some  $\tau \in (0, \tau_*)$  with  $\tau_* > 0$  small. Then, there exists a family of strong subdifferentials  $(\gamma_\tau) \subset (\partial_S\phi(\mu_\tau))$  which converges towards  $\partial^\circ\phi(\mu)$  in the  $W_2$ -metric along a suitable vanishing sequence  $\tau_n \downarrow 0$ .*

The main reason for resorting to the abstract notion of measure subdifferentials in the context of our argument is that the approximation property of the minimal selection by a sequence of strong subdifferentials plays a key role in the proof of the general Wasserstein chainrule of Proposition 2.14 below. In the sequel however, we will mainly use the simpler notion of *classical Wasserstein differentials* which we introduce in the following definition.

**Definition 2.11** (Classical Wasserstein subdifferentials and superdifferentials). Let  $\mu \in D(\phi)$ . We say that a map  $\xi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu)$  belongs to the *classical subdifferential*  $\partial^-\phi(\mu)$  of  $\phi(\cdot)$  at  $\mu$  provided that

$$\phi(\nu) - \phi(\mu) \geq \sup_{\gamma \in \Gamma_o(\mu, \nu)} \int_{\mathbb{R}^{2d}} \langle \xi(x), y - x \rangle d\gamma(x, y) + o(W_2(\mu, \nu))$$

for all  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ . Similarly, we say that a map  $\xi \in L^2(\mathbb{R}^d, \mathbb{R}^d; \mu)$  belongs to the *classical superdifferential*  $\partial^+\phi(\mu)$  of  $\phi(\cdot)$  at  $\mu$  if  $(-\xi) \in \partial^-(-\phi)(\mu)$ .

It has been proven recently in [37] that the definition of classical Wasserstein subdifferential involving a supremum taken over the set of optimal transport plans is equivalent to the usual one introduced in [9] which involves an infimum. This allows for the elaboration of a convenient notion of differentiability in Wasserstein spaces as detailed below.

**Definition 2.12** (Differentiable functionals in  $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ ). A functional  $\phi : \mathcal{P}_2(\mathbb{R}^d) \mapsto \mathbb{R}$  is said to be *Wasserstein-differentiable* at some  $\mu \in D(\phi)$  if  $\partial^-\phi(\mu) \cap \partial^+\phi(\mu) \neq \emptyset$ . In this case, there exists a unique element  $\nabla_\mu\phi(\mu) \in \partial^-\phi(\mu) \cap \partial^+\phi(\mu)$  called the *Wasserstein gradient* of  $\phi(\cdot)$  at  $\mu$ , which satisfies

$$\phi(\nu) - \phi(\mu) = \int_{\mathbb{R}^{2d}} \langle \nabla_\mu\phi(\mu)(x), y - x \rangle d\gamma(x, y) + o(W_2(\mu, \nu)), \quad (2.6)$$

for any  $\nu \in \mathcal{P}_2(\mathbb{R}^d)$  and any  $\gamma \in \Gamma_o(\mu, \nu)$ .

**Remark 2.13** (Consistency of Defs. 2.8 and 2.12). Let it be noted that the general statements of Definition 2.8 and Theorem 2.10 are consistent with the simpler ones provided in Definition 2.12. Indeed, if a functional  $\phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is differentiable at some  $\mu \in D(\phi)$ , then the plan  $(\text{Id} \times \nabla_\mu \phi(\mu))_{\#} \mu$  is the minimal selection  $\partial^\circ \phi(\mu)$  in its extended Fréchet subdifferential at  $\mu$ . This follows from the fact that by definition  $(\text{Id} \times \nabla_\mu \phi(\mu))_{\#} \mu \in \partial \phi(\mu)$ , and from the estimate  $\|\nabla_\mu \phi(\mu)\|_{L^2(\mu)} \leq |\partial \phi|(\mu)$ . The latter can be obtained by taking derivatives of  $\phi(\cdot)$  along elements of the tangent space to  $\mathcal{P}_2(\mathbb{R}^d)$  (see [9], Chaps. 8 and 10) and then by using the fact that  $\nabla_\mu \phi(\mu)(\cdot)$  is a tangent vector as shown in ([37], Thm. 3.10).

We conclude these recalls by stating in Proposition 2.14 below a new chainrule formula for Wasserstein-differentiable functionals along suitable multi-dimensional perturbations of a measure.

**Proposition 2.14** (Chainrule along multidimensional perturbations by smooth vector fields). *Let  $K \subset \mathbb{R}^d$  be a compact set and  $\mu \in \mathcal{P}(K)$ . Suppose that  $\phi : \mathcal{P}(K) \rightarrow \mathbb{R}$  is Lipschitz in the  $W_2$ -metric, regular in the sense of Definition 2.9 and Wasserstein-differentiable over  $\mathcal{P}(K)$ . Given  $N \geq 1$  and a small parameter  $\epsilon > 0$ , suppose that  $\mathcal{G} \in C^0([-\epsilon, \epsilon]^N \times K, \mathbb{R}^d)$  is a map satisfying the following assumptions.*

- (i)  $\mathcal{G}(0, \cdot) = \text{Id}$  and  $e \mapsto \mathcal{G}(e, x)$  is Fréchet-differentiable at  $e = 0$  uniformly with respect to  $x \in K$ .
- (ii)  $\text{supp}(\mathcal{G}(e, \cdot)_{\#} \mu) \subset K$  for all  $e \in [-\epsilon, \epsilon]^N$ .
- (iii) The directional derivative map

$$\mathcal{F}_\sigma : x \in K \mapsto D_e \mathcal{G}(0, x) \sigma = \sum_{k=1}^N \sigma_k \mathcal{F}_k(x)$$

is continuous for all  $\sigma \in [-\epsilon, \epsilon]^N$ .

Then, the map  $e \in [-\epsilon, \epsilon]^N \mapsto \phi(\mathcal{G}(e, \cdot)_{\#} \mu)$  is Fréchet-differentiable at  $e = 0$  and

$$\nabla_e (\phi(\mathcal{G}(e, \cdot)_{\#} \mu))|_{e=0}(\sigma) = \int_{\mathbb{R}^d} \langle \nabla_\mu \phi(\mu)(x), D_e \mathcal{G}(0, x) \sigma \rangle d\mu(x) = \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \langle \nabla_\mu \phi(\mu)(x), \mathcal{F}_k(x) \rangle d\mu(x),$$

for any  $\sigma \in [-\epsilon, \epsilon]^N$ .

*Proof.* By assumption,  $\phi(\cdot)$  is Lipschitz over  $\mathcal{P}(K)$  in the  $W_2$ -metric, regular in the sense of Definition 2.9 and Wasserstein-differentiable in the sense of Definition 2.12 at  $\mu$ . Hence,  $|\partial \phi|(\mu) < +\infty$  and  $\partial^\circ \phi(\mu) = (\text{Id} \times \nabla_\mu \phi(\mu))_{\#} \mu$ .

Let  $(\tau_n) \subset (0, \tau_*)$  be a suitable vanishing sequence,  $(\mu_{\tau_n}) \subset D(\phi)$  be a sequence of minimizers of the Moreau-Yosida functional (2.5) and  $(\gamma_{\tau_n}) \subset (\partial_S \phi(\mu_{\tau_n}))$  be the corresponding sequence of strong subdifferentials converging in the  $W_2$ -metric towards  $(\text{Id} \times \nabla_\mu \phi(\mu))_{\#} \mu$ . For any  $\sigma \in [-\epsilon, \epsilon]^N$ , define the family of 3-plans  $\mu_\sigma^{\tau_n} = (\pi^1, \pi^2, \mathcal{G}(\sigma, \cdot) \circ \pi^1)_{\#} \gamma_{\tau_n}$ . By (2.4), it holds that

$$\frac{\phi(\mathcal{G}(\sigma, \cdot)_{\#} \mu_{\tau_n}) - \phi(\mu_{\tau_n})}{|\sigma|} \geq \int_{\mathbb{R}^{2d}} \left\langle r, \frac{\mathcal{G}(\sigma, x) - x}{|\sigma|} \right\rangle d\gamma_{\tau_n}(x, r) + o(1),$$

since  $o(W_2, \mu_\sigma^{\tau_n}(\mu_{\tau_n}, \mathcal{G}(\sigma, \cdot)_{\#} \mu_{\tau_n})) = o(\|\mathcal{G}(\sigma, \cdot) - \text{Id}\|_{L^2(\mu_{\tau_n})}) = o(|\sigma|)$ . For all  $\sigma \in [-\epsilon, \epsilon]^N$ , the maps  $(x, r) \mapsto \langle r, (\mathcal{G}(\sigma, x) - x)/|\sigma| \rangle$  are continuous and uniformly integrable with respect to the family of measures  $\{\gamma_{\tau_n}\}_{n=1}^{+\infty}$ .

By a classical convergence result for sequences of measures (see *e.g.* [9], Lem. 5.1.7), we recover that

$$\begin{aligned} \int_{\mathbb{R}^{2d}} \left\langle r, \frac{\mathcal{G}(\sigma, x) - x}{|\sigma|} \right\rangle d\gamma_{\tau_n}(x, r) &\xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}^{2d}} \left\langle r, \frac{\mathcal{G}(\sigma, x) - x}{|\sigma|} \right\rangle d(\mathcal{D}^\circ \phi(\mu))(x, r) \\ &= \int_{\mathbb{R}^d} \left\langle \nabla_\mu \phi(\mu)(x), \frac{\mathcal{G}(\sigma, x) - x}{|\sigma|} \right\rangle d\mu(x). \end{aligned} \quad (2.7)$$

Furthermore, the continuity of  $x \mapsto \mathcal{G}(\sigma, x)$  uniformly with respect to  $\sigma \in [-\epsilon, \epsilon]^N$  together with the Lipschitzianity of  $\phi(\cdot)$  with respect to the  $W_2$ -topology implies that

$$\frac{\phi(\mathcal{G}(\sigma, \cdot) \# \mu_{\tau_n}) - \phi(\mu_{\tau_n})}{|\sigma|} \xrightarrow{n \rightarrow +\infty} \frac{\phi(\mathcal{G}(\sigma, \cdot) \# \mu_\tau) - \phi(\mu_\tau)}{|\sigma|}. \quad (2.8)$$

Merging together (2.7) and (2.8) and applying Vitali's Convergence Theorem to the  $\mu$ -uniformly integrable family of maps  $\sigma \mapsto \langle \nabla_\mu \phi(\mu)(\cdot), (\mathcal{G}(\sigma, \cdot) - \text{Id})/|\sigma| \rangle$  (see *e.g.* [7], Exercise 1.18), we obtain that

$$\frac{\phi(\mathcal{G}(\sigma, \cdot) \# \mu) - \phi(\mu)}{|\sigma|} \geq \int_{\mathbb{R}^d} \langle \nabla_\mu \phi(\mu)(x), D_e \mathcal{G}(0, x) \sigma \rangle d\mu(x) + o(1) = \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \langle \nabla_\mu \phi(\mu)(x), \mathcal{F}_k(x) \rangle d\mu(x) + o(1).$$

Using the fact that  $(\text{Id} \times (-\nabla_\mu \phi(\mu))) \# \mu$  is the minimal selection in the extended Fréchet subdifferential of  $(-\phi(\cdot))$  as a consequence of Remark 2.13, we recover the converse inequality.  $\square$

We list in Appendix A a series of commonly encountered functionals which are both regular in the sense of Definition 2.9 differentiable in the sense of Definition 2.12, and we provide their Wasserstein gradients. This list was already presented along the same lines in our previous work [14], and we add it here for self-containedness.

### 2.3. The continuity equation with non-local velocities in $\mathbb{R}^d$

In this section, we introduce the continuity equation with non-local velocities in  $(\mathcal{P}_c(\mathbb{R}^d), W_1)$ . This equation is commonly written as

$$\partial_t \mu(t) + \nabla \cdot (v[\mu(t)](t, \cdot) \mu(t)) = 0, \quad (2.9)$$

where  $t \mapsto \mu(t)$  is a narrowly continuous family of probability measures on  $\mathbb{R}^d$  and  $(t, x) \mapsto v[\mu](t, x)$  is a Borel family of vector fields satisfying the condition

$$\int_0^T \int_{\mathbb{R}^d} |v[\mu(t)](t, x)| d\mu(t)(x) dt < +\infty. \quad (2.10)$$

Equation (2.9) has to be understood in duality with smooth and compactly supported functions, *i.e.*

$$\int_0^T \int_{\mathbb{R}^d} \left( \partial_t \xi(t, x) + \left\langle \nabla_x \xi(t, x), v[\mu(t)](t, x) \right\rangle \right) d\mu(t)(x) dt = 0 \quad (2.11)$$

for all  $\xi \in C_c^\infty([0, T] \times \mathbb{R}^d)$ . This definition can be alternatively written as

$$\frac{d}{dt} \int_{\mathbb{R}^d} \xi(x) d\mu(t)(x) = \int_{\mathbb{R}^d} \langle \nabla \xi(x), v[\mu(t)](t, x) \rangle d\mu(t)(x) \quad (2.12)$$

for all  $\xi \in C_c^\infty(\mathbb{R}^d)$  and  $\mathcal{L}^1$ -almost every  $t \in [0, T]$ .

We recall in Theorem 2.15 the classical existence, uniqueness and representation formula for solutions of non-local PDEs. Although these results were first derived in [8], we state here a version explored in [45, 46] which is better suited to our smoother control-theoretic framework.

**Theorem 2.15** (Existence, uniqueness and representation of solutions for (2.9)). *Consider a non-local velocity field  $v[\cdot](\cdot, \cdot)$  defined as*

$$v : \mu \in \mathcal{P}_c(\mathbb{R}^d) \mapsto v[\mu](\cdot, \cdot) \in L^1(\mathbb{R}, C^1(\mathbb{R}^d, \mathbb{R}^d)), \quad (2.13)$$

and satisfying the following assumptions

**(H')**

(i) *There exist positive constants  $L_1$  and  $M$  such that*

$$|v[\mu](t, x)| \leq M(1 + |x|) \quad \text{and} \quad |v[\mu](t, x) - v[\mu](t, y)| \leq L_1|x - y|$$

*for every  $\mu \in \mathcal{P}_c(\mathbb{R}^d)$ ,  $t \in \mathbb{R}$  and  $(x, y) \in \mathbb{R}^{2d}$ .*

(ii) *There exists a positive constant  $L_2$  such that*

$$\|v[\mu](t, \cdot) - v[\nu](t, \cdot)\|_{C^0(\mathbb{R}^d)} \leq L_2 W_1(\mu, \nu)$$

*for every  $\mu, \nu \in \mathcal{P}_c(\mathbb{R}^d)$  and  $t \in \mathbb{R}$ .*

Then, for every initial datum  $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$ , the Cauchy problem

$$\begin{cases} \partial_t \mu(t) + \nabla \cdot (v[\mu(t)](t, \cdot) \mu(t)) = 0 \\ \mu(0) = \mu^0, \end{cases} \quad (2.14)$$

admits a unique solution  $\mu(\cdot) \in \text{Lip}_{\text{loc}}(\mathbb{R}_+, \mathcal{P}_c(\mathbb{R}^d))$ . If  $\mu^0$  is absolutely continuous with respect to  $\mathcal{L}^d$ , then  $\mu(t)$  is absolutely continuous with respect to  $\mathcal{L}^d$  as well for all times  $t \in \mathbb{R}_+$ . Furthermore for every  $T > 0$  and every  $\mu^0, \nu^0 \in \mathcal{P}_c(\mathbb{R}^d)$ , there exists positive constants  $R_T, L_T > 0$  such that

$$\text{supp}(\mu(t)) \subset \overline{B(0, R_T)} \quad \text{and} \quad W_1(\mu(t), \nu(t)) \leq L_T W_1(\mu^0, \nu^0),$$

for all times  $t \in [0, T]$ , where  $\mu(\cdot), \nu(\cdot)$  are solutions of (2.14) with initial conditions  $\mu^0, \nu^0$ , respectively.

Let  $\mu(\cdot)$  be the unique solution of (2.14) and  $(\Phi_{(0,t)}^v[\mu^0](\cdot))_{t \geq 0}$  be the family of flows of diffeomorphisms generated by the non-autonomous velocity field  $(t, x) \mapsto v[\mu(t)](t, x)$ , i.e.

$$\begin{cases} \partial_t \Phi_{(0,t)}^v[\mu^0](x) = v[\mu(t)](t, \Phi_{(0,t)}^v[\mu^0](x)), \\ \Phi_{(0,0)}^v[\mu^0](x) = x \quad \text{for all } x \text{ in } \mathbb{R}^d. \end{cases} \quad (2.15)$$

Then, the curve  $\mu(\cdot)$  is given explicitly by the pushforward formula

$$\mu(t) = \Phi_{(0,t)}^v[\mu^0](\cdot) \# \mu^0, \quad (2.16)$$

for all times  $t \in [0, T]$ .

In the following proposition, we recall a standard result which links the differential in space of the flow of diffeomorphisms of an ODE to the solution of the corresponding linearized Cauchy problem (see e.g. [15], Thm. 2.3.1).

**Proposition 2.16** (Classical differential of a non-local flow of diffeomorphisms). *Let  $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$  and let  $(\Phi_{(0,t)}^v[\mu^0](\cdot))_{t \in [0,T]}$  be the flows of diffeomorphisms generated by a non-local velocity field  $v[\cdot](\cdot, \cdot)$  satisfying hypotheses **(H')**. Then, the flow map  $x \mapsto \Phi_{(s,t)}^v[\mu(s)](x)$  is Fréchet differentiable over  $\mathbb{R}^d$  for any  $s, t \in [0, T]$ . Moreover, its differential  $D_x \Phi_{(s,t)}^v[\mu(s)](x)h$  in a direction  $h \in \mathbb{R}^d$  is the unique solution of the linearized Cauchy problem*

$$\begin{cases} \partial_t w(t, x) = D_x v[\mu(t)] \left( t, \Phi_{(s,t)}^v[\mu(s)](x) \right) w(t, x), \\ w(s, x) = h, \end{cases} \quad (2.17)$$

for any  $s, t \in [0, T]$ .

In our previous work [14], we extended the classical result of Proposition 2.16 to the Wasserstein setting in order to compute derivatives of the flow maps  $\Phi_{(0,t)}^v[\mu^0](\cdot)$  with respect to their initial measure  $\mu^0$ . In the following proposition, we state a further refinement of this result to the case in which the initial measure is perturbed by a multi-dimensional family of maps, in the spirit of the chainrule stated in Proposition 2.14.

**Proposition 2.17** (Wasserstein differential of a non-local flow of diffeomorphisms). *Let  $K \subset \mathbb{R}^d$  be a compact set,  $\mu^0 \in \mathcal{P}(K)$ ,  $v[\cdot](\cdot, \cdot)$  be a non-local velocity field satisfying hypotheses **(H')** of Theorem 2.15 and  $(\Phi_{(0,t)}^v[\mu^0](\cdot))_{t \in [0,T]}$  be its associated flow of diffeomorphisms. Suppose moreover that for any  $i \in \{1, \dots, d\}$ , the maps  $\mu \mapsto v^i[\mu](t, x)$  are regular and differentiable over  $\mathcal{P}(K)$  uniformly with respect to  $(t, x) \in [0, T] \times K$ . Given  $N \geq 1$  and a small parameter  $\epsilon > 0$ , let  $\mathcal{G} \in C^0([-\epsilon, \epsilon]^N \times K, \mathbb{R}^d)$  be a function satisfying the hypotheses of Proposition 2.14.*

*Then, the map  $e \in [-\epsilon, \epsilon]^N \mapsto \Phi_{(s,t)}^v[\mathcal{G}(e, \cdot) \# \mu](x)$  is Fréchet-differentiable at  $e = 0$  and its differential  $w_\sigma(\cdot, x)$  in an arbitrary direction  $\sigma \in [-\epsilon, \epsilon]^N$  can be expressed as*

$$w_\sigma(t, x) = \sum_{k=1}^N \sigma_k w_k(t, x),$$

where for any  $k \in \{1, \dots, N\}$ , the map  $w_k(\cdot, x)$  is the unique solution of the non-local Cauchy problem

$$\begin{cases} \partial_t w_k(t, x) = D_x v[\mu(t)] \left( t, \Phi_{(s,t)}^v[\mu](x) \right) w_k(t, x) \\ \quad + \int_{\mathbb{R}^d} \mathbb{I}^v_{(t, \Phi_{(s,t)}^v[\mu](x))} \left( \Phi_{(s,t)}^v[\mu](y) \right) \left( D_x \Phi_{(s,t)}^v[\mu](y) \mathcal{F}_k(y) + w_k(t, y) \right) d\mu(y) \\ w_k(s, x) = 0. \end{cases} \quad (2.18)$$

Here,  $(t, x, y) \mapsto \mathbb{I}^v_{(t,x)}(y) \in \mathbb{R}^{d \times d}$  is the matrix-valued map whose rows are the Wasserstein gradients of the components  $v^i[\cdot](t, x)$  of the non-local velocity field at  $\mu(t)$ , i.e.

$$\left( \mathbb{I}^v_{(t,x)}(y) \right)_{i,j} = \left( \nabla_\mu (v^i[\cdot](t, x)) (\mu(t))(y) \right)_j, \quad (2.19)$$

for any  $i, j \in \{1, \dots, d\}$ .

*Proof.* By Proposition 2.14 and as a consequence of our hypotheses on  $v[\cdot](\cdot, \cdot)$ , we know that the map  $e \in [-\epsilon, \epsilon]^N \mapsto \Phi_{(s,t)}^v[\mathcal{G}(e, \cdot) \# \mu](x)$  is Fréchet-differentiable at  $e = 0$ . Therefore, the action of its differential on a given direction  $\sigma \in [-\epsilon, \epsilon]^N$  can be expressed in coordinates using partial derivatives, *i.e.*

$$w_\sigma(t, x) = \sum_{k=1}^N \sigma_k \partial_{e_k} \left( \Phi_{(s,t)}^v[\mathcal{G}(e, \cdot) \# \mu](x) \right) (0).$$

Moreover, it has been proven in ([14], Prop. 5) that such one-dimensional variations could be characterized as the unique solution of the linearized Cauchy problems (2.18).  $\square$

## 2.4. Non-smooth multiplier rule and differentiable extension of functions

In this section, we recall some facts of non-smooth analysis as well as a non-smooth Lagrange multiplier rule which is instrumental in the proof of our main result. This multiplier rule is expressed in terms of the so-called *Michel-Penot subdifferential*, see *e.g.* [40, 44]. In the sequel, we denote by  $(X, \|\cdot\|_X)$  a separable Banach space and by  $X^*$  its topological dual associated with the duality bracket  $\langle \cdot, \cdot \rangle_X$ . Given a map  $f : X \rightarrow \mathbb{R}$ , we denote by  $D(f) = \{x \in X \text{ s.t. } f(x) < +\infty\}$  its effective domain.

**Definition 2.18** (Michel-Penot subdifferential). Given a map  $f : X \rightarrow \mathbb{R}$ , the *Michel-Penot subdifferential* (MP-subdifferential in the sequel) of  $f(\cdot)$  at some  $x \in D(f)$  is defined by

$$\partial_{\text{MP}} f(x) = \left\{ \xi \in X^* \text{ s.t. } \langle \xi, h \rangle_X \leq d_{\text{MP}} f(x; h) \text{ for all } h \in X \right\},$$

where

$$d_{\text{MP}} f(x; h) = \sup_{e \in X} \limsup_{t \downarrow 0} \left[ \frac{f(x + t(e + h)) - f(x + te)}{t} \right]$$

denotes the so-called *Michel-Penot derivative* of  $f(\cdot)$  at  $x$  in the direction  $h$ . Moreover, if  $f : X \rightarrow \mathbb{R}$  is locally convex around  $x \in X$ , then its Michel-Penot and convex subdifferentials coincide, *i.e.*  $\partial_{\text{MP}} f(x) = \partial f(x)$ .

The MP-subdifferential – smaller than the Clarke subdifferential – bears the nice property of shrinking to a singleton whenever the functional  $f(\cdot)$  is merely Fréchet-differentiable. It also enjoys a summation rule and a chained-derivative formula for compositions of locally Lipschitz and Fréchet-differentiable maps. We list these properties in the following proposition.

**Proposition 2.19** (Properties of the Michel-Penot subdifferentials). *Let  $x \in X$ ,  $f, g : X \rightarrow \mathbb{R}$  and  $\mathcal{G} : X^N \rightarrow X$ .*

- (a) *If  $f(\cdot)$  is Fréchet-differentiable at  $x$ , then  $\partial_{\text{MP}} f(x) = \{\nabla f(x)\}$ .*
- (b) *If  $d_{\text{MP}} f(x; h) < +\infty$  and  $d_{\text{MP}} g(x; h) < +\infty$  for any  $h \in X$ , it holds that*

$$\partial_{\text{MP}}(f + g)(x) \subseteq \left( \partial_{\text{MP}} f(x) + \partial_{\text{MP}} g(x) \right).$$

- (c) *If  $\mathcal{G}(\cdot)$  is Fréchet-differentiable at  $0 \in \mathbb{R}^N$  and  $f(\cdot)$  is Lipschitz in a neighbourhood of  $\mathcal{G}(0)$ , one has that*

$$d_{\text{MP}}(f \circ \mathcal{G})(0; \sigma) = \langle \xi, D\mathcal{G}(0)\sigma \rangle_X, \tag{2.20}$$

*for some  $\xi \in \partial_{\text{MP}} f(\mathcal{G}(0))$ . In other words,  $\partial_{\text{MP}}(f \circ \mathcal{G})(0) \subseteq D\mathcal{G}(0)^* \circ \partial_{\text{MP}} f(\mathcal{G}(0))$ .*

These properties can be proven easily by computing explicitly the Michel-Penot derivatives of the corresponding maps and using the definition of the set  $\partial_{\text{MP}}(\bullet)$ , see *e.g.* [51]. Another useful feature of this notion of

subdifferential is that it allows to write Lagrange multiplier rules for locally Lipschitz functions. This family of optimality conditions was initially derived in [40] and refined in [51] where the author extended the result to the class of so-called *calm* functions.

**Definition 2.20** (Calm functions). A map  $f : X \rightarrow \mathbb{R}$  is *calm* at  $x \in X$  provided that the following holds.

- (i) There exists a constant  $L > 0$  such that for any  $\delta \in X$  with  $\|\delta\|_X$  sufficiently small, it holds that

$$\|f(x + \delta) - f(x)\|_X \leq L \|\delta\|_X.$$

- (ii)  $d_{\text{MP}}f(x; h) < +\infty$  for any  $h \in X$ .

**Theorem 2.21** (Multiplier rule for Michel-Penot subdifferentials). *Let  $f_0, \dots, f_n, g_1, \dots, g_m : X \rightarrow \mathbb{R}$  and  $\Omega \subset X$  be a closed and convex set. Suppose that  $x_*$  is a local solution of the non-linear optimization problem*

$$\begin{cases} \min_{x \in \Omega} [f_0(x)] \\ \text{s.t.} \begin{cases} f_i(x) \leq 0 & \text{for all } i \in \{1, \dots, n\}, \\ g_j(x) = 0 & \text{for all } j \in \{1, \dots, m\}, \end{cases} \end{cases}$$

and that the maps  $f_0(\cdot), \dots, f_n(\cdot), g_1(\cdot), \dots, g_m(\cdot)$  are calm at  $x_*$ . Then, there exist Lagrange multipliers  $(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m) \in \{0, 1\} \times \mathbb{R}_+^n \times \mathbb{R}^m$  such that the following stationarity (S), non-triviality (NT) and complementary-slackness (CS) conditions hold

$$\begin{cases} 0 \in \partial_{\text{MP}} \left( \lambda_0 f_0(\cdot) + \sum_{i=1}^n \lambda_i f_i(\cdot) + \sum_{j=1}^m \eta_j g_j(\cdot) \right) (x_*) + \mathcal{N}(\Omega, x_*), & \text{(S)} \\ \lambda_0 + \sum_{i=1}^n \lambda_i + \sum_{j=1}^m |\eta_j| = 1, & \text{(NT)} \\ \lambda_i f_i(x_*) = 0 \text{ for all } i \in \{1, \dots, n\}, & \text{(CS)} \end{cases}$$

where  $\mathcal{N}(\Omega, x_*)$  denotes the normal cone of convex analysis to  $\Omega$  at  $x_*$ .

We end this introductory section by stating a Lusin-type lemma for vector-valued functions and a derivative-preserving continuous extension result that will both prove to be useful in the sequel. We refer the reader *e.g.* to [28] for notions on Bochner integrals and abstract integration in Banach spaces.

**Lemma 2.22** (Pointwise convergence and restriction). *Let  $f : [0, T] \rightarrow X$  be an  $L^1$ -function in the sense of Bochner and  $\mathcal{F}$  be any subset of  $[0, T]$  with full Lebesgue measure. Then, there exist  $\mathcal{A}, \mathcal{M} \subset \mathcal{F}$  respectively with null and full Lebesgue measure satisfying the property that for any  $\tau \in \mathcal{M}$ , there exists  $(\tau_k) \subset \mathcal{A}$  such that*

$$\tau_k \xrightarrow[k \rightarrow +\infty]{} \tau \quad \text{and} \quad \|f(\tau) - f(\tau_k)\|_X \xrightarrow[k \rightarrow +\infty]{} 0.$$

*Proof.* This result is a consequence of ([51], Lem. 4.1) along with Lusin's Theorem for vector valued maps.  $\square$

**Lemma 2.23** (A continuous extension preserving the derivative). *Let  $\epsilon > 0$  and  $f : [0, \epsilon]^N \rightarrow X$  be a continuous map differentiable at  $e = 0$  relatively to  $\mathbb{R}_+^N$ . Then, there exists a continuous extension  $\tilde{f} : [-\bar{\epsilon}_N, \bar{\epsilon}_N]^N \rightarrow X$  of  $f(\cdot)$  which is Fréchet-differentiable at  $e = 0$  and such that  $D_e \tilde{f}(0) = D_e f(0)$ .*

*Proof.* We adapt here a simple proof that can be found *e.g.* in ([51], Lem. 2.11). Define the map

$$g : e \in \mathbb{R}_+^N \mapsto \frac{1}{|e|} \left( f(e) - f(0) - D_e f(0)e \right) \in X.$$

By definition,  $g(\cdot)$  is continuous over  $\mathbb{R}_+^N \setminus \{0\}$  and can be extended to  $\mathbb{R}_+^N$  by imposing that  $g(0) = 0$  since  $f(\cdot)$  is differentiable at  $e = 0$  relatively to  $\mathbb{R}_+^N$ . Invoking Dugundji's extension theorem (see [29]), we can define a continuous extension  $\tilde{g}(\cdot)$  of  $g(\cdot)$  on the whole of  $\mathbb{R}^N$ .

We now define the auxiliary map  $\tilde{f} : e \in \mathbb{R}^N \mapsto f(0) + D_e f(0)e + |e|\tilde{g}(e)$ . By construction,  $\tilde{f}(\cdot)$  is continuous and coincides with  $f(\cdot)$  over  $\mathbb{R}_+^N$ . Moreover, one has for any  $e \in \mathbb{R}^N$  that

$$\tilde{f}(e) - \tilde{f}(0) = D_e f(0)e + |e|\tilde{g}(e) = D_e f(0)e + o(|e|)$$

by continuity of  $\tilde{g}(\cdot)$  at 0. Therefore, the map  $\tilde{f}(\cdot)$  is Fréchet-differentiable at  $e = 0$  with  $D_e \tilde{f}(0) = D_e f(0)$ .  $\square$

### 3. PROOF OF THE MAIN RESULT

In this section we prove the main result of this article, that is Theorem 3.1 below. We recall that this result is a first-order necessary optimality condition of Pontryagin-type for the general Wasserstein optimal control problem  $(\mathcal{P})$  defined in the Introduction.

**Theorem 3.1** (Pontryagin Maximum Principle for  $(\mathcal{P})$ ). *Let  $(u^*(\cdot), \mu^*(\cdot)) \in \mathcal{U} \times \text{Lip}([0, T], \mathcal{P}_c(\mathbb{R}^d))$  be an optimal pair control-trajectory for  $(\mathcal{P})$  and assume that the set of hypotheses **(H)** below holds.*

#### (H)

**(H1)** *The set of admissible controls is defined as  $\mathcal{U} = L^\infty([0, T], U)$  where  $U$  is any  $C^0$ -closed subset of  $\{\omega \in C^1(\mathbb{R}^d, \mathbb{R}^d) \text{ s.t. } \|\omega(\cdot)\|_{C^0(\mathbb{R}^d, \mathbb{R}^d)} + \text{Lip}(\omega(\cdot); \mathbb{R}^d) \leq L_U\}$  for a given constant  $L_U > 0$ .*

**(H2)** *The non-local velocity field  $\mu \in \mathcal{P}_c(\mathbb{R}^d) \mapsto v[\mu] \in L^1([0, T], C^1(\mathbb{R}^d, \mathbb{R}^d))$  satisfies the classical Cauchy-Lipschitz assumptions in Wasserstein spaces, i.e. there exist positive constants  $M, L_1$  and  $L_2$  such that*

$$\begin{aligned} |v[\mu](t, x)| &\leq M(1 + |x|), \quad |v[\mu](t, x) - v[\mu](t, y)| \leq L_1|x - y|, \\ \|v[\mu](t, \cdot) - v[\nu](t, \cdot)\|_{C^0(\mathbb{R}^d)} &\leq L_2 W_1(\mu, \nu) \end{aligned}$$

for all  $(x, y) \in \mathbb{R}^d$  and  $\mathcal{L}^1$ -almost every  $t \in [0, T]$ .

**(H3)** *The maps  $\mu \mapsto v^i[\mu](t, x) \in \mathbb{R}$  are regular in the sense of Definition 2.9 over  $\mathcal{P}(K)$  for any compact set  $K \subset \mathbb{R}^d$  and Wasserstein-differentiable at  $\mu^*(t)$ , uniformly with respect to  $(t, x) \in [0, T] \times \mathbb{R}^d$ . The maps  $\mu \mapsto D_x v[\mu](t, x) \in \mathbb{R}^d$  and  $(y, z) \mapsto \nabla_\mu (v^i[\cdot](t, y))(\mu)(z) \in \mathbb{R}^d$  are continuous for all  $i \in \{1, \dots, d\}$ , uniformly with respect to  $(t, x) \in [0, T] \times \mathbb{R}^d$ .*

- (H4) The final cost  $\mu \mapsto \varphi(\mu) \in \mathbb{R}$  and the boundary constraints maps  $\mu \mapsto (\Psi_i^I(\mu), \Psi_j^E(\mu))_{i,j} \in \mathbb{R}^{n+m}$  are bounded and Lipschitz continuous in the  $W_2$ -metric over  $\mathcal{P}(K)$  for any compact set  $K \subset \mathbb{R}^d$ . They are furthermore regular in the sense of Definition 2.9 and Wasserstein-differentiable at  $\mu^*(T)$ . The Wasserstein gradients  $\nabla_\mu \varphi(\mu^*(T))(\cdot)$ ,  $\nabla_\mu \Psi_i^I(\mu^*(T))(\cdot)$  and  $\nabla_\mu \Psi_j^E(\mu^*(T))(\cdot)$  are continuous for any  $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$ .
- (H5) The running cost  $(t, \mu, \omega) \mapsto L(t, \mu, \omega) \in \mathbb{R}$  is  $\mathcal{L}^1$ -measurable with respect to  $t \in [0, T]$ , bounded, and Lipschitz continuous with respect to the product  $W_2 \times C^0$ -metric defined over  $\mathcal{P}(K) \times U$  for any compact set  $K \subset \mathbb{R}^d$ . It is furthermore regular in the sense of Definition 2.9 and Wasserstein-differentiable at  $\mu^*(t)$  for any  $(t, \omega) \in [0, T] \times U$ , and its Wasserstein gradient  $\nabla_\mu L(t, \mu^*(t), \omega)(\cdot)$  is continuous.
- (H6) The state constraints maps  $(t, \mu) \mapsto \Lambda_l(t, \mu) \in \mathbb{R}$  are bounded and Lipschitz-continuous over  $[0, T] \times \mathcal{P}(K)$  and regular in the sense of Definition 2.9, for any compact set  $K \subset \mathbb{R}^d$  and any  $l \in \{1, \dots, r\}$ . Moreover, the maps  $(t, \mu) \mapsto \partial_t \Lambda_l(t, \mu)$  and  $(t, \mu) \mapsto \nabla_\mu \Lambda_l(t, \mu)(\cdot)$  are well-defined and continuously differentiable at  $(t, \mu^*(t))$  with

$$\begin{aligned} \nabla_\mu \partial_t \Lambda_l(t, \mu^*(t))(\cdot) &= \partial_t \nabla_\mu \Lambda_l(t, \mu^*(t))(\cdot) \in C^0(K, \mathbb{R}^d), \quad D_x \nabla_\mu \Lambda_l(t, \mu^*(t))(\cdot) \in C^0(K, \mathbb{R}^d), \\ \nabla_\mu [\nabla_\mu \Lambda_l(t, \mu^*(t))(\cdot)](\cdot) &\in C^0(K^2, \mathbb{R}^d) \end{aligned}$$

Then there exists a constant  $R'_T > 0$ , Lagrange multipliers  $(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1, \dots, \varpi_r) \in \{0, 1\} \times \mathbb{R}_+^n \times \mathbb{R}^m \times \mathcal{M}_+([0, T])^r$  and a curve  $\nu^*(\cdot) \in \text{Lip}([0, T], \overline{\mathcal{P}(B_{2d}(0, R'_T))})$  such that the following holds.

(i) The map  $t \mapsto \nu^*(t)$  is a solution of the forward-backward Hamiltonian system of continuity equations

$$\begin{cases} \partial_t \nu^*(t) + \nabla \cdot (\mathbb{J}_{2d} \nabla_\nu \mathcal{H}_{\lambda_0}(t, \nu^*(t), \zeta^*(t), u^*(t))(\cdot, \cdot) \nu^*(t)) = 0 & \text{in } [0, T] \times \mathbb{R}^{2d}, \\ \pi_{\#}^1 \nu^*(t) = \mu^*(t) & \text{for all } t \in [0, T], \\ \pi_{\#}^2 \nu^*(T) = (-\nabla_\mu \mathcal{S}(\mu^*(T)))_{\#} \mu^*(T), \end{cases} \quad (3.1)$$

where  $\mathbb{J}_{2d}$  is the symplectic matrix of  $\mathbb{R}^{2d}$ . The augmented infinite-dimensional Hamiltonian  $\mathcal{H}_{\lambda_0}(\cdot, \cdot, \cdot, \cdot)$  of the system is defined by

$$\mathcal{H}_{\lambda_0}(t, \nu, \zeta, \omega) = \int_{\mathbb{R}^{2d}} \langle r, v[\pi_{\#}^1 \nu](t, x) + \omega(x) \rangle d\nu(x, r) - \lambda_0 L(t, \pi_{\#}^1 \nu, \omega) - \mathcal{C}(t, \pi_{\#}^1 \nu, \zeta, \omega), \quad (3.2)$$

for any  $(t, \nu, \zeta, \omega) \in [0, T] \times \overline{\mathcal{P}(B_{2d}(0, R'_T))} \times \mathbb{R}^r \times U$ . The penalized state constraints and final gradient maps are given respectively by

$$\mathcal{C}(t, \mu, \zeta, \omega) = \sum_{l=1}^r \zeta_l \left( \partial_t \Lambda_l(t, \mu) + \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(t, \mu)(x), v[\mu](t, x) + \omega(x) \rangle d\mu(x) \right), \quad (3.3)$$

and

$$\nabla_\mu \mathcal{S}(\mu) = \lambda_0 \nabla_\mu \varphi(\mu) + \sum_{i=1}^n \lambda_i \nabla_\mu \Psi_i^I(\mu) + \sum_{j=1}^m \eta_j \nabla_\mu \Psi_j^E(\mu). \quad (3.4)$$

For any  $l \in \{1, \dots, r\}$ , the map  $t \in [0, T] \mapsto \zeta_l^*(t) \in \mathbb{R}_+$  denotes the cumulated state constraints multiplier associated with  $\varpi_l$ , defined by

$$\zeta_l^*(t) = \mathbf{1}_{[0, T)}(t) \int_t^T d\varpi_l(s),$$

(ii) The Lagrange multipliers satisfy the non-degeneracy condition

$$(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1, \dots, \varpi_r) \neq 0. \quad (3.5)$$

as well as the complementary slackness conditions

$$\begin{cases} \lambda_i \Psi_i^I(\mu^*(T)) = 0 & \text{for all } i \in \{1, \dots, n\}, \\ \text{supp}(\varpi_l) = \left\{ t \in [0, T] \text{ s.t. } \Lambda_l(t, \mu(t)) = 0 \right\} & \text{for all } l \in \{1, \dots, r\}. \end{cases}$$

(iii) The Pontryagin maximization condition

$$\mathcal{H}_{\lambda_0}(t, \nu^*(t), \zeta^*(t), u^*(t)) = \max_{\omega \in U} [\mathcal{H}_{\lambda_0}(t, \nu^*(t), \zeta^*(t), \omega)] \quad (3.6)$$

holds for  $\mathcal{L}^1$ -almost every  $t \in [0, T]$ .

**Remark 3.2** (The Gamkrelidze Maximum Principle). The so-called Gamkrelidze formulation of the PMP corresponds to the case in which one includes the derivative of the state constraints inside the Hamiltonian function. The consequence of this choice is that the costate variables are absolutely continuous in time instead of being merely BV as it is the case in the more classical formulation of the constrained PMP (see *e.g.* [54], Chap. 9).

As already mentioned in the Introduction, this fact is crucial for our purpose since absolutely continuous curves in Wasserstein spaces are exactly those curves which solve a continuity equation. Hence, one cannot derive a Hamiltonian system such as (3.1) by sticking to the classical formulation of the PMP.

**Remark 3.3** (On the regularity hypothesis **(H1)**). One of the distinctive features of continuity equations in Wasserstein spaces, compared to other families of PDEs is that they require Cauchy-Lipschitz assumptions on the driving velocity fields in order to be classically well-posed for arbitrary initial data. Even though the existence theory has gone far beyond this basic setting, notably through the DiPerna-Lions-Ambrosio theory (see [6, 27] or [9], Sect. 8.2), such extensions come at the price of losing the strict micro-macro correspondence of the solutions embodied by the underlying flow-structure. Therefore, from a mean-field control-theoretic viewpoint, it seemed more meaningful for the author to work in a setting where classical well-posedness holds for the optimal trajectory.

Furthermore, the proof of Theorem 3.1 relies heavily on the geometric flow of diffeomorphism structure of the underlying characteristic system of ODEs, both forward and backward in time. For this reason, the Lipschitz-regularity assumption **(H1)** is instrumental in our argumentation.

Let it be remarked however that there exists common examples of Wasserstein optimal control problems for which the optimal control is  $C^1$ -smooth in space. Such a situation is given *e.g.* by controlled vector fields of the form  $u(t, x) = \sum_{k=1}^m u_k(t) X_k(\cdot)$  where  $X_1, \dots, X_m \in C^1(\mathbb{R}^d, \mathbb{R}^d)$  and  $u_1, \dots, u_m \in L^\infty([0, T], \mathbb{R})$ , or by non-linear controlled vector field  $(t, x, \mu, u) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_c(\mathbb{R}^d) \times \mathbb{R}^m \mapsto v[\mu](t, x, u) \in \mathbb{R}^d$ .

We divide the proof of Theorem 3.1 into four steps. In Step 1, we introduce the concept of *packages of needle-like variations* of an optimal control and compute the corresponding perturbations induced on the optimal trajectory. In Step 2, we apply the non-smooth Lagrange multiplier rule of Theorem 2.21 to the sequence of finite-dimensional optimization problem formulated on the length of the needle variations, to obtain a family

of finite-dimensional optimality conditions at time  $T$ . We introduce in Step 3 a suitable notion of costate, allowing to propagate this family of optimality condition backward in time, yielding the PMP with a relaxed maximization condition restricted to a countable subset of needle parameters. The full result is then recovered in Step 4 through a limiting procedure combined with several approximation arguments.

### Step 1: Packages of needle-like variations.

We start by considering an optimal pair control-trajectory  $(u^*(\cdot), \mu^*(\cdot)) \in \mathcal{U} \times \text{Lip}([0, T], \mathcal{P}(\overline{B(0, R_T)}))$  where  $R_T > 0$  is given by Theorem 2.15. Let  $\mathcal{T} \subset [0, T]$  be the set of Lebesgue points of  $t \in [0, T] \mapsto (v[\mu^*(t)](t, \cdot), u^*(t, \cdot), L(t, \mu^*(t), \cdot)) \in C^0(\overline{B(0, R_T)}, \mathbb{R}^d) \times U \times C^0(U, \mathbb{R})$  in the sense of Bochner's integral (see e.g. [28], Thm. 9). This set has full Lebesgue measure in  $[0, T]$ , and Lemma 2.22 yields the existence of two subsets  $\mathcal{A}, \mathcal{M} \subset \mathcal{T}$ , having respectively null and full Lebesgue measure, such that for any  $\tau \in \mathcal{M}$ , there exists  $(\tau_k) \subset \mathcal{A}$  converging towards  $\tau$  and such that

$$\|u^*(\tau, \cdot) - u^*(\tau_k, \cdot)\|_{C^0(\mathbb{R}^d, \mathbb{R}^d)} \xrightarrow[k \rightarrow +\infty]{} 0 \quad \text{and} \quad \|L(\tau, \mu^*(\tau), \cdot) - L(\tau_k, \mu^*(\tau_k), \cdot)\|_{C^0(U, \mathbb{R})} \xrightarrow[k \rightarrow +\infty]{} 0.$$

We further denote by  $U_D$  a countable and dense subset of the set of admissible control values  $U$  which is compact and separable in the  $C^0$ -topology as a consequence of **(H1)**.

**Definition 3.4** (Package of needle-like variations). Let  $(u^*(\cdot), \mu^*(\cdot)) \in \mathcal{U} \times \text{Lip}([0, T], \mathcal{P}_c(\mathbb{R}^d))$  be an optimal pair control-trajectory. Given  $N \geq 1$ , a family of elements  $\{(\omega_k, \tau_k)\}_{k=1}^N \subset U_D \times \mathcal{A}$  and  $e = (e_1, \dots, e_N) \in [0, \bar{\epsilon}_N]^N$  such that  $[\tau_i - e_i, \tau_i] \cap [\tau_j - e_j, \tau_j] = \emptyset$  for all distinct pairs  $i, j \in \{1, \dots, N\}$ , we define the  $N$ -package of needle-like variations of  $u^*(\cdot)$  by

$$\tilde{u}_e^N \equiv \tilde{u}_e : t \mapsto \begin{cases} \omega_k & \text{if } t \in [\tau_k - e_k, \tau_k], \\ u^*(t) & \text{otherwise.} \end{cases} \quad (3.7)$$

We also denote by  $t \mapsto \tilde{\mu}_e(\cdot)$  the corresponding perturbed trajectory, i.e. the solution of (2.14) associated with the controlled non-local velocity field  $v[\cdot](\cdot, \cdot) + \tilde{u}_e(\cdot, \cdot)$ .

This class of variations is known in the literature of control theory to generate admissible perturbations of the optimal control without any assumption on the structure of the control set  $\mathcal{U}$ , while allowing for an explicit and tractable computation of the relationship between the perturbed and optimal states (see e.g. [15]).

In the following lemma, we make use of the geometric structure of solutions to non-local transport equations presented in Theorem 2.15 together with some notations borrowed from Proposition 2.17 to express  $\tilde{\mu}_e(t)$  as a function of  $\mu^*(t)$  for all times  $t \in [0, T]$ . In the sequel, we denote by  $\Phi_{(s,t)}^{v,u}[\mu(s)](\cdot)$  the flow map generated by the non-local velocity field  $v[\cdot](\cdot, \cdot) + u(\cdot, \cdot)$  between times  $s$  and  $t$ , defined as in (2.15).

**Lemma 3.5** (First-order perturbation induced by a package of needle-like variations in the non-local case). *There exists a family of maps  $(\mathcal{G}_t^N(\cdot, \cdot))_{t \in [0, T]} \subset C^0([-\bar{\epsilon}_N, \bar{\epsilon}_N]^N \times \mathbb{R}^d, \mathbb{R}^d)$  such that*

(i) *For all  $(t, e) \in [0, T] \times [0, \bar{\epsilon}_N]^N$ , the perturbed measures  $\tilde{\mu}_e(t)$  satisfy*

$$\tilde{\mu}_e(t) = \mathcal{G}_t^N(e, \cdot) \# \mu^*(t). \quad (3.8)$$

- (ii) *For all  $(t, e) \in [0, T] \times [0, \bar{\epsilon}_N]^N$ , the maps  $\mathcal{G}_t^N(e, \cdot)$  are  $C^1$ -diffeomorphisms over  $\overline{B(0, R_T)}$ .*  
 (iii) *There exists a constant  $R_T^\Phi > 0$  depending on  $R_T, L_U$  such that for all  $(t, e) \in [0, T] \times [0, \bar{\epsilon}_N]^N$  one has  $\text{supp}(\mathcal{G}_t^N(e, \cdot) \# \mu^*(T)) \subset \overline{B(0, R_T^\Phi)}$*   
 (iv) *The map  $e \in [0, \bar{\epsilon}_N]^N \mapsto \mathcal{G}_t^N(e, \cdot)$  is Fréchet-differentiable at  $e = 0$  with respect to the  $C^0(\overline{B(0, R_T)}, \mathbb{R}^d)$ -norm, uniformly with respect to  $t \in [0, T]$ . The corresponding Taylor expansion can be written explicitly*

as

$$\mathcal{G}_t^N(e, \cdot) = \text{Id} + \sum_{k=1}^{\iota(t)} e_k \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}[\mu^*(t)](\cdot) + o(|e|), \quad (3.9)$$

where  $\iota(t) \in \{1, \dots, N\}$  is the biggest index such that  $\tau_{\iota(t)} \leq t \leq \tau_{\iota(t)+1} - e_{\iota(t)+1}$ . For all  $x \in \overline{B(0, R_T)}$  and  $k \in \{1, \dots, N\}$ , the map  $t \in [\tau_k, T] \mapsto \mathcal{F}_t^{\omega_k, \tau_k}(x)$  is the unique solution of the non-local Cauchy problem

$$\begin{cases} \partial_t \mathcal{F}_t^{\omega_k, \tau_k}(x) = \left( D_x u^* \left( t, \Phi_{(\tau_k, t)}^{v, u^*}(x) \right) + D_x v[\mu^*(t)] \left( t, \Phi_{(\tau_k, t)}^{v, u^*}(x) \right) \right) \mathcal{F}_t^{\omega_k, \tau_k}(x) \\ \quad + \int_{\mathbb{R}^d} \Pi^v \left( t, \Phi_{(\tau_k, t)}^{v, u^*}(x) \right) \left( \Phi_{(\tau_k, t)}^{v, u^*}(y) \right) \mathcal{F}_t^{\omega_k, \tau_k}(y) d\mu^*(\tau_k)(y), \\ \mathcal{F}_{\tau_k}^{\omega_k, \tau_k}(x) = \omega_k(x) - u^*(\tau_k, x). \end{cases} \quad (3.10)$$

*Proof.* The proof of this result is similar to that of ([14], Lem. 5), with some extra technicalities arising from the induction argument performed on the non-local terms. By definition of a package of needle-like variations, the perturbed controls  $\tilde{u}_e(\cdot, \cdot)$  generate well-defined flows of diffeomorphisms  $(\Phi_{(0, t)}^{v, \tilde{u}_e}[\mu^0](\cdot))_{t \in [0, T]}$  which are such that

$$\tilde{\mu}_e(t) = \Phi_{(0, t)}^{v, \tilde{u}_e}[\mu^0] \circ \Phi_{(t, 0)}^{v, u^*}[\mu^*(t)](\cdot) \# \mu^*(t).$$

Hence, items (i), (ii) and (iii) hold for any  $e \in [0, \epsilon_N]^N$  with  $\mathcal{G}_t^N(e, \cdot) = \Phi_{(0, t)}^{v, \tilde{u}_e}[\mu^0] \circ \Phi_{(t, 0)}^{v, u^*}[\mu^*(t)](\cdot)$ .

We focus our attention on the proof by induction of (iv). Let  $t \in [0, T]$  be such that  $\iota(t) = 1$ . By (3.7), one has that

$$\begin{aligned} \tilde{\mu}_e(t) &= \Phi_{(\tau_1, t)}^{v, u^*}[\tilde{\mu}_e(\tau_1)] \circ \Phi_{(\tau_1 - e_1, \tau_1)}^{v, \omega_1}[\mu^*(\tau_1 - e_1)] \circ \Phi_{(t, \tau_1 - e_1)}^{v, u^*}[\mu^*(t)](\cdot) \# \mu^*(t) \\ &= \Phi_{(\tau_1, t)}^{v, u^*}[\tilde{\mu}_e(\tau_1)] \circ \Phi_{(\tau_1 - e_1, \tau_1)}^{v, \omega_1}[\mu^*(\tau_1 - e_1)] \circ \Phi_{(\tau_1, \tau_1 - e_1)}^{v, u^*}[\mu^*(\tau_1)] \circ \Phi_{(t, \tau_1)}^{v, u^*}[\mu^*(t)](\cdot) \# \mu^*(t). \end{aligned}$$

Invoking Lebesgue's Differentiation Theorem (see *e.g.* [7], Cor. 2.23) along with the continuity of  $e \mapsto v[\tilde{\mu}_e(t)](t, \cdot)$  in the  $C^0$ -norm topology, it holds that

$$\begin{aligned} &\Phi_{(\tau_1 - e_1, \tau_1)}^{v, \omega_1}[\mu^*(\tau_1 - e_1)](x) \\ &= x + \int_{\tau_1 - e_1}^{\tau_1} \left( v[\tilde{\mu}_e(t)] \left( t, \Phi_{(\tau_1 - e_1, t)}^{v, \omega_1}[\mu^*(\tau_1 - e_1)](x) \right) + \omega_1 \left( \Phi_{(\tau_1 - e_1, t)}^{v, \omega_1}[\mu^*(\tau_1 - e_1)](x) \right) \right) dt \\ &= x + e_1 \left( v[\mu^*(\tau_1)](\tau_1, x) + \omega_1(x) \right) + o(e_1), \end{aligned}$$

as well as

$$\begin{aligned} \Phi_{(\tau_1, \tau_1 - e_1)}^{v, u^*}[\mu^*(\tau_1)](x) &= x - \int_{\tau_1 - e_1}^{\tau_1} \left( v[\tilde{\mu}_e(t)] \left( t, \Phi_{(\tau_1, t)}^{v, u^*}[\mu^*(\tau_1)](x) \right) + u^* \left( t, \Phi_{(\tau_1, t)}^{v, u^*}[\mu^*(\tau_1)](x) \right) \right) dt \\ &= x - e_1 \left( v[\mu^*(\tau_1)](\tau_1, x) + u^*(\tau_1, x) \right) + o(e_1). \end{aligned}$$

Chaining these two expansions, we obtain that

$$\begin{aligned} \tilde{\mu}_e(\tau_1) &= \Phi_{(\tau_1 - e_1, \tau_1)}^{v, \omega_1}[\mu^*(\tau_1 - e_1)] \circ \Phi_{(\tau_1, \tau_1 - e_1)}^{v, u^*}[\mu^*(\tau_1)](\cdot) \# \mu^*(\tau_1) \\ &= (\text{Id} + e_1[\omega_1(\cdot) - u^*(\tau_1, \cdot)] + o(e_1)) \# \mu^*(\tau_1). \end{aligned}$$

We can now proceed to compute the induced first-order expansion on the non-local flows  $\Phi_{(\tau_1, t)}^{v, u^*}[\tilde{\mu}_e(\tau_1)](\cdot)$  as follows

$$\begin{aligned} & \Phi_{(\tau_1, t)}^{v, u^*}[\tilde{\mu}_e(\tau_1)]\left(x + e_1 [\omega_1(x) - u^*(\tau_1, x)] + o(e_1)\right) \\ &= \Phi_{(\tau_1, t)}^{v, u^*}[\tilde{\mu}_e(\tau_1)](x) + e_1 D_x \Phi_{(\tau_1, t)}^{v, u^*}[\tilde{\mu}_e(\tau_1)](x) \cdot [\omega_1(x) - u^*(\tau_1, x)] + o(e_1) \\ &= \Phi_{(\tau_1, t)}^{v, u^*}[\mu^*(\tau_1)](x) + e_1 \left( D_x \Phi_{(\tau_1, t)}^{v, u^*}[\mu^*(\tau_1)](x) \cdot [\omega_1(x) - u^*(\tau_1, x)] + w_1(t, x) \right) + o(e_1), \end{aligned}$$

where  $w_1(\cdot, \cdot)$  is defined as in Proposition 2.17, and where we used the fact that the  $e \mapsto D_x \Phi_{(\tau_1, \cdot)}^{v, u^*}[\tilde{\mu}_e(\cdot)](\cdot)$  is continuous as a consequence of hypothesis **(H1)** and **(H2)**. Introducing for all times  $t \in [\tau_1, T]$  the map

$$\mathcal{F}_t^{\omega_1, \tau_1} : x \in \overline{B(0, R_T)} \mapsto D_x \Phi_{(\tau_1, t)}^{v, u^*}[\mu^*(\tau_1)](x) [\omega_1(x) - u^*(\tau_1, x)] + w_1(t, x)$$

and invoking the statements of Propositions 2.16 and 2.17, we have that both (3.9) and (3.10) hold for any  $e_1 \in [0, \bar{\epsilon}_N]$  and all times  $t \in [0, T]$  such that  $\iota(t) = 1$ .

Let us now assume that (3.9) and (3.10) hold for all times  $t \in [0, T]$  such that  $\iota(t) = k - 1$ , *i.e.*

$$\tilde{\mu}_e(t) = \mathcal{G}_t^N(e, \cdot) \# \mu^*(t) = \left( \text{Id} + \sum_{l=1}^{k-1} e_l \mathcal{F}_t^{\omega_l, \tau_k} \circ \Phi_{(t, \tau_l)}^{v, u^*}[\mu^*(t)](\cdot) + o(|e|) \right) \# \mu^*(t), \quad (3.11)$$

for  $e \in [0, \bar{\epsilon}_N]^N$ . By definition (3.7) of an  $N$ -package of needle-like variations, we have that

$$\tilde{\mu}_e(\tau_k) = \Phi_{(\tau_k - e_k, \tau_k)}^{v, \omega_k}[\tilde{\mu}_e(\tau_k - e_k)] \circ \Phi_{(\tau_k, \tau_k - e_k)}^{v, u^*}[\tilde{\mu}_e(\tau_k)] \circ \Phi_{(\tau_{k-1}, \tau_k)}^{v, u^*}[\tilde{\mu}_e(\tau_{k-1})](\cdot) \# \tilde{\mu}_e(\tau_{k-1}).$$

As in the initialization step, we can write using Lebesgue's Differentiation Theorem that

$$\Phi_{(\tau_k - e_k, \tau_k)}^{v, \omega_k}[\tilde{\mu}_e(\tau_k - e_k)] \circ \Phi_{(\tau_k, \tau_k - e_k)}^{v, u^*}[\tilde{\mu}_e(\tau_k)](x) = x + e_k [\omega_k(x) - u^*(\tau_k, x)] + o(e_k). \quad (3.12)$$

Furthermore, invoking the induction hypothesis (3.11) and the results of Proposition 2.14, we obtain that

$$\begin{aligned} & \Phi_{(\tau_{k-1}, \tau_k)}^{v, u^*}[\tilde{\mu}_e(\tau_{k-1})] \left( x + \sum_{l=1}^{k-1} \left( e_l \mathcal{F}_{\tau_{k-1}}^{\omega_l, \tau_l} \circ \Phi_{(\tau_{k-1}, \tau_l)}^{v, u^*}(x) + o(e_l) \right) \right) \\ &= \Phi_{(\tau_{k-1}, \tau_k)}^{v, u^*}[\tilde{\mu}_e(\tau_{k-1})](x) + \sum_{l=1}^{k-1} \left( e_l D_x \Phi_{(\tau_{k-1}, \tau_k)}^{v, u^*}[\tilde{\mu}_e(\tau_{k-1})](x) \mathcal{F}_{\tau_{k-1}}^{\omega_l, \tau_l} \circ \Phi_{(\tau_{k-1}, \tau_l)}^{v, u^*}(x) + o(e_l) \right) \\ &= \Phi_{(\tau_{k-1}, \tau_k)}^{v, u^*}[\mu^*(\tau_{k-1})](x) + \sum_{l=1}^{k-1} e_l \left( D_x \Phi_{(\tau_{k-1}, \tau_k)}^{v, u^*}[\mu^*(\tau_{k-1})](x) \mathcal{F}_{\tau_{k-1}}^{\omega_l, \tau_l} \circ \Phi_{(\tau_{k-1}, \tau_l)}^{v, u^*}(x) + w_l(\tau_k, x) + o(e_l) \right) \end{aligned} \quad (3.13)$$

where the maps  $(w_l(\cdot, \cdot))_{1 \leq l \leq k-1}$  are defined as in Proposition 2.17 with  $\mathcal{F}_l(\cdot) \equiv \mathcal{F}_{\tau_{k-1}}^{\omega_l, \tau_l} \circ \Phi_{(\tau_{k-1}, \tau_l)}^{v, u^*}[\mu^*(\tau_{k-1})](\cdot)$ . Plugging together equation (2.17) of Proposition 2.16 and equation (2.18) of Proposition 2.17, one can see that the maps

$$t \in [\tau_{k-1}, \tau_k] \mapsto D_x \Phi_{(\tau_{k-1}, t)}^{v, u^*}[\mu^*(\tau_{k-1})] \left( \Phi_{(\tau_l, \tau_{k-1})}^{v, u^*}[\mu^*(\tau_l)](x) \right) \mathcal{F}_{\tau_{k-1}}^{\omega_l, \tau_l}(x) + w_l \left( t, \Phi_{(\tau_l, \tau_{k-1})}^{v, u^*}[\mu^*(\tau_l)](x) \right)$$

are solutions of (3.10) on  $[\tau_{k-1}, \tau_k]$  with initial condition  $\mathcal{F}_{\tau_{k-1}}^{\omega_l, \tau_l}(\cdot)$  at time  $\tau_{k-1}$  for any  $l \in \{1, \dots, k-1\}$ . By Cauchy-Lipschitz uniqueness, we can therefore extend the definition of the maps  $t \mapsto \mathcal{F}_t^{\omega_l, \tau_l}(x)$  to the whole of  $[\tau_l, \tau_k]$  for any  $l \in \{1, \dots, k-1\}$ .

Chaining the expansions (3.12) and (3.13) along with our previous extension argument, we obtain that both (3.9) and (3.10) hold up to time  $\tau_k$ , *i.e.*

$$\mathcal{G}_{\tau_k}^N(e, \cdot) = \text{Id} + \sum_{l=1}^k e_l \mathcal{F}_{\tau_k}^{\omega_l, \tau_l} \circ \Phi_{(\tau_k, \tau_l)}^{u^*} [\mu^*(\tau_k)](\cdot) + o(|e|)$$

for any  $e \in [0, \bar{\epsilon}_N]^N$ . Performing yet another coupled Taylor expansion of the same form on the expression

$$\tilde{\mu}_e(t) = \Phi_{(\tau_k, t)}^{v, u^*} [\tilde{\mu}_e(\tau_k)](\cdot) \# \tilde{\mu}_e(\tau_k),$$

and invoking the same extension argument yields the full induction step for all times  $t \in [0, T]$  such that  $\iota(t) = k$ . Hence, we have proven that item (iv) holds for all  $e \in [0, \bar{\epsilon}_N]^N$ . Using Lemma 2.23, we can now extend the map  $e \in [0, \bar{\epsilon}_N]^N \mapsto \mathcal{G}_t^N(e, \cdot) \in C^0(\overline{B(0, R_T)}, \mathbb{R}^d)$  to the whole of  $[-\bar{\epsilon}_N, \bar{\epsilon}_N]^N$  in a continuous and bounded way, uniformly with respect to  $t \in [0, T]$ , while preserving its differential at  $e = 0$ .  $\square$

In the sequel, we drop the explicit dependence of the flow maps on their starting measures and adopt the simplified notation  $\Phi_{(s, t)}^{v, u^*}(x) \equiv \Phi_{(s, t)}^{v, u^*}[\mu(s)](x)$  for clarity and conciseness.

## Step 2: First-order optimality condition.

In Lemma 3.5, we derived the analytical expression of the first-order perturbation induced by a  $N$ -package of needle-like variations on the solution of a controlled non-local continuity equation. By the very definition of an  $N$ -package of needle-like variations, we know that the finite-dimensional optimization problem

$$(\mathcal{P}_N) \quad \begin{cases} \min_{e \in [0, \bar{\epsilon}_N]^N} \left[ \int_0^T L(t, \tilde{\mu}_e(t), \tilde{u}_e(t)) dt + \varphi(\tilde{\mu}_e(T)) \right] \\ \text{s.t.} \quad \begin{cases} \Psi^E(\tilde{\mu}_e(T)) = 0, & \Psi^I(\tilde{\mu}_e(T)) \leq 0, \\ \max_{t \in [0, T]} \Lambda(t, \tilde{\mu}_e(t)) \leq 0, \end{cases} \end{cases}$$

admits  $e = 0$  as an optimal solution in  $[0, \bar{\epsilon}_N]^N$  for  $\bar{\epsilon}_N$  small enough.

In the following lemma, we check that the functionals involved in  $(\mathcal{P}_N)$  meet the requirements of the Lagrange multiplier rule stated in Theorem 2.21. We also compute their first-order variation induced by the package of needle-like variations at  $e = 0$ .

**Lemma 3.6** (Differentiability and calmness of the functionals involved in  $(\mathcal{P}_N)$ ). *The maps  $e \in [-\bar{\epsilon}_N, \bar{\epsilon}_N]^N \mapsto \varphi(\tilde{\mu}_e(T))$ ,  $\Psi^E(\tilde{\mu}_e(T))$ ,  $\Psi^I(\tilde{\mu}_e(T))$  and  $e \in [-\bar{\epsilon}_N, \bar{\epsilon}_N]^N \mapsto \int_0^T L(t, \tilde{\mu}_e(t), \tilde{u}_e(t))dt$  are calm and Fréchet-differentiable at  $e = 0$ . Their Fréchet derivative in a direction  $\sigma \in [0, \bar{\epsilon}_N]^N$  are respectively given by*

$$\left\{ \begin{array}{l} \nabla_e(\varphi(\tilde{\mu}_e(T)))|_{e=0}(\sigma) = \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \varphi(\mu^*(T))(x), \mathcal{F}_T^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x), \\ \nabla_e(\Psi_i^I(\tilde{\mu}_e(T)))|_{e=0}(\sigma) = \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Psi_i^I(\mu^*(T))(x), \mathcal{F}_T^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x), \\ \nabla_e(\Psi_j^E(\tilde{\mu}_e(T)))|_{e=0}(\sigma) = \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Psi_j^E(\mu^*(T))(x), \mathcal{F}_T^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x), \end{array} \right. \quad (3.14)$$

for any  $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$  and

$$\begin{aligned} \nabla_e \left( \int_0^T L(t, \tilde{\mu}_e(t), \tilde{u}_e(t)) dt \right) |_{e=0}(\sigma) &= \sum_{k=1}^N \sigma_k \int_{\tau_k}^T \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} L(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x) dt \\ &\quad + \sum_{k=1}^N \sigma_k \left( L(\tau_k, \mu^*(\tau_k), u^*(\tau_k)) - L(\tau_k, \mu^*(\tau_k), \omega_k) \right). \end{aligned} \quad (3.15)$$

The maps  $e \in [-\bar{\epsilon}_N, \bar{\epsilon}_N]^N \mapsto \max_{t \in [0, T]} \Lambda_1(t, \tilde{\mu}_e(t)), \dots, \max_{t \in [0, T]} \Lambda_r(t, \tilde{\mu}_e(t))$  are calm and locally Lipschitz around  $e = 0$ . Their Michel-Penot derivatives in a direction  $\sigma \in [0, \bar{\epsilon}_N]^N$  are given by

$$d_{\text{MP}} \left( \max_{t \in [0, T]} \Lambda_l(t, \tilde{\mu}_e(t)) \right) (0; \sigma) = \sum_{k=1}^N \sigma_k \int_0^T \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t) d\varpi_l^N(t), \quad (3.16)$$

where the Borel measures  $\varpi_l^N \in \mathcal{M}_+([0, T])$  are such that  $\text{supp}(\varpi_l^N) = \{t \in [0, T] \text{ s.t. } \Lambda_l(t, \mu^*(t)) = 0\}$  and  $\|\varpi_l^N\|_{TV} = 1$  for any  $l \in \{1, \dots, r\}$ .

*Proof.* The calmness property of the maps  $e \mapsto \varphi(\tilde{\mu}_e(T))$ ,  $\Psi^E(\tilde{\mu}_e(T))$ ,  $\Psi^I(\tilde{\mu}_e(T))$  and  $e \mapsto \int_0^T L(t, \tilde{\mu}_e(t), \tilde{u}_e(t))dt$  at  $e = 0$  stems from the fact that they are compositions of Fréchet-differentiable and locally Lipschitz mappings as a by-product of hypotheses **(H4)** and **(H5)** and Lemma 3.5. The differentials of the final cost and boundary constraints can be computed with a direct application of Proposition 2.14.

We split the computation of the first-order variation at  $e = 0$  of the running cost functional into two parts. One can first derive that

$$\begin{aligned} \int_0^T \left( L(t, \tilde{\mu}_e(t), \tilde{u}_e(t)) - L(t, \tilde{\mu}_e(t), u^*(t)) \right) dt &= \sum_{k=1}^N \int_{\tau_k - e_k}^{\tau_k} \left( L(t, \tilde{\mu}_e(t), \omega_k) - L(t, \tilde{\mu}_e(t), u^*(t)) \right) dt \\ &= \sum_{k=1}^N e_k \left( L(\tau_k, \tilde{\mu}_e(\tau_k), \omega_k) - L(\tau_k, \tilde{\mu}_e(\tau_k), u^*(\tau_k)) \right) + o(|e|) \\ &= \sum_{k=1}^N e_k \left( L(\tau_k, \mu^*(\tau_k), \omega_k) - L(\tau_k, \mu^*(\tau_k), u^*(\tau_k)) \right) + o(|e|), \end{aligned}$$

by Lebesgue's Differentiation Theorem, since  $\tau_k$  is a Lebesgue points of  $t \in [0, T] \mapsto L(t, \mu^*(t), \cdot) \in C^0(U, \mathbb{R})$  and the maps  $e \mapsto L(\tau_k, \tilde{\mu}_e(\tau_k), u^*(\tau_k))$  and  $e \mapsto L(\tau_k, \tilde{\mu}_e(\tau_k), \omega_k)$  are continuous for all  $k \in \{1, \dots, N\}$ . Furthermore, invoking the Wasserstein chainrule of Proposition 2.14 along with the results of Lemma 3.5, we have that

$$L(t, \tilde{\mu}_e(t), u^*(t)) - L(t, \mu^*(t), u^*(t)) = \sum_{k=1}^N e_k \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} L(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x) + o(|e|).$$

for  $\mathcal{L}^1$ -almost every  $t \in [0, T]$ . Combining these expansions with an application of Lebesgue's Dominated Convergence Theorem yields (3.15).

We now turn our attention to the state constraints functionals. By hypothesis **(H6)** and Proposition 2.14, the maps  $e \mapsto \Lambda_l(t, \tilde{\mu}_e(t))$  are Fréchet-differentiable at  $e = 0$  for any  $l \in \{1, \dots, r\}$ , uniformly with respect to  $t \in [0, T]$ . Moreover, the functional

$$\gamma \in C^0([0, T], \mathbb{R}) \mapsto \max_{t \in [0, T]} \gamma(t)$$

is locally convex and therefore locally Lipschitz over  $C^0([0, T], \mathbb{R})$ . Hence, the maps  $e \mapsto \max_{t \in [0, T]} \Lambda_l(t, \tilde{\mu}_e(t))$  are calm at  $e = 0$  as compositions of Fréchet-differentiable and locally Lipschitz mappings. By Proposition 2.19-(c), we can compute their Michel-Penot derivatives in a direction  $\sigma \in [0, \bar{\epsilon}_N]^N$  as follows

$$d_{\text{MP}} \left( \max_{t \in [0, T]} \Lambda_l(t, \tilde{\mu}_e(t)) \right) (0; \sigma) = \int_0^T \nabla_e \left( \Lambda_l(t, \tilde{\mu}_e(t)) \right) \Big|_{e=0} (\sigma) d\varpi_l^N(t),$$

where  $\varpi_l^N \in \mathcal{M}_+([0, T])$  belongs to the convex subdifferential of the  $C^0([0, T], \mathbb{R})$ -norm evaluated at  $\Lambda_l(\cdot, \mu^*(\cdot))$ . This subdifferential can be classically characterized (see e.g. [41], Sect. 4.5.1) as the set of Borel regular measures such that

$$\|\varpi_l^N\|_{TV} = 1 \text{ and } \text{supp}(\varpi_l^N) = \left\{ t \in [0, T] \text{ s.t. } \Lambda_l(t, \mu^*(t)) = 0 \right\}.$$

Invoking again Proposition 2.14 and Lemma 3.5, we can write the differential of  $e \mapsto \Lambda(t, \tilde{\mu}_e(t))$  at  $e = 0$  evaluated in a direction  $\sigma \in [0, \bar{\epsilon}_N]^N$  as

$$\nabla_e \left( \Lambda_l(t, \tilde{\mu}_e(t)) \right) \Big|_{e=0} (\sigma) = \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x)$$

for all times  $t \in [0, T]$ . Using the measure-theoretic version of Lebesgue's Dominated Convergence Theorem (see e.g. [7], Thm. 1.21), we conclude that (3.16) holds as well, which ends the proof of Lemma 3.6.  $\square$

Using the results of Lemma 3.6, we can apply the Lagrange multiplier rule of Theorem 2.21 to  $(\mathcal{P}_N)$  and obtain the existence of scalar multipliers  $(\lambda_0^N, \lambda_1^N, \dots, \lambda_n^N, \eta_1^N, \dots, \eta_m^N, \theta_1^N, \dots, \theta_r^N) \in \{0, 1\} \times \mathbb{R}_+^n \times \mathbb{R}^m \times \mathbb{R}_+^r$  such

that

$$\left\{ \begin{array}{l} 0 \in \partial_{\text{MP}} \left( \lambda_0^N \varphi(\tilde{\mu}_e(T)) + \lambda_0^N \int_0^T L(t, \tilde{\mu}_e(t), \tilde{u}_e(t)) dt + \sum_{l=1}^r \theta_l^N \max_{t \in [0, T]} \Lambda_l(t, \tilde{\mu}_e(t)), \right. \\ \qquad \qquad \qquad \left. + \sum_{i=1}^n \lambda_i^N \Psi_i^I(\tilde{\mu}_e(T)) + \sum_{j=1}^m \eta_j^N \Psi_j^E(\tilde{\mu}_e(T)) \right)_{|e=0} + \mathcal{N}([0, \bar{\epsilon}_N]^N, 0) \\ \lambda_0^N + \sum_{i=1}^n \lambda_i^N + \sum_{j=1}^m |\eta_j^N| + \sum_{l=1}^r \theta_l^N = 1, \\ \theta_l^N \max_{t \in [0, T]} \Lambda_l(t, \mu^*(t)) = 0 \text{ for all } l \in \{1, \dots, r\}, \\ \lambda_i^N \Psi_i^I(\mu^*(T)) = 0 \text{ for all } i \in \{1, \dots, n\}. \end{array} \right. \quad \begin{array}{l} \text{(S)} \\ \text{(NT)} \\ \text{(CS)} \end{array}$$

Since all the functions involved in the subdifferential inclusion (S) are calm, we can use the summation rule of Proposition 2.19-(b) along with the characterization of MP-subdifferentials for Fréchet-differentiable functionals stated in Proposition 2.19-(a) to obtain that

$$\begin{aligned} -\nabla_e \left( \lambda_0^N \varphi(\tilde{\mu}_e(T)) + \lambda_0^N \int_0^T L(t, \tilde{\mu}_e(t), \tilde{u}_e(t)) dt + \sum_{i=1}^n \lambda_i^N \Psi_i^I(\tilde{\mu}_e(T)) + \sum_{j=1}^m \eta_j^N \Psi_j^E(\tilde{\mu}_e(T)) \right)_{|e=0} \\ \in \sum_{l=1}^r \theta_l^N \partial_{\text{MP}} \left( \max_{t \in [0, T]} \Lambda_l(t, \tilde{\mu}_e(t)) \right)_{|e=0} + \mathcal{N}([0, \bar{\epsilon}_N]^N, 0). \end{aligned}$$

By combining the expressions of the gradients (3.14), (3.15) and the MP-derivative (3.16) derived in Lemma 3.6, along with the composition rule of Proposition 2.19-(c) for MP-subdifferentials, we obtain that

$$\begin{aligned} & \sum_{k=1}^N \sigma_k \int_{\mathbb{R}^d} \langle -\nabla_{\mu} \mathcal{S}_N(\mu^*(T))(x), \mathcal{F}_T^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(T)(x) \\ & - \sum_{k=1}^N \sigma_k \lambda_0^N \left( L(\tau_k, \mu^*(\tau_k), u^*(\tau_k)) - L(\tau_k, \mu^*(\tau_k), \omega_k) \right) \\ & - \sum_{k=1}^N \sigma_k \int_{\tau_k}^T \int_{\mathbb{R}^d} \left\langle \lambda_0^N \nabla_{\mu} L(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x) dt \\ & - \sum_{k=1}^N \sigma_k \sum_{l=1}^r \theta_l^N \int_{\tau_k}^T \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x) d\varpi_l^N(t) \leq 0 \end{aligned}$$

for any direction  $\sigma \in [0, \bar{\epsilon}_N]^N$ , where  $\nabla_{\mu} \mathcal{S}_N(\mu^*(T))(\cdot)$  is defined as in (3.4). By choosing particular vectors  $\sigma \in [0, \bar{\epsilon}_N]^N$  which have all their components except one equal to 0, this family of inequalities can be rewritten

as

$$\begin{aligned}
& \int_{\mathbb{R}^d} \langle -\nabla_{\mu} \mathcal{S}_N(\mu^*(T))(x), \mathcal{F}_T^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(T)(x) - \lambda_0^N \left( L(\tau_k, \mu^*(\tau_k), u^*(\tau_k)) - L(\tau_k, \mu^*(\tau_k), \omega_k) \right) \\
& - \int_{\tau_k}^T \int_{\mathbb{R}^d} \langle \lambda_0^N \nabla_{\mu} L(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x) dt \\
& - \sum_{l=1}^r \int_{\tau_k}^T \int_{\mathbb{R}^d} \langle \nabla_{\mu} \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x) d\varpi_l^N(t) \leq 0
\end{aligned} \tag{3.17}$$

for all  $k \in \{1, \dots, N\}$ , where we redefined the notation  $\varpi_l^N \equiv \theta_l^N \varpi_l^N$ .

### Step 3: Backward dynamics and partial Pontryagin maximization condition.

The next step of our proof is to introduce a suitable notion of state-costate variable transporting the family of inequalities (3.17) derived at time  $T$  to the base points  $(\tau_1, \dots, \tau_N)$  of the needle-like variations while generating a Hamiltonian dynamical structure. To this end, we build for all  $N \geq 1$  a curve  $\nu_N^*(\cdot) \in \text{Lip}([0, T], \mathcal{P}_c(\mathbb{R}^{2d}))$  solution of the forward-backward system of continuity equations

$$\begin{cases} \partial_t \nu_N^*(t) + \nabla \cdot (\mathcal{V}_N^*[\nu_N^*(t)](t, \cdot, \cdot) \nu_N^*(t)) = 0 & \text{in } [0, T] \times \mathbb{R}^{2d}, \\ \pi_{\#}^1 \nu_N^*(t) = \mu^*(t) & \text{for all } t \in [0, T], \\ \nu_N^*(T) = (\text{Id} \times (-\nabla_{\mu} \mathcal{S}_N(\mu^*(T))))_{\#} \mu^*(T). \end{cases} \tag{3.18}$$

Here, the non-local velocity field  $\mathcal{V}_N^*[\cdot](\cdot, \cdot, \cdot)$  is given for  $\mathcal{L}^1$ -almost every  $t \in [0, T]$  any  $(x, r, \nu) \in \mathbb{R}^{2d} \times \mathcal{P}_c(\mathbb{R}^{2d})$  by

$$\mathcal{V}_N^*[\nu](t, x, r) = \begin{pmatrix} v[\pi_{\#}^1 \nu](t, x) + u^*(t, x) \\ \lambda_0^N \nabla_{\mu} L(t, \pi_{\#}^1 \nu, u^*(t))(x) + \nabla_{\mu} \mathcal{E}(t, \pi_{\#}^1 \nu, \zeta_N^*(t), u^*(t))(x) \\ -\mathbb{I}^v[\nu](t, x) - D_x v[\pi_{\#}^1 \nu](t, x)^{\top} r - D_x u^*(t, x)^{\top} r \end{pmatrix},$$

where we introduced the notation

$$\mathbb{I}^v[\nu](t, x) = \int_{\mathbb{R}^{2d}} \left( \mathbb{I}_{(t, y)}^v(x) \right)^{\top} p d\nu(y, p). \tag{3.19}$$

Notice that the transport equation (3.18) does not satisfy the classical hypotheses of Theorem 2.15. Following a methodology introduced in our previous work [14], it is possible to circumvent this difficulty by building explicitly a solution of (3.18) relying on the *cascade structure* of the equations.

**Lemma 3.7** (Definition and well-posedness of solutions of (3.18)). *Let  $(u^*(\cdot), \mu^*(\cdot))$  be an optimal pair control-trajectory for  $(\mathcal{P})$ . For  $\mu^*(T)$ -almost every  $x \in \mathbb{R}^d$ , we consider the family of backward flows  $(\Psi_{(T, t)}^{x, N}(\cdot))_{t \leq T}$*

solution of the non-local Cauchy problems

$$\left\{ \begin{array}{l} \partial_t w_x(t, r) = \lambda_0^N \nabla_\mu L(t, \mu^*(t), u^*(t)) \left( \Phi_{(T,t)}^{v,u^*}(x) \right) + \nabla_\mu \mathcal{C}(t, \mu^*(t), \zeta_N^*(t), u^*(t)) \left( \Phi_{(T,t)}^{v,u^*}(x) \right) \\ \quad - D_x u^* \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right)^\top w_x(t, r) - D_x v[\mu^*(t)] \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right)^\top w_x(t, r) \\ \quad - \int_{\mathbb{R}^{2d}} \Pi^v_{(t, \Phi_{(T,t)}^{v,u^*}(y))} \left( \Phi_{(T,t)}^{v,u^*}(x) \right)^\top w_y(t, p) d((\text{Id} \times (-\nabla_\mu \mathcal{S}_N(\mu^*(T))))_{\#} \mu^*(T))(y, p) \\ w_x(T, r) = r, \end{array} \right. \quad (3.20)$$

and the associated curves of measures

$$\sigma_{x,N}^* : t \mapsto \Psi_{(T,t)}^{x,N}(\cdot)_{\#} \delta_{(-\nabla_\mu \mathcal{S}_N(\mu^*(T)))(x)}.$$

Define the map  $\nu_{T,N}^* : t \in [0, T] \mapsto \int \sigma_{x,N}^*(t) d\mu^*(T)(x) \in \mathcal{P}_c(\mathbb{R}^{2d})$ . Then, the curve  $\nu_N^* : t \in [0, T] \mapsto (\Phi_{(T,t)}^{v,u^*}(\cdot), \text{Id})_{\#} \nu_{T,N}^*(t)$  is the unique solution of (3.18). Moreover, there exists two constants  $R'_T, L'_T > 0$  such that

$$\text{supp}(\nu_N^*(t)) \subset \overline{B_{2d}(0, R'_T)} \quad \text{and} \quad W_1(\nu_N^*(t), \nu_N^*(s)) \leq L'_T |t - s|,$$

for all  $s, t \in [0, T]$ .

*Proof.* Let us denote by  $\Omega \subset \mathbb{R}^{2d}$  a compact set such that

$$\bigcup_{N \geq 1} \text{supp} \left( (\text{Id} \times (-\nabla_\mu \mathcal{S}_N(\mu^*(T))))_{\#} \mu^*(T) \right) \subseteq \Omega.$$

Such a set exists since the maps  $(\nabla_\mu \mathcal{S}_N(\mu^*(T)))(\cdot)$  are continuous by **(H4)**, as well as uniformly bounded as a consequence of the non-triviality condition (NT) on the Lagrange multipliers  $(\lambda_0^N, \dots, \lambda_n^N, \eta_1^N, \dots, \eta_m^N)$ .

The existence and uniqueness of the maps  $(t, x, r) \mapsto w_x(t, r)$  solving the family of non-local Cauchy problems (3.20) can be obtained under hypotheses **(H)**, as a consequence of Banach fixed point Theorem in the spirit of ([14], Prop. 5). In this context, the Banach space under consideration is that of all maps  $w : [0, T] \times \Omega \rightarrow \mathbb{R}^d$  endowed with the norm

$$\|w(\cdot, \cdot)\|_* = \inf \left\{ M > 0 \text{ s.t. } \|w_x(\cdot, \cdot)\|_{C^0([0, T] \times \mathbb{R}^d, \mathbb{R}^d)} \leq M \text{ for } \mu^*(T)\text{-almost every } x \in \mathbb{R}^d \right\}.$$

By an application of Grönwall's Lemma to (3.20), it holds that  $(t, r) \in [0, T] \times \pi^2(\Omega) \mapsto \Psi_{(T,t)}^{x,N}(r)$  is bounded by a positive constant, uniformly with respect to  $x \in \text{supp}(\mu^*(T))$  and  $N \geq 1$ . This follows in particular from the uniform boundedness of the sequences of multipliers  $(\lambda_N^0)$  and  $(\zeta_N^*(\cdot))$ . Therefore, there exists a uniform constant  $R'_T > 0$  such that

$$\text{supp}(\nu_N^*(t)) \subset \overline{B_{2d}(0, R'_T)},$$

for all  $t \in [0, T]$ . This in turn implies that the right-hand side of (3.20) is uniformly bounded, so that the maps  $t \in [0, T] \mapsto \Psi_{(T,t)}^{x,N}(r)$  are Lipschitz, uniformly with respect to  $(x, r) \in \Omega$  and  $N \geq 1$ . By applying again Grönwall's Lemma to the difference  $|\Psi_{(T,t)}^{x,N}(r_2) - \Psi_{(T,t)}^{x,N}(r_1)|$  with  $r_1, r_2 \in \pi^2(\Omega)$ , we further obtain that  $(t, r) \in [0, T] \times \pi^2(\Omega) \mapsto \Psi_{(T,t)}^{x,N}(r)$  is also Lipschitz regular, uniformly with respect to  $x \in \text{supp}(\mu^*(T))$  and  $N \geq 1$ . It can be checked by leveraging Kantorovich duality in the spirit of ([14], Lem. 6) that this in turn yields the Lipschitz

regularity of  $t \in [0, T] \mapsto \sigma_{x,N}^*(t)$  uniformly with respect to  $x \in \text{supp}(\mu^*(T))$  and  $N \geq 1$ . An application of Proposition 2.7 combined with the uniform Lipschitz regularity of  $(t, x) \in [0, T] \times \pi^1(\Omega) \mapsto \Phi_{(T,t)}^{v,u^*}(x)$  to  $\nu_N^*(\cdot)$  provides the existence of a uniform constant  $L'_T > 0$  such that

$$W_1(\nu_N^*(t), \nu_N^*(s)) \leq L'_T |t - s|,$$

for any  $s, t \in [0, T]$ .

In order to prove that  $\nu_N^*(\cdot)$  is indeed a solution of (3.18), take  $\xi \in C_c^\infty(\mathbb{R}^{2d})$  and compute the time derivative

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^{2d}} \xi(x, r) d\nu_N^*(t)(x, r) &= \frac{d}{dt} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \xi \left( \Phi_{(T,t)}^{v,u^*}(x), r \right) d\sigma_{x,N}^*(t)(r) d\mu^*(T)(x) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\langle \nabla_x \xi \left( \Phi_{(T,t)}^{v,u^*}(x), r \right), v[\mu^*(t)] \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right) + u^* \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right) \right\rangle d\sigma_{x,N}^*(t)(r) d\mu^*(T)(x) \\ &\quad + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\langle \nabla_r \xi \left( \Phi_{(T,t)}^{v,u^*}(x), r \right), \lambda_0^N \nabla_\mu L(t, \mu^*(t), u^*(t)) \left( \Phi_{(T,t)}^{v,u^*}(x) \right) \right\rangle d\sigma_{x,N}^*(t)(r) d\mu^*(T)(x) \\ &\quad + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\langle \nabla_r \xi \left( \Phi_{(T,t)}^{v,u^*}(x), r \right), \nabla_\mu \mathcal{C}(t, \mu^*(t), \zeta_N^*(t), u^*(t)) \left( \Phi_{(T,t)}^{v,u^*}(x) \right) \right\rangle d\sigma_{x,N}^*(t)(r) d\mu^*(T)(x) \\ &\quad - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\langle \nabla_r \xi \left( \Phi_{(T,t)}^{v,u^*}(x), r \right), D_x v[\mu^*(t)] \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right)^\top r + D_x u^* \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right)^\top r \right\rangle d\sigma_{x,N}^*(t)(r) d\mu^*(T)(x) \\ &\quad - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\langle \nabla_r \xi \left( \Phi_{(T,t)}^{v,u^*}(x), r \right), \Pi^v[\nu_N^*(t)] \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right) \right\rangle d\sigma_{x,N}^*(t)(r) d\mu^*(T)(x) \end{aligned}$$

where we used the fact that by Fubini's Theorem

$$\begin{aligned} \Pi^v[\nu_N^*(t)] \left( t, \Phi_{(T,t)}^{v,u^*}(x) \right) &= \int_{\mathbb{R}^{2d}} \Pi_{(t,y)}^v \left( \Phi_{(T,t)}^{v,u^*}(x) \right)^\top p d\nu_N^*(t)(y, p) \\ &= \int_{\mathbb{R}^{2d}} \Pi_{(t, \Phi_{(T,t)}^{v,u^*}(y))}^v \left( \Phi_{(T,t)}^{v,u^*}(x) \right)^\top p d\nu_{T,N}^*(t)(y, p) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Pi_{(t, \Phi_{(T,t)}^{v,u^*}(y))}^v \left( \Phi_{(T,t)}^{v,u^*}(x) \right)^\top p d\sigma_{y,N}^*(t)(p) d\mu^*(T)(y) \\ &= \int_{\mathbb{R}^{2d}} \Pi_{(t, \Phi_{(T,t)}^{v,u^*}(y))}^v \left( \Phi_{(T,t)}^{v,u^*}(x) \right)^\top \Psi_{(T,t)}^{y,N}(p) d((\text{Id} \times (-\nabla_\mu \mathcal{L}_N(\mu^*(T))))_{\#} \mu^*(T))(y, p). \end{aligned}$$

This can in turn be reformulated into the more concise expression

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^{2d}} \xi(x, r) d\nu_N^*(t)(x, r) &= \int_{\mathbb{R}^{2d}} \langle \nabla_x \xi(x, r), v[\mu^*(t)](t, x) + u^*(t, x) \rangle d\nu_N^*(t)(x, r) \\ &\quad + \int_{\mathbb{R}^{2d}} \langle \nabla_r \xi(x, r), \lambda_0^N \nabla_\mu L(t, \mu^*(t), u^*(t))(x) \rangle d\nu_N^*(t)(x, r) \\ &\quad + \int_{\mathbb{R}^{2d}} \langle \nabla_r \xi(x, r), \nabla_\mu \mathcal{C}(t, \mu^*(t), \zeta_N^*(t), u^*(t))(x) \rangle d\nu_N^*(t)(x, r) \\ &\quad - \int_{\mathbb{R}^{2d}} \langle \nabla_r \xi(x, r), D_x v[\mu^*(t)](t, x)^\top r + D_x u^*(t, x)^\top r + \Pi^v[\nu_N^*(t)](t, x) \rangle d\nu_N^*(t)(x, r), \end{aligned}$$

which by (2.12) precisely corresponds to the fact that  $\nu_N^*(\cdot)$  is a solution of (3.18).  $\square$

**Remark 3.8** (Wasserstein and classical costates). In the finite-dimensional proof of the Gamkrelidze PMP, the optimal costates  $p_N^*(\cdot)$  are defined as the solutions of the backward equations

$$\dot{p}_N^*(t) = -\nabla_x \mathcal{H}_{\lambda_0^N}(t, x_N^*(t), p_N^*(t), \zeta_N^*(t), u^*(t)), \quad p_N^*(T) = -\nabla \mathcal{S}_N(x^*(T)), \quad (3.21)$$

where  $\mathcal{H}_{\lambda_0^N}(\cdot, \cdot, \cdot, \cdot, \cdot)$  and  $\mathcal{S}_N(\cdot)$  are the counterparts of  $\mathcal{H}_{\lambda_0^N}(\cdot, \cdot, \cdot, \cdot)$  and  $\mathcal{S}_N(\cdot)$  associated with the finite-dimensional optimal control problem

$$\left\{ \begin{array}{l} \min_{u \in \mathcal{U}} \left[ \int_0^T L(t, x(t), u(t)) dt + \varphi(x(T)) \right], \\ \text{s.t.} \left\{ \begin{array}{l} \dot{x}(t) = v(t, x(t)) + u(t), \\ x(0) = x^0 \in \mathbb{R}^d, \end{array} \right. \\ \text{and} \left\{ \begin{array}{l} \Psi^I(x(T)) \leq 0, \quad \Psi^E(x(T)) = 0, \\ \Lambda(t, x(t)) \leq 0 \text{ for all } t \in [0, T], \end{array} \right. \end{array} \right.$$

and defined by

$$\left\{ \begin{array}{l} \mathcal{H}_{\lambda_0^N}(t, x, p, \zeta, \omega) = \langle p, v(t, x) + \omega \rangle - \lambda_0^N L(t, x, \omega) - \mathcal{C}(t, x, \zeta, \omega), \\ \mathcal{S}_N(x) = \lambda_0^N \varphi(x) + \sum_{i=1}^n \lambda_i^N \Psi_i^I(x) + \sum_{j=1}^m \eta_j^N \Psi_j^E(x). \end{array} \right.$$

In our statement of the PMP, one should think of  $\pi_{\#}^2 \nu^*(\cdot)$  as being concentrated on the characteristic curves of the backward costate dynamics. Indeed in Lemma 3.7, the curves  $\sigma_{x,N}^*(\cdot)$  are concentrated on the unique characteristic of the linearized backward non-local dynamics (3.20) starting from  $(-\nabla_{\mu} \mathcal{S}_N(\mu^*(T))(x))$  for  $\mu^*(T)$ -almost every  $x \in \mathbb{R}^d$ . The second marginal of the curve  $\nu_{T,N}^*(\cdot) = \int \sigma_{x,N}^*(\cdot) d\mu^*(T)(x)$  can then be seen as a Lagrangian superposition of integral curves of (3.21) depending on the starting point of the curve in  $\text{supp}(\mu^*(T))$ .

Now that we have built a suitable notion of solution for (3.18), let us prove that  $\nu_N^*(\cdot)$  is such that the PMP holds with a relaxed maximization condition formulated over the collection of needle parameters  $\{(\omega_k, \tau_k)\}_{k=1}^N \subset U_D \times \mathcal{A}$ .

**Lemma 3.9** (A partial Pontryagin Maximum Principle). *For any  $N \geq 1$ , the curve of measures  $\nu_N^*(\cdot)$  introduced in Lemma 3.7 is a solution of the Hamiltonian flow (3.1) associated to the Hamiltonian  $\mathcal{H}_{\lambda_0^N}(\cdot, \cdot, \cdot, \cdot)$  and to the Lagrange multipliers  $(\lambda_0^N, \dots, \lambda_n^N, \dots, \eta_1^N, \dots, \eta_m^N, \varpi_1^N, \dots, \varpi_r^N) \in \{0, 1\} \times \mathbb{R}_+^n \times \mathbb{R}^m \times \mathcal{M}_+([0, T])^r$ . Moreover, the relaxed maximization condition*

$$\begin{aligned} & \mathcal{H}_{\lambda_0^N}(\tau_k, \nu_N^*(\tau_k), \zeta_N^*(\tau_k), \omega_k) - \sum_{l=1}^r \varpi_l^N(\{\tau_k\}) \int_{\mathbb{R}^d} \langle \nabla_{\mu} \Lambda_l(\tau_k, \mu^*(\tau_k))(x), \omega_k \rangle d\mu^*(\tau_k)(x) \\ & \leq \mathcal{H}_{\lambda_0^N}(\tau_k, \nu_N^*(\tau_k), \zeta_N^*(\tau_k), u^*(\tau_k)) - \sum_{l=1}^r \varpi_l^N(\{\tau_k\}) \int_{\mathbb{R}^d} \langle \nabla_{\mu} \Lambda_l(\tau_k, \mu^*(\tau_k))(x), u^*(\tau_k, x) \rangle d\mu^*(\tau_k)(x) \end{aligned} \quad (3.22)$$

holds for all  $\{(\omega_k, \tau_k)\}_{k=1}^N \subset U_D \times \mathcal{A}$ .

*Proof.* Using the expression (3.2) of the augmented infinite-dimensional Hamiltonian along with the definition (2.6) of Wasserstein gradients and the results given in Propositions A.2 and B.1, one can check that

$$\mathcal{V}_N^*[\nu_N^*(t)](t, x, r) = \mathbb{J}_{2d} \nabla_\nu \mathcal{H}_{\lambda_0^N}(t, \nu_N^*(t), \zeta_N^*(t), u^*(t))(x, r),$$

for  $\mathcal{L}^1$ -almost every  $t \in [0, T]$  and every  $(x, r) \in \overline{B_{2d}(0, R'_T)}$ .

For  $k \in \{1, \dots, N\}$ , we introduce the collection of maps  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  defined for  $\mathcal{L}^1$ -almost all  $t \in [\tau_k, T]$  by

$$\begin{aligned} \mathcal{K}_{\omega_k, \tau_k}^N(t) &= \int_{\mathbb{R}^{2d}} \langle r, \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\nu_N^*(t)(x, r) + \lambda_0^N \left( L(\tau_k, \mu^*(\tau_k), \omega_k) - L(\tau_k, \mu^*(\tau_k), u^*(\tau_k)) \right) \\ &\quad - \int_{\tau_k}^t \int_{\mathbb{R}^d} \left\langle \lambda_0^N \nabla_\mu L(s, \mu^*(s), u^*(s))(x), \mathcal{F}_s^{\omega_k, \tau_k} \circ \Phi_{(s, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(s)(x) ds \\ &\quad - \sum_{l=1}^r \int_{\tau_k}^t \int_{\mathbb{R}^d} \left\langle \nabla_\mu \Lambda_l(s, \mu^*(s))(x), \mathcal{F}_s^{\omega_k, \tau_k} \circ \Phi_{(s, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(s)(x) d\varpi_l^N(s) \\ &\quad - \sum_{l=1}^r \zeta_{l, N}^*(t) \int_{\mathbb{R}^d} \left\langle \nabla_\mu \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x). \end{aligned} \tag{3.23}$$

$k \in \{1, \dots, N\}$ . By construction, the maps  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  satisfy

$$\mathcal{K}_{\omega_k, \tau_k}^N(T) \leq 0, \tag{3.24}$$

since it can be checked that the evaluation of  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  at  $T$  coincides with the left-hand side of (3.17), which has been shown to be non-positive for all  $k \in \{1, \dots, N\}$ . Moreover, the evaluation of the maps  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  at  $\tau_k$  can be written explicitly as

$$\begin{aligned} \mathcal{K}_{\omega_k, \tau_k}^N(\tau_k) &= \mathcal{H}_{\lambda_0^N}(\tau_k, \nu_N^*(\tau_k), \zeta_N^*(\tau_k), \omega_k) - \mathcal{H}_{\lambda_0^N}(\tau_k, \nu_N^*(\tau_k), \zeta_N^*(\tau_k), u^*(\tau_k)) \\ &\quad - \sum_{l=1}^r \varpi_l^N(\{\tau_k\}) \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(\tau_k, \mu^*(\tau_k)), \omega_k(x) - u^*(\tau_k, x) \rangle d\mu^*(\tau_k)(x). \end{aligned} \tag{3.25}$$

We now aim at showing that the maps  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  are constant over  $[\tau_k, T]$ . By definition, these functions are in  $BV([0, T], \mathbb{R})$  and therefore admit a distributional derivative in the form of finite Borel regular measures (see *e.g.* [7], Chap. 3). A simple computation of the time derivatives of the last two terms in (3.23) shows that the non-absolutely continuous parts of the derivatives of the maps  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  cancel each other out, since the weak derivatives of the maps  $\zeta_{N, l}^*(\cdot)$  are such that  $d\zeta_{N, l}^* = -\varpi_l^N$ . Hence, the maps  $\mathcal{K}_{\omega_k, \tau_k}^N(\cdot)$  are absolutely continuous and therefore differentiable  $\mathcal{L}^1$ -almost everywhere. One can then compute their derivative at  $\mathcal{L}^1$ -almost every  $t \in [\tau_k, T]$  as follows.

$$\begin{aligned} \frac{d}{dt} \mathcal{K}_{\omega_k, \tau_k}^N(t) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \langle r, \partial_t \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \rangle d\sigma_{x, N}^*(t)(r) d\mu^*(T)(x) \\ &\quad + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \langle \mathcal{V}_N^*[\nu_N^*(t)](t, x, r), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \rangle d\sigma_{x, N}^*(t)(r) d\mu^*(T)(x) \\ &\quad - \int_{\mathbb{R}^d} \left\langle \lambda_0^N \nabla_\mu L(t, \mu^*(t), u^*(t)) \left( \Phi_{(T, t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\ &\quad - \sum_{l=1}^r \zeta_{l, N}^*(t) \frac{d}{dt} \left[ \int_{\mathbb{R}^d} \left\langle \nabla_\mu \Lambda_l(t, \mu^*(t)) \left( \Phi_{(T, t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \right] \end{aligned} \tag{3.26}$$

The time-derivatives of the summands of the last term can be computed as follows using Proposition 2.14 and the geometric structure (2.15) of solutions of (2.14) associated with the non-local velocity field  $v[\mu^*(t)](t, \cdot) + u^*(t, \cdot)$ .

$$\begin{aligned}
& \frac{d}{dt} \left[ \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Lambda_l(t, \mu^*(t)) \left( \Phi_{(T,t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \right] \\
&= \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \Lambda_l(t, \mu^*(t)) \left( \Phi_{(T,t)}^{v, u^*}(x) \right), \partial_t \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\
&+ \int_{\mathbb{R}^d} \left\langle \partial_t \nabla_{\mu} \Lambda_l(t, \mu^*(t)) \left( \Phi_{(T,t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\
&+ \int_{\mathbb{R}^d} \left\langle D_x \nabla_{\mu} \Lambda_l(t, \mu^*(t)) \left( \Phi_{(T,t)}^{v, u^*}(x) \right) v[\mu^*(t)] \left( t, \Phi_{(T,t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\
&+ \int_{\mathbb{R}^d} \left\langle D_x \nabla_{\mu} \Lambda_l(t, \mu^*(t)) \left( \Phi_{(T,t)}^{v, u^*}(x) \right) u^* \left( t, \Phi_{(T,t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\
&+ \int_{\mathbb{R}^d} \left\langle \int_{\mathbb{R}^d} \Pi_{\left( t, \Phi_{(T,t)}^{v, u^*}(x) \right)}^{\nabla_{\mu} \Lambda_l} \left( \Phi_{(T,t)}^{v, u^*}(y) \right) v[\mu^*(t)] \left( t, \Phi_{(T,t)}^{v, u^*}(y) \right) d\mu^*(T)(y), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\
&+ \int_{\mathbb{R}^d} \left\langle \int_{\mathbb{R}^d} \Pi_{\left( t, \Phi_{(T,t)}^{v, u^*}(x) \right)}^{\nabla_{\mu} \Lambda_l} \left( \Phi_{(T,t)}^{v, u^*}(y) \right) u^* \left( t, \Phi_{(T,t)}^{v, u^*}(y) \right) d\mu^*(T)(y), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x) \\
&= \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \mathcal{G}_l(t, \mu^*(t), u^*(t)) \left( \Phi_{(T,t)}^{v, u^*}(x) \right), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(T)(x)
\end{aligned}$$

by applying Fubini's Theorem and identifying the analytical expressions of the Wasserstein gradients of the summands  $\mathcal{G}_l(t, \mu^*(t), u^*(t))$  of  $\mathcal{C}(t, \mu^*(t), \zeta^*(t), u^*(t))$  derived in Proposition B.1. Plugging this expression into (3.26) along with the characterization (3.10) of  $\partial_t \mathcal{F}_t^{\omega_k, \tau_k}(\cdot)$  derived in Lemma 3.5, we obtain that

$$\begin{aligned}
\frac{d}{dt} \mathcal{K}_{\omega_k, \tau_k}^N(t) &= \int_{\mathbb{R}^{2d}} \left\langle D_x u^*(t, x)^\top r + D_x v[\mu^*(t)](t, x)^\top r + \Pi^v[\nu_N^*(t)](t, x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\nu_N^*(t)(x, r) \\
&+ \int_{\mathbb{R}^{2d}} \left\langle \mathcal{V}_N^*[\nu_N^*(t)](t, x, r), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\nu_N^*(t)(x) \\
&- \int_{\mathbb{R}^d} \left\langle \lambda_0^N \nabla_{\mu} L(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x) \\
&- \sum_{l=1}^r \zeta_{l, N}^*(t) \int_{\mathbb{R}^d} \left\langle \nabla_{\mu} \mathcal{G}_l(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x).
\end{aligned}$$

In the first line we used the fact that

$$\begin{aligned}
& \int_{\mathbb{R}^{2d}} \left\langle r, \int_{\mathbb{R}^d} \Pi_{(t, x)}^v(y) \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(y) d\mu^*(t)(y) \right\rangle d\nu_N^*(t)(x, r) \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^{2d}} \left\langle \Pi_{(t, y)}^v(x)^\top p, \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\nu_N^*(t)(x, r) d\mu^*(t)(y) \\
&= \int_{\mathbb{R}^d} \left\langle \Pi^v[\nu_N^*(t)](t, x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \right\rangle d\mu^*(t)(x).
\end{aligned}$$

as a consequence of Fubini's Theorem, where  $\Pi^v[\nu_N^*(\cdot)](\cdot, \cdot)$  is defined as in (3.19). Recalling the definition of the vector field  $\mathcal{V}_N^*[\cdot](\cdot, \cdot, \cdot)$  given in (3.18), we therefore observe that  $\frac{d}{dt} \mathcal{K}_{\omega_k, \tau_k}^N(t) = 0$  for  $\mathcal{L}^1$ -almost every  $t \in [\tau_k, T]$ , so that it is constant over this time interval. Merging this fact with (3.24) and (3.25) yields (3.22) and concludes the proof of our claim.  $\square$

#### Step 4: Limiting procedure.

##### *The PMP for absolutely continuous state constraints multipliers*

In Step 3, we have built for any  $N \geq 1$  a suitable state-costate curve  $\nu_N^*(\cdot)$  solution of the Hamiltonian system (3.1), and such that the relaxed Pontryagin maximization condition (3.22) holds on an  $N$ -dimensional subset of needle parameters. The last step in the proof of Theorem 3.1 is to take the limit as  $N$  goes to infinity of the previous optimality conditions in order to recover the PMP formulated on the whole set of needle parameters.

By the non-triviality condition (NT), the sequence of Lagrange multipliers  $((\lambda_0^N, \dots, \lambda_n^N, \eta_1^N, \dots, \eta_m^N)) \subset \{0, 1\} \times \mathbb{R}_+^n \times \mathbb{R}^m$  is bounded uniformly with respect to  $N$ . Hence, we can extract a subsequence of multipliers such that  $(\lambda_0^N, \dots, \lambda_n^N, \eta_1^N, \dots, \eta_m^N) \rightarrow (\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m)$  as  $N \rightarrow +\infty$ . A straightforward passage to the limit shows that these limit multipliers satisfy the complementary slackness condition

$$\lambda_i \Psi_i^I(\mu^*(T)) = 0 \quad \text{for all } i \in \{1, \dots, n\}.$$

Similarly, the sequence of measures  $(\varpi_1^N, \dots, \varpi_r^N) \subset \mathcal{M}_+([0, T])^r$  is uniformly bounded with respect to the total variation norm as a consequence of Lemma 3.6. By Banach-Alaoglu's Theorem (see *e.g.* [16], Thm. 3.16), it therefore admits a weakly-\* converging subsequence to some  $(\varpi_1, \dots, \varpi_r) \in \mathcal{M}_+([0, T])^r$ . Moreover for any  $l \in \{1, \dots, r\}$ , the measures  $(\varpi_l^N)$  are equi-supported in the sets  $\{t \in [0, T] \text{ s.t. } \Lambda_l(t, \mu^*(t)) = 0\}$ . It can then be shown by standard convergence properties on the supports of sequences of measures (see *e.g.* [9], Prop. 5.1.8) that the limit measures  $(\varpi_1, \dots, \varpi_r)$  are supported on these sets as well. Therefore, it holds that

$$\text{supp}(\varpi_l) \subseteq \left\{ t \in [0, T] \text{ s.t. } \Lambda_l(t, \mu^*(t)) = 0 \right\},$$

for any  $l \in \{1, \dots, r\}$ . Furthermore, if all the scalar Lagrange multipliers  $(\lambda_0^N, \dots, \lambda_n^N, \eta_1^N, \dots, \eta_m^N)$  vanish for large  $N$ , it follows from the non-triviality condition (NT) that  $\|\varpi_l^N\|_{TV} > 0$  so that  $\varpi_l \neq 0$  in the limit, at least for some  $l \in \{1, \dots, r\}$ . Hence, we recover the non-degeneracy condition (3.5), *i.e.*

$$(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1, \dots, \varpi_r) \neq 0.$$

In Lemma 3.7, we have shown that the curves of measures  $\nu_N^*(\cdot)$  are uniformly equi-compactly supported and equi-Lipschitz. Hence,  $(\nu_N^*(\cdot))$  admits converging subsequences in the  $C^0$ -topology by Ascoli-Arzelà Theorem (see *e.g.* [49], Thm. 11.28).

We now prove that there exists an accumulation point  $\nu^*(\cdot)$  of  $(\nu_N^*(\cdot))$  which solves the system of equations (3.1) associated with the limit multipliers  $(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1, \dots, \varpi_r)$ . To this end, we start by making an extra simplifying assumption on the state constraints multipliers.

**(H7)** : The measures  $(\varpi_1, \dots, \varpi_r)$  are absolutely continuous with respect to  $\mathcal{L}^1$  on  $[0, T]$ .

We shall see in the sequel how this extra assumption can be lifted at the price of an extra approximation argument by absolutely continuous measures. Let  $\nu^*(\cdot) \in \text{Lip}([0, T], \mathcal{P}(\overline{B_{2d}(0, R_T)}))$  be an accumulation point of  $(\nu_N^*(\cdot))$  along a suitable subsequence. As a direct consequence of the convergence of the scalar Lagrange multipliers, one recovers the uniform convergence of the final gradient map

$$\nabla_{\mu} \mathcal{S}_N(\mu^*(T))(\cdot) \xrightarrow[N \rightarrow +\infty]{C^0} \nabla_{\mu} \mathcal{S}(\mu^*(T))(\cdot).$$

This implies by standard convergence results for pushforwards of measures (see *e.g.* [9], Lem. 5.2.1) that  $\nu^*(\cdot)$  satisfies the boundary condition

$$\pi_{\#}^2 \nu^*(T) = (-\nabla_{\mu} \mathcal{S}(\mu^*(T)))_{\#} \mu^*(T).$$

Moreover, the weak- $*$  convergence of  $(\varpi_1^N, \dots, \varpi_r^N)$  towards  $(\varpi_1, \dots, \varpi_r)$  along with **(H7)** implies by Proposition 2.1 that

$$\zeta_{l,N}^*(t) = \mathbb{1}_{[0,T]}(t) \int_t^T d\varpi_l^N(s) \xrightarrow{N \rightarrow +\infty} \mathbb{1}_{[0,T]}(t) \int_t^T d\varpi_l(s) = \zeta_l^*(t)$$

for all times  $t \in [0, T]$ . By definition (2.11) of distributional solutions to transport equations, the fact that  $\nu_N^*(\cdot)$  is a solution of (3.18) can be written as

$$\int_0^T \int_{\mathbb{R}^{2d}} \left( \partial_t \xi(t, x, r) + \left\langle \nabla_{(x,r)} \xi(t, x, r), \mathbb{J}_{2d} \nabla_{\nu} \mathcal{H}_{\lambda_0^N}(t, \nu_N^*(t), \zeta_N^*(t), u^*(t))(x, r) \right\rangle \right) d\nu_N^*(t)(x, r) dt = 0 \quad (3.27)$$

for any  $\xi \in C_c^\infty([0, T] \times \mathbb{R}^{2d})$ . Since all the functionals involved in the definition of the Wasserstein gradient of the augmented infinite-dimensional Hamiltonian are continuous and bounded, we have that

$$\nabla_{\nu} \mathcal{H}_{\lambda_0^N}(t, \nu_N^*(t), \zeta_N^*(t), u^*(t))(\cdot, \cdot) \xrightarrow[N \rightarrow +\infty]{C^0} \nabla_{\nu} \mathcal{H}_{\lambda_0}(t, \nu^*(t), \zeta^*(t), u^*(t))(\cdot, \cdot)$$

uniformly with respect to  $t \in [0, T]$ , as a by-product of the convergence of the Lagrange multipliers. By using this fact along with the uniform equi-compactness of the supports of  $(\nu_N^*(\cdot))$ , we can take the limit as  $N \rightarrow +\infty$  in (3.27) and apply Lebesgue's Dominated Convergence Theorem to recover that

$$\int_0^T \int_{\mathbb{R}^{2d}} \left( \partial_t \xi(t, x, r) + \left\langle \nabla_{(x,r)} \xi(t, x, r), \nabla_{\nu} \mathcal{H}_{\lambda_0}(t, \nu^*(t), \zeta^*(t), u^*(t))(x, r) \right\rangle \right) d\nu^*(t)(x, r) dt = 0$$

for any  $\xi \in C_c^\infty([0, T] \times \mathbb{R}^{2d})$ . Hence, the accumulation point  $\nu^*(\cdot)$  of  $(\nu_N^*(\cdot))$  in the  $C^0$ -topology is a solution of the Hamiltonian flow (3.18) associated with the limit multipliers  $(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1, \dots, \varpi_r)$ .

In order to complete our proof of Theorem 3.1, there remains to show that the limit curve  $\nu^*(\cdot)$  is such that the maximization condition (3.6) holds for  $\mathcal{L}^1$ -almost every  $t \in [0, T]$ . We know that for any  $(\omega_k, \tau_k) \in U_D \times \mathcal{A}$ , the modified maximization condition (3.22) holds. By the preliminary assumption **(H7)** that the limit measures  $(\varpi_1, \dots, \varpi_r)$  are absolutely continuous with respect to  $\mathcal{L}^1$ , we can apply Proposition 2.1 to recover that

$$\varpi_l^N(\{\tau_k\}) \xrightarrow{N \rightarrow +\infty} \varpi_l(\{\tau_k\}) = 0,$$

for any  $l \in \{1, \dots, r\}$ . Since the infinite-dimensional Hamiltonian is continuous with respect to its second argument in the  $W_1$ -metric and linear with respect to its third argument, it holds that

$$\mathcal{H}_{\lambda_0^N}(\tau_k, \nu_N^*(\tau_k), \zeta_N^*(\tau_k), \omega_k) \xrightarrow{N \rightarrow +\infty} \mathcal{H}_{\lambda_0}(\tau_k, \nu^*(\tau_k), \zeta^*(\tau_k), \omega_k)$$

and

$$\mathcal{H}_{\lambda_0^N}(\tau_k, \nu_N^*(\tau_k), \zeta_N^*(\tau_k), u^*(\tau_k)) \xrightarrow{N \rightarrow +\infty} \mathcal{H}_{\lambda_0}(\tau_k, \nu^*(\tau_k), \zeta^*(\tau_k), u^*(\tau_k))$$

uniformly with respect to  $k \leq N$ . We can therefore pass to the limit as  $N \rightarrow +\infty$  in the partial maximization condition (3.22) to obtain that

$$\mathcal{H}_{\lambda_0}(\tau_k, \nu^*(\tau_k), \zeta^*(\tau_k), \omega_k) \leq \mathcal{H}_{\lambda_0}(\tau_k, \nu^*(\tau_k), \zeta^*(\tau_k), u^*(\tau_k)) \quad (3.28)$$

for any  $(\omega_k, \tau_k) \in U_D \times \mathcal{A}$ .

Given an arbitrary pair  $(\omega, \tau) \in U \times \mathcal{M}$ , it is possible to choose a sequence of elements  $\{(\omega_k, \tau_k)\}_k \subset U_D \times \mathcal{A}$  such that

$$\tau_k \xrightarrow[k \rightarrow +\infty]{} \tau, \quad \omega_k \xrightarrow[k \rightarrow +\infty]{C^0} \omega,$$

and

$$\|u^*(\tau, \cdot) - u^*(\tau_k, \cdot)\|_{C^0(\overline{B(0, R_T)}, \mathbb{R}^d)} \xrightarrow[k \rightarrow +\infty]{} 0, \quad \|L(\tau, \mu^*(\tau), \cdot) - L(\tau_k, \mu^*(\tau_k), \cdot)\|_{C^0(U, \mathbb{R})} \xrightarrow[k \rightarrow +\infty]{} 0. \quad (3.29)$$

Remark first that under assumption **(H7)**, the maps  $t \rightarrow \zeta_l^*(t)$  are continuous on  $[0, T]$ . By (3.29) along with the continuity of the augmented infinite-dimensional Hamiltonian in the  $C^0$ -norm topology with respect to its fourth argument, we can pass to the limit as  $k \rightarrow +\infty$  in (3.28). This yields the Pontryagin Maximization condition

$$\mathcal{H}_{\lambda_0}(\tau, \nu^*(\tau), \zeta^*(\tau), \omega) \leq \mathcal{H}_{\lambda_0}(\tau, \nu^*(\tau), \zeta^*(\tau), u^*(\tau))$$

for any pair  $(\omega, \tau) \in U \times \mathcal{M}$ .

### *Lifting the absolute continuity hypothesis (H7)*

In order to recover the full statement of Theorem 3.1, we now show how to relax the absolute continuity assumption **(H7)** made on the state constraints multipliers. For a given small parameter  $\epsilon > 0$ , we consider a sequence of mollifiers  $\rho_\epsilon : t \in [0, T] \mapsto \epsilon^{-1} \rho(t/\epsilon)$  where  $\rho \in C_c^\infty([0, T])$  is such that  $\int_0^T \rho(t) dt = 1$ . Given  $N \geq 1$ , we define the mollified measure  $(\varpi_1^{N, \epsilon}, \dots, \varpi_r^{N, \epsilon})$  by

$$\varpi_l^{N, \epsilon} = (\rho_\epsilon * \varpi_l^N)(t) \cdot \mathcal{L}^1,$$

where for any  $l \in \{1, \dots, r\}$ , the convolution maps are defined by  $\rho_\epsilon * \varpi_l^N : t \in [0, T] \mapsto \int_0^T \rho_\epsilon(t-s) d\varpi_l^N(s)$ . Using the fact that the functions

$$t \in [\tau_k, T] \mapsto \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x)$$

are Lipschitz and bounded as a by-product of **(H6)** and Lemma 3.5, one can assert using the definition of the dual norm in the Banach space  $\mathcal{M}_+([0, T])$  that

$$\begin{aligned} & - \int_{\tau_k}^T \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x) d\varpi_l^N(t) \\ & \geq - \int_{\tau_k}^T \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x) d\varpi_l^{N, \epsilon}(t) - C\epsilon \end{aligned}$$

for some uniform constant  $C > 0$ . This allows us to rewrite the optimality conditions (3.17) derived at time  $T$  as

$$\begin{aligned} & \int_{\mathbb{R}^d} \langle -\nabla_{\mu} \mathcal{S}_N(\mu^*(T))(x), \mathcal{F}_T^{\omega_k, \tau_k} \circ \Phi_{(T, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(T)(x) - \lambda_N^0 \left( L(\tau_k, \mu^*(\tau_k), u^*(\tau_k)) - L(\tau_k, \mu^*(\tau_k), \omega_k) \right) \\ & - \int_{\tau_k}^T \int_{\mathbb{R}^d} \langle \lambda_N^0 \nabla_{\mu} L(t, \mu^*(t), u^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x) dt \\ & - \sum_{l=1}^r \int_{\tau_k}^T \int_{\mathbb{R}^d} \langle \nabla_{\mu} \Lambda_l(t, \mu^*(t))(x), \mathcal{F}_t^{\omega_k, \tau_k} \circ \Phi_{(t, \tau_k)}^{v, u^*}(x) \rangle d\mu^*(t)(x) d\varpi_l^{N, \epsilon}(t) \leq C\epsilon. \end{aligned} \tag{3.30}$$

By defining the family of measure curves  $(\nu_{N, \epsilon}^*(\cdot))$  as in Lemma 3.7, we can prove that the corresponding maps  $\mathcal{K}_{\omega_k, \tau_k}^{N, \epsilon}(\cdot)$  defined as in (3.26) are constant over  $[\tau_k, T]$  and that the partial maximization conditions

$$\mathcal{H}_{\lambda_0^N}(\tau_k, \nu_{N, \epsilon}^*(\tau_k), \zeta_{N, \epsilon}^*(\tau_k), \omega_k) \leq \mathcal{H}_{\lambda_0^N}(\tau_k, \nu_{N, \epsilon}^*(\tau_k), \zeta_{N, \epsilon}^*(\tau_k), u^*(\tau_k)) + C\epsilon$$

hold for any  $\epsilon > 0$ . Performing again the limiting arguments of Step 4 as  $N \rightarrow +\infty$  and remarking that

$$\varpi_l^{N, \epsilon} \xrightarrow[N \rightarrow +\infty]{*} \varpi_l^{\epsilon} = (\rho_{\epsilon} * \varpi_l) \cdot \mathcal{L}^1,$$

we recover the statement of the PMP with a measure curve  $\nu_{\epsilon}^*(\cdot)$  associated to the Lagrange multipliers  $(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1^{\epsilon}, \dots, \varpi_r^{\epsilon})$ . This limit curve is such that the relaxed maximization condition

$$\mathcal{H}_{\lambda_0}(\tau, \nu_{\epsilon}^*(\tau), \zeta_{\epsilon}^*(\tau), \omega) \leq \mathcal{H}_{\lambda_0}(\tau, \nu_{\epsilon}^*(\tau), \zeta_{\epsilon}^*(\tau), u^*(\tau)) + C\epsilon \tag{3.31}$$

holds for any  $(\omega, \tau) \in U \times \mathcal{M}$ . There now remains to perform a last limiting argument as  $\epsilon \downarrow 0$  to recover the full maximum principle.

By Lebesgue's Decomposition Theorem for finite Borel measures on the real line (see *e.g.* [7], Rem. 3.32, Cor. 3.33), the sets  $\{t \in [0, T] \text{ s.t. } \varpi_l(\{t\}) > 0\}$  are at most countable for all  $l \in \{1, \dots, r\}$ . Therefore, the set  $\mathcal{M}^{\circ} = \mathcal{M} \setminus \bigcup_{l=1}^r \{t \in [0, T] \text{ s.t. } \varpi_l(\{t\}) > 0\}$  has full Lebesgue measure in  $[0, T]$ , and by Proposition 2.1 it is such that

$$\zeta_{l, \epsilon}^*(\tau) = \int_{\tau}^T d\varpi_l^{\epsilon}(s) \xrightarrow[\epsilon \downarrow 0]{} \int_{\tau}^T d\varpi_l(s) = \zeta_l^*(\tau).$$

for any  $\tau \in \mathcal{M}^{\circ}$ . By Ascoli-Arzelà Theorem, it holds that  $\nu_{\epsilon}^*(\cdot) \rightarrow \nu^*(\cdot)$  in the  $C^0$ -norm topology and it can be checked that this limit curve solves the forward-backward system of continuity equations (3.1) associated with the multipliers  $(\lambda_0, \dots, \lambda_n, \eta_1, \dots, \eta_m, \varpi_1, \dots, \varpi_r)$ . Moreover, letting  $\epsilon \downarrow 0$  in (3.31) implies that the Pontryagin maximization condition (3.6) holds on the restricted subset  $U \times \mathcal{M}^{\circ}$ .

## APPENDIX A. EXAMPLES OF FUNCTIONALS SATISFYING HYPOTHESES **(H)**

In this appendix, we show that the rather long list of hypotheses **(H)** is not too restrictive and that a good score of relevant functionals for applications fit into the framework of Theorem 3.1. This list of example is partly borrowed from our previous work [14].

**Proposition A.1** (Example of non-local velocity field). *Let  $(t, x, y) \mapsto H(t, x, y) \in \mathbb{R}^d$  be measurable with respect to  $t \in [0, T]$ , sublinear, Lipschitz and  $C^1$ -with respect to  $(x, y) \in \mathbb{R}^{2d}$ . Then, the map  $\mu \in \mathcal{P}_c(\mathbb{R}^d) \mapsto$*

$v[\mu](\cdot, \cdot)$  defined by

$$v[\mu](t, x) = \int_{\mathbb{R}^d} H(t, x, y) d\mu(y)$$

for  $\mathcal{L}^1$ -almost every  $t \in [0, T]$  and any  $x \in \mathbb{R}^d$  satisfies the hypotheses **(H2)** and **(H3)** of Theorem 3.1. Moreover, its first-order variations  $D_x v[\mu](t, x)$  and  $\int_{\mathbb{R}^d} \Pi_{(t,x)}^v(y) d\mu(y)$  are given by

$$D_x v[\mu](t, x) = \int_{\mathbb{R}^d} D_x H(t, x, y) d\mu(y) \quad , \quad \int_{\mathbb{R}^d} \Pi_{(t,x)}^v(y) d\mu(y) = \int_{\mathbb{R}^d} D_y H(t, x, y) d\mu(y).$$

**Proposition A.2** (Example of cost and constraint functions). *Let  $n \geq 1$  and  $W \in C^1(\mathbb{R}^{nd}, \mathbb{R})$ . Then, the functional*

$$\varphi : \mu \in \mathcal{P}_c(\mathbb{R}^d) \mapsto \int_{\mathbb{R}^{nd}} W(x_1, \dots, x_n) d\mu^{\otimes n}(x_1, \dots, x_n)$$

with  $\mu^{\otimes n} = \mu \times \dots \times \mu$  satisfies **(H4)** of Theorem 3.1 and its Wasserstein gradient at some  $\mu \in \mathcal{P}(K)$  is given by

$$\nabla_{\mu} \varphi(\mu)(x_1, \dots, x_n) = \sum_{j=1}^n \nabla_{x_j} W(x_1, \dots, x_n).$$

Let  $m \in C^1(\mathbb{R}^d, \mathbb{R}^n)$  and  $(t, x, v, r) \mapsto l(t, x, v, r) \in \mathbb{R}$  be  $\mathcal{L}^1$ -measurable with respect to  $t \in [0, T]$  and  $C^1$ -smooth with respect to  $(x, v, r) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^n$ . Then, the functional

$$L : (t, \mu, \omega) \in [0, T] \times \mathcal{P}(K) \times U \mapsto \int_{\mathbb{R}^d} l(t, x, \omega(x), \int m d\mu) d\mu(x),$$

satisfies the hypotheses **(H5)** of Theorem 3.1 and its Wasserstein gradient is given by

$$\begin{aligned} \nabla_{\mu} L(t, \mu, \omega)(x) &= \nabla_x l(t, x, \omega(x), \int m d\mu) + D_x \omega(x)^{\top} \nabla_v l(t, x, \omega(x), \int m d\mu) \\ &\quad + D_x m(x)^{\top} \int_{\mathbb{R}^d} \nabla_r l(t, y, \omega(y), \int m d\mu) d\mu(y). \end{aligned}$$

**Proposition A.3** (Example of state constraints). *Let  $m \in C^2(\mathbb{R}^d, \mathbb{R}^k)$  and  $\lambda \in C^2([0, T] \times \mathbb{R}^d \times \mathbb{R}^k, \mathbb{R}^r)$ . Then for any  $l \in \{1, \dots, r\}$ , the functionals*

$$\Lambda_l(t, \mu) \in [0, T] \times \mathcal{P}(K) \mapsto \int_{\mathbb{R}^d} \lambda_l(t, x, \int m d\mu) d\mu(x)$$

satisfy the hypotheses **(H6)** of Theorem 3.1 and their derivatives can be computed using Proposition A.2.

**Remark A.4.** Particular cases of functionals which are of great interest for applications are for instance the variance functional  $\mu \mapsto \frac{1}{2} \int_{\mathbb{R}^d} |x - \bar{\mu}|^2 d\mu(x)$  where  $\bar{\mu} = \int y d\mu(y)$  or the target-support map to a closed set  $S \subset \mathbb{R}^d$   $\mu \mapsto \frac{1}{2} \int_{\mathbb{R}^d} d_S(x)^2 d\mu(x)$ .

APPENDIX B. WASSERSTEIN DIFFERENTIAL OF THE RUNNING  
 CONSTRAINT PENALIZATION

In this section, we give the analytical expression of the Wasserstein differential of the running constraint penalization map  $(t, \mu, \zeta, \omega) \mapsto \mathcal{C}(t, \mu, \zeta, \omega)$  defined in (3.3).

**Proposition B.1** (Wasserstein differential of the state constraints penalization map). *Let  $K \subset \mathbb{R}^d$  be a compact set and  $\omega \in U$ . Under hypotheses (H6), the map*

$$\begin{aligned} \mu \in \mathcal{P}(K) \mapsto \mathcal{C}(t, \mu, \zeta^*(t), \omega) &= \sum_{l=1}^r \zeta_l^*(t) \left( \partial_t \Lambda_l(t, \mu) + \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(t, \mu)(x), v[\mu](t, x) + \omega(x) \rangle d\mu(x) \right) \\ &= \sum_{l=1}^r \zeta_l^*(t) \mathcal{C}_l(t, \mu, \omega) \end{aligned}$$

is regular and Wasserstein-differentiable at any  $\mu \in \mathcal{P}(K)$ . The Wasserstein gradients of its summands  $\mathcal{C}_l(t, \cdot, \omega)$  can be computed explicitly as

$$\begin{aligned} \nabla_\mu \mathcal{C}_l(t, \mu, \omega)(x) &= \partial_t \nabla_\mu \Lambda_l(t, \mu)(x) + D_x \nabla_\mu \Lambda_l(t, \mu)(x)^\top (v[\mu](t, x) + \omega(x)) \\ &\quad + (D_x v[\mu](t, x)^\top + D_x u^*(t, x)^\top) \nabla_\mu \Lambda_l(t, \mu)(x) + \int_{\mathbb{R}^d} \mathbb{I}_{(t, y)}^v(x)^\top \nabla_\mu \Lambda_l(t, \mu)(y) d\mu(y) \quad (\text{B.1}) \\ &\quad + \int_{\mathbb{R}^d} \mathbb{I}_{(t, y)}^{\nabla_\mu \Lambda_l}(x)^\top (v[\mu](t, y) + u^*(t, y)) d\mu(y), \end{aligned}$$

where  $(t, x, y) \mapsto \mathbb{I}_{(t, y)}^{\nabla_\mu \Lambda_l}(x) = \left( \mathbb{I}_{(t, y)}^{\nabla_\mu \Lambda_l, i}(x) \right)_{1 \leq i \leq d}$  are the matrix-valued maps in which rows are the Wasserstein gradients of the components of  $\nabla_\mu \Lambda_l^i(t, \mu)(x)$ , i.e.

$$\mathbb{I}_{(t, y)}^{\nabla_\mu \Lambda_l, i}(x) = \nabla_\mu \left[ \nabla_\mu \Lambda_l^i(t, \cdot)(y) \right](\mu)(x).$$

*Proof.* In order to lighten the coming computations, we introduce the auxiliary functions

$$\mathcal{C}_l^1(t, \mu) = \partial_t \Lambda_l(t, \mu) \quad \text{and} \quad \mathcal{C}_l^2(t, \mu, \omega) = \int_{\mathbb{R}^d} \langle \nabla_\mu \Lambda_l(t, \mu)(x), v[\mu](t, x) + \omega(x) \rangle d\mu(x).$$

Let  $t \in [0, T]$  and  $\mu \in \mathcal{P}(K)$ . The Wasserstein gradient of  $\mathcal{C}_l^1(t, \cdot)$  at  $\mu$  is given by

$$\nabla_\mu \mathcal{C}_l^1(t, \mu) = \nabla_\mu (\partial_t \Lambda_l(t, \cdot))(\mu) = \partial_t \nabla_\mu \Lambda_l(t, \mu). \quad (\text{B.2})$$

We turn our attention to  $\mathcal{C}_l^2(t, \cdot, \omega)$ . For any  $\nu \in \mathcal{P}(K)$  and  $\gamma \in \Gamma_o(\mu, \nu)$ , it holds that

$$\begin{aligned} &\mathcal{C}_l^2(t, \nu, \omega) - \mathcal{C}_l^2(t, \mu, \omega) \\ &= \int_{\mathbb{R}^{2d}} \left( \langle \nabla_\mu \Lambda_l(t, \nu)(y), v[\nu](t, y) + u^*(t, y) \rangle - \langle \nabla_\mu \Lambda_l(t, \mu)(x), v[\mu](t, x) + \omega(x) \rangle \right) d\gamma(x, y) \\ &= \int_{\mathbb{R}^d} \left( \langle \nabla_\mu \Lambda_l(t, \nu)(x), v[\nu](t, x) + \omega(x) \rangle - \langle \nabla_\mu \Lambda_l(t, \mu)(x), v[\mu](t, x) + \omega(x) \rangle \right) d\mu(x) \quad (\text{B.3}) \\ &\quad + \int_{\mathbb{R}^{2d}} \langle D_x \nabla_\mu \Lambda_l(t, \mu)(x)^\top (v[\mu](x) + \omega(x)), y - x \rangle d\gamma(x, y) \\ &\quad + \int_{\mathbb{R}^{2d}} \langle (D_x v[\mu](t, x)^\top + D_x \omega(x)^\top) \nabla_\mu \Lambda_l(t, \mu)(x), y - x \rangle d\gamma(x, y) + \int_{\mathbb{R}^{2d}} o(|x - y|) d\gamma(x, y). \end{aligned}$$

By definition of the Landau notation  $o(\cdot)$ , for any  $\epsilon > 0$  there exists  $\eta > 0$  such that whenever  $|x - y| \leq \eta$ , one has that  $o(|x - y|) \leq \epsilon|x - y|$ . Therefore,

$$\begin{aligned} \int_{\mathbb{R}^{2d}} o(|x - y|) d\gamma(x, y) &\leq \epsilon \int_{\{|x-y|\leq\eta\}} |x - y| d\gamma(x, y) + C \int_{\{|x-y|>\eta\}} |x - y| d\gamma(x, y) \\ &\leq \epsilon W_2(\mu, \nu) + 2C \text{diam}(K) \gamma(\{(x, y) \in \mathbb{R}^{2d} \text{ s.t. } |x - y| > \eta\}) \\ &\leq \epsilon W_2(\mu, \nu) + \frac{2C}{\eta^2} \text{diam}(K) W_2^2(\mu, \nu) \end{aligned}$$

by Chebyshev's inequality and where the constant  $C > 0$  exists because  $o(|x - y|)$  is in particular a  $O(|x - y|)$  on the compact set  $\text{supp}(\gamma) \subset \mathbb{R}^{2d}$  for  $|x - y| > \eta$ . Upon choosing  $\eta' = \eta^2 \epsilon / (2C \text{diam}(K))$ , we recover that

$$\int_{\mathbb{R}^{2d}} o(|x - y|) d\gamma(x, y) \leq 2\epsilon W_2(\mu, \nu),$$

whenever  $W_2(\mu, \nu) \leq \eta'$ . By definition, this estimate precisely amounts to the fact that  $\int_{\mathbb{R}^{2d}} o(|x - y|) d\gamma(x, y) = o(W_2(\mu, \nu))$  as  $W_2(\mu, \nu) \rightarrow 0$ .

We further compute the first-order variations arising from the remaining measure terms as follows

$$\begin{aligned} &\int_{\mathbb{R}^d} \left( \langle \nabla_{\mu} \Lambda_l(t, \nu)(x), v[\nu](t, x) + \omega(x) \rangle - \langle \nabla_{\mu} \Lambda_l(t, \mu)(x), v[\mu](t, x) + \omega(x) \rangle \right) d\mu(x) \\ &= \int_{\mathbb{R}^d} \left\langle \int_{\mathbb{R}^{2d}} \mathbb{I}_{(t,x)}^{\nabla_{\mu} \Lambda_l}(x')(y' - x') d\gamma'(x', y'), v[\mu](t, x) + \omega(x) \right\rangle d\mu(x) \\ &\quad + \int_{\mathbb{R}^d} \left\langle \int_{\mathbb{R}^{2d}} \mathbb{I}_{(t,x)}^v(x')(y' - x') d\gamma'(x', y'), \nabla_{\mu} \Lambda_l(t, \mu)(x) \right\rangle d\mu(x) + o(W_2(\mu, \nu)) \\ &= \int_{\mathbb{R}^{2d}} \left\langle \int_{\mathbb{R}^d} \left( \mathbb{I}_{(t,y')}^v(x)^\top \nabla_{\mu} \Lambda_l(t, \mu)(y') + \mathbb{I}_{(t,y')}^{\nabla_{\mu} \Lambda_l}(x)^\top \nabla_{\mu} \Lambda_l(t, \mu)(y') \right) d\mu(y'), y - x \right\rangle d\gamma(x, y) + o(W_2(\mu, \nu)) \end{aligned} \tag{B.4}$$

by a standard application of Fubini's Theorem. Merging equations (B.2), (B.3) and (B.4), we recover the characterization (2.6) of the Wasserstein gradient  $\nabla_{\mu} \mathcal{E}_l(t, \mu, \omega)(\cdot)$  of  $\mathcal{E}_l(t, \cdot, \omega)$  at  $\mu$  given by (B.1).  $\square$

## REFERENCES

- [1] Y. Achdou and M. Laurière, On the system of partial differential equations arising in mean field type control. *Discrete Contin. Dyn. Syst.* **35** (2015) 3879–3900.
- [2] Y. Achdou and M. Laurière, Mean field type control with congestion. *Appl. Math. Optim.* **73** (2016) 393–418.
- [3] A. Agrachev and Y. Sachkov, in Control Theory from the Geometric Viewpoint, Vol. 87 of *Encyclopaedia of Mathematical Sciences*. Springer, Berlin (2004).
- [4] G. Albi, M. Bongini, E. Cristiani and D. Kalise, Invisible control of self-organizing agents leaving unknown environments. *SIAM J. Appl. Math.* **76** (2016) 1683–1710.
- [5] G. Albi, L. Pareschi and M. Zanella, Boltzmann type control of opinion consensus through leaders. *Proc. of the Roy. Soc. A.* **372** (2014) 20140138.
- [6] L. Ambrosio, Transport equation and Cauchy problem for BV vector fields. *Invent. Math.* **158** (2004) 227–260.
- [7] L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variations and Free Discontinuity Problems. *Oxford Mathematical Monographs*. Clarendon Press, Oxford (2000).
- [8] L. Ambrosio and W. Gangbo, Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.* **61** (2008) 18–53.
- [9] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd. edn. *Lectures in Mathematics, ETH Zürich*. Birkhäuser Verlag, Basel (2008).
- [10] A.V. Arutyunov, D.Y. Karamzin and F.L. Pereira, The maximum principle for optimal control problems with state constraints by R.V. Gamkrelidze: Revisited. *J. Optim. Theor. Appl.* **149** (2011) 474–493.
- [11] A.V. Arutyunov and R. Vinter, A simple “finite approximations” proof of the Pontryagin Maximum Principle under reduced differentiability hypotheses. *Set-Valued Anal.* **12** (2004) 5–24.

- [12] N. Bellomo, M.A. Herrero and A. Tosin, On the dynamics of social conflicts: looking for the black swan. *Kinet. Relat. Models* **6** (2013) 459–479.
- [13] M. Bongini, M. Fornasier, F. Rossi and F. Solombrino, Mean field Pontryagin maximum principle. *J. Optim. Theor. Appl.* **175** (2017) 1–38.
- [14] B. Bonnet and F. Rossi, The Pontryagin Maximum Principle in the Wasserstein Space. *Calc. Var. Partial Differ. Equ.* **58** (2019) 11.
- [15] A. Bressan and B. Piccoli, Introduction to the Mathematical Theory of Control, Vol. 2 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO (2007).
- [16] H. Brézis, Functional Analysis, Sobolev Spaces and Partial Differential Equations. In *Universitext*. Springer, Berlin (2010).
- [17] M. Burger, R. Pinnau, O. Totzeck and O. Tse, Mean-field optimal control and optimality conditions in the space of probability measures. arXiv:1902.05339.
- [18] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, The master equation and the convergence problem in mean field games. Preprint <https://arxiv.org/abs/1509.02505> (2015).
- [19] P. Cardaliaguet and L. Silvester, Hölder continuity to Hamilton-Jacobi equations with super-quadratic growth in the gradient and unbounded right-hand side. *Commun. Partial Differ. Equ.* **37** (2012) 1668–1688.
- [20] R. Carmona and F. Delarue, Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. *Ann. Probab.* **43** (2015) 2647–2700.
- [21] R. Carmona, F. Delarue and A. Lachapelle, Control of McKean–Vlasov dynamics versus mean field games. *Math. Financial Econ.* **7** (2013) 131–166.
- [22] J.A. Carrillo, M. Fornasier, J. Rosado and G. Toscani, Asymptotic flocking for the kinetic Cucker-Smale model. *SIAM J. Math. Anal.* **42** (2010) 218–236.
- [23] G. Cavagnari, A. Marigonda, K.T. Nguyen and F.S. Priuli, Generalized control systems in the space of probability measures. *Set-Valued Var. Anal.* **26** (2018) 663–691.
- [24] G. Cavagnari, A. Marigonda and B. Piccoli, Averaged time-optimal control problem in the space of positive Borel measures. *ESAIM: COCV* **24** (2018) 721–740.
- [25] F. Clarke, Functional Analysis, Calculus of Variations and Optimal Control. Springer, Berlin (2013).
- [26] F. Cucker and S. Smale, On the mathematics of emergence. *Jpn. J. Math.* **2** (2007) 197–227.
- [27] R.L. Di Perna and P.-L. Lions, Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.* **98** (1989) 511–548.
- [28] J. Diestel and J.J. Jr Uhl. Vector Measures, Vol. 15. American Mathematical Society, Rhode Island (1977).
- [29] J. Dugundji. An extension of Tietze’s theorem. *Pac. J. Math.* **1** (1951) 353–367.
- [30] M. Duprez, M. Morancey and F. Rossi, Approximate and exact controllability of the continuity equation with a localized vector field. *SIAM J. Control Optim.* **57** (2019) 1284–1311.
- [31] K. Elamvazhuthi and S. Berman, Optimal control of stochastic coverage strategies for robotic swarms. *IEEE International Conference on Robotics and Automation* (2015).
- [32] A. Ferscha and K. Zia, Lifebelt: crowd evacuation based on vibro-tactile guidance. *IEEE Pervasive Comput.* **9** (2010) 33–42.
- [33] M. Fornasier, S. Lisini, C. Orrieri and G. Savaré, Mean-field optimal control as gamma-limit of finite agent controls. *Eur. J. Appl. Math.* (2019) 1–34.
- [34] M. Fornasier, B. Piccoli and F. Rossi, Mean-field sparse optimal control. *Phil. Trans. R. Soc. A* **372** (2014) 20130400.
- [35] M. Fornasier and F. Solombrino, Mean field optimal control. *ESAIM: COCV* **20** (2014) 1123–1152.
- [36] W. Gangbo, T. Nguyen and A. Tudorascu, Hamilton-Jacobi equations in the Wasserstein space. *Methods Appl. Anal.* **15** (2008) 155–184.
- [37] W. Gangbo and A. Tudorascu, On differentiability in Wasserstein spaces and well-posedness for Hamilton-Jacobi equations, *Technical Report*, 2017.
- [38] S.Y. Ha and J.G. Liu, A simple proof of the Cucker-Smale flocking dynamics and mean-field limit. *Commun. Math. Sci.* **7** (2009) 297–325.
- [39] R. Hegselmann and U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simul.* **5** (2002) 1–33.
- [40] A.D. Ioffe, A Lagrange multiplier rule with small convex-valued subdifferentials for nonsmooth problems of mathematical programming involving equality and nonfunctional constraints. *Math. Program.* **58** (1993) 137–145.
- [41] A.D. Ioffe and V.M. Tihomirov, Theory of Extremal Problems. North Holland Publishing Company, Elsevier (1979).
- [42] L.V. Kantorovich, On the translocation of mass. *Dokl. Akad. Nauk. USSR* **37** (1942) 199–201.
- [43] J.-M. Lasry and P.-L. Lions, Mean field games. *Jpn. J. Math.* **2** (2007) 229–260.
- [44] J.P. Penot and P. Michel, Calcul sous-différentiel pour les fonction Lipschitziennes et non-Lipschitziennes. *C. R. Acad. Sci. Paris Sér. I* **298** (1984) 269–272.
- [45] B. Piccoli and F. Rossi, Transport equation with nonlocal velocity in Wasserstein spaces: convergence of numerical schemes. *Acta Appl. Math.* **124** (2013) 73–105.
- [46] B. Piccoli, F. Rossi and E. Trélat, Control of the kinetic Cucker-Smale model. *SIAM J. Math. Anal.* **47** (2015) 4685–4719.
- [47] N. Pogodaev, Numerical algorithm for optimal control of continuity equations. Preprint [arXiv:1708.05516](https://arxiv.org/abs/1708.05516) (2017).
- [48] N. Pogodaev, Optimal control of continuity equations. *Nonlinear Differ. Equ. Appl.* **23** (2016) 21.
- [49] W. Rudin, Real and Complex Analysis. *Mathematical Series*. McGraw-Hill International Editions (1987).

- [50] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Vol. 87. Birkhauser, Basel (2015).
- [51] I.A. Shvartsman, New approximation method in the proof of the maximum principle for nonsmooth optimal control problems with state constraints. *J. Math. Anal. Appl.* **326** (2006) 974–1000.
- [52] F. Tröltzsch, *Optimal Control of Partial Differential Equations*. American Mathematical Society, Rhode Island (2010).
- [53] C. Villani, *Optimal Transport: Old and New*. Springer-Verlag, Berlin (2009).
- [54] R.B. Vinter, *Optimal Control. Modern Birkhauser Classics*. Birkhauser, Basel (2000).
- [55] A.A. Vlasov, *Many-Particle Theory and its Application to Plasma*. Gordon and Breach, New York (1961).