

# Monge's transport problem in the Heisenberg group

Luigi De Pascale and Severine Rigot

Communicated by G. Mingione

**Abstract.** We prove the existence of a solution to Monge's transport problem between two compactly supported Borel probability measures in the Heisenberg group equipped with its Carnot–Carathéodory distance assuming that the initial measure is absolutely continuous with respect to the Haar measure of the group.

**Keywords.** Monge transport problem, optimal transport map, Monge–Kantorovich problem, Heisenberg group, sub-Riemannian distance.

**2010 Mathematics Subject Classification.** 49Q20; 53C17.

## 1 Introduction

The classical Monge's transport problem refers to the problem of moving one distribution of mass onto another as efficiently as possible, where the efficiency criterion is expressed in terms of the average distance transported. It originates from a paper by G. Monge, *Mémoire sur la théorie des déblais et des remblais*, in 1781. Rephrased and generalized in modern mathematical terms, we are given two Borel probability measures  $\mu$  and  $\nu$  on a metric space  $(X, d)$  and we want to minimize

$$T \mapsto \int_X d(x, T(x)) d\mu(x)$$

among all transport maps  $T$  from  $\mu$  to  $\nu$ , i.e., all  $\mu$ -measurable maps  $T : X \rightarrow X$  such that  $T_{\#}\mu = \nu$ , meaning that  $\nu(B) = \mu(T^{-1}(B))$  for all Borel sets  $B$ .

In this paper, we are interested in Monge's transport problem in the Heisenberg group  $(\mathbb{H}^n, d)$  equipped with its Carnot–Carathéodory distance. We prove the existence of an optimal transport map between two compactly supported Borel probability measures  $\mu$  and  $\nu$  on  $\mathbb{H}^n$  assuming that the first measure  $\mu$  is absolutely continuous with respect to the Haar measure  $\mathcal{L}^{2n+1}$  of  $\mathbb{H}^n$ .

**Theorem 1.1.** *Let  $\mu$  and  $\nu$  be two compactly supported Borel probability measures on  $\mathbb{H}^n$ . Assume that  $\mu \ll \mathcal{L}^{2n+1}$ . Then there exists an optimal transport map*

solution to Monge's transport problem between  $\mu$  and  $\nu$ , i.e., a  $\mu$ -measurable map  $T : \mathbb{H}^n \rightarrow \mathbb{H}^n$  such that  $T_{\#}\mu = \nu$  and

$$\int_{\mathbb{H}^n} d(x, T(x)) d\mu(x) = \inf_{S_{\#}\mu=\nu} \int_{\mathbb{H}^n} d(x, S(x)) d\mu(x).$$

Monge's transport problem in  $\mathbb{R}^n$  equipped with a distance induced by a norm has already been widely investigated. A first attempt to solve this problem goes back to the work of Sudakov [25]. It was however discovered some years later that the proof in [25] was not completely correct. In [14] a PDE-based alternative to Sudakov's approach has been developed. The authors prove the existence of an optimal transport map in  $\mathbb{R}^n$  equipped with the Euclidean norm under the assumptions that  $\text{spt } \mu \cap \text{spt } \nu = \emptyset$ ,  $\mu, \nu \ll \mathcal{L}^n$  with Lipschitz densities with compact support. Existence results for general absolutely continuous measures  $\mu, \nu$  with compact support have been obtained independently in [11] and [26] and have been extended to a Riemannian setting in [15]. The existence of a solution to Monge's transport problem assuming only that the initial measure  $\mu$  is absolutely continuous has been proved in [1], see also [3], [4], [10]. All these later proofs roughly involve a Sudakov-type dimension reduction argument, via different technical implementations though, and require some regularity assumptions about the norm  $\mathbb{R}^n$  is endowed with. For some time, it seemed that there were indeed some borderline cases about the norms that could not be attacked through these techniques.

Recently another approach that does not go through Sudakov-type arguments and in particular does not require disintegration of measures has been developed in [13], see also [12], to solve Monge's transport problem for general norms in  $\mathbb{R}^n$ . This approach relies on rather simple but powerful density results. In the present paper we follow closely this approach. We basically show that a very similar strategy can be implemented in the context of the Heisenberg group equipped with its Carnot–Carathéodory distance. The main features that play a role in this approach are that  $(\mathbb{H}^n, d, \mathcal{L}^{2n+1})$  is a doubling polish metric measure space, a non-branching geodesic space and satisfies a so-called Measure Contraction Property. It is very likely that this approach can be extended to more general metric measure spaces, see Section 9. We have chosen however to present the particular case of the Heisenberg group for simplicity, this space being moreover an instructive explicit example of non-Riemannian space.<sup>1</sup>

The strategy starts by considering the nowadays classical relaxation of Monge's transport problem proposed by Kantorovich. In Kantorovich's formulation one

<sup>1</sup> The preprint [7] which appeared during the completion of the present paper also addresses Monge's transport problem in metric spaces with a geodesic distance using a Sudakov-type dimension reduction argument and disintegration of measures.

considers transport plans, i.e., Borel probability measures on  $X \times X$  with first and second marginals  $\mu$  and  $\nu$  respectively. Denoting by  $\Pi(\mu, \nu)$  the class of all transport plans, one wants to minimize

$$\gamma \mapsto \int_{X \times X} d(x, y) d\gamma(x, y)$$

among all  $\gamma \in \Pi(\mu, \nu)$ . Due to the linearity of the constraint  $\gamma \in \Pi(\mu, \nu)$ , weak topologies provide existence of optimal transport plans. As a classical fact it turns out that whenever an optimal transport plan is induced by a  $\mu$ -measurable map  $T$ , i.e., can be written in the form  $(I \otimes T)_\# \mu$  where  $(I \otimes T)(x) := (x, T(x))$ , then  $T$  is an optimal transport map between  $\mu$  and  $\nu$  solution to Monge's transport problem. We follow here this scheme seeking after optimal transport plans that will be shown to be eventually induced by  $\mu$ -measurable maps.

In our present context we first prove that any optimal transport plan is concentrated on a set of pairs of points that are connected by a unique minimal curve and that these transport rays cannot bifurcate (see Section 4). Next, following ideas already introduced in the literature and more specifically here inspired by [24], one introduces variational approximations (see Section 5). This procedure allows to select optimal transport plans with specific properties. These transport plans will eventually be proved to be induced by  $\mu$ -measurable maps. This procedure is here essentially twofold. On one hand it allows to select optimal transport plans that are solution to a secondary variational problem. This secondary variational problem prescribes the geometry of transport rays. The selected transport plans are indeed shown to be monotonic along transport rays (see Lemma 4.3). On the other hand, given a transport plan, one can in our context interpolate between its first and second marginal in a natural way (see Subsection 3.3). Absolute continuity and more importantly  $L^\infty$ -estimates on the density of the interpolations will play an important role and one can indeed prove  $L^\infty$ -estimates on the interpolations in the approximating variational problems (see Proposition 5.7). These estimates rely on the so-called Measure Contraction Property of  $\mathbb{H}^n$ . In the limit one will eventually get suitable  $L^\infty$ -estimates on the interpolations constructed from optimal transport plans selected through the variational approximation procedure. Next we note that some properties of plans with absolutely continuous first marginal proved in [13] can be easily generalized to our setting (see Section 6). These properties are independent of the transport problem. They rely on the notion of Lebesgue points of functions and Lebesgue points of sets, notions which make sense for instance in any doubling metric measure space. Together with the above mentioned  $L^\infty$ -estimates on the interpolations, one can in particular prove density estimate on the transport set of selected optimal transport plans in the same way as in [13]

(see Section 7). All together, namely combining this later density estimate on the transport set (Lemma 7.1) with Lemma 6.2 and remembering the monotonicity along transport rays (Lemma 4.3) it turns out that the selected transport plans are necessarily induced by a transport map as eventually proved in Theorem 8.1.

The paper is organized as follows. In Section 2 we recall classical facts about optimal transportation for later use. In Section 3 we describe the Heisenberg group focusing on the features that will be needed in this paper. In Section 4 we prove geometric properties of optimal transport plans and the monotonicity along transport rays of solutions to the secondary variational problem. The variational approximations are introduced and studied in Section 5. In Section 6 we state in our framework properties of plans with absolutely continuous first marginal proved in [13] in  $\mathbb{R}^n$ . This section, independent of the transport problem, contains density results that play an essential role in the strategy followed here. In Section 7 we prove lower bounds on the density, in some suitable sense, of the transport set of optimal transport plans selected through the variational approximations. We conclude in Section 8 proving that the selected transport plans are induced by a transport map. We discuss in the final Section 9 some possible extensions of this approach to other spaces.

## 2 Preliminaries on optimal transportation

We recall some well-known facts about optimal transportation confining ourselves to statements that will fit our needs in the rest of the paper. More general versions of these results hold in more general contexts. We refer to e.g. [27] and the references therein.

Let  $(X, d)$  be a Polish space, i.e., a complete and separable metric space. We denote by  $\mathcal{P}(X)$  the set of all Borel probability measures on  $X$  and by  $\mathcal{P}_c(X)$  the set of Borel probability measures on  $X$  with compact support. The weak topology we consider on  $\mathcal{P}(X)$  is the topology induced by convergence against bounded and continuous test functions (or narrow topology).

### 2.1 Kantorovich transport problem

Let  $\mu, \nu \in \mathcal{P}(X)$ . We denote by

$$\Pi(\mu, \nu) := \{\gamma \in \mathcal{P}(X \times X); (\pi_1)_\# \gamma = \mu, (\pi_2)_\# \gamma = \nu\}$$

the set of all transport plans between  $\mu$  and  $\nu$ . Here  $\pi_1, \pi_2 : X \times X \rightarrow X$  denote the canonical projections on the first and second factor respectively.

Given  $c : X \times X \rightarrow [0, +\infty]$  a lower semicontinuous cost function, we look at Kantorovich transport problem between  $\mu$  and  $\nu$  with cost  $c$ :

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) d\gamma(x, y). \quad (2.1)$$

As a classical fact, existence of solutions to (2.1) follows from the weak compactness of  $\Pi(\mu, \nu)$  together with the lower semicontinuity of the functional to be minimized. We call them optimal transport plans.

**Cyclical monotonicity.** We say that a set  $\Gamma \subset X \times X$  is  $c$ -cyclically monotone if

$$\sum_{i=1}^N c(x_i, y_i) \leq \sum_{i=1}^N c(x_{i+1}, y_i)$$

whenever  $N \geq 2$  and  $(x_1, y_1), \dots, (x_N, y_N) \in \Gamma$ .

**Theorem 2.1.** *Let  $\gamma \in \Pi(\mu, \nu)$  be an optimal transport plan and assume that  $\int_{X \times Y} c(x, y) d\gamma < +\infty$ . Then  $\gamma$  is concentrated on a  $c$ -cyclically monotone Borel set.*

**Dual formulation – Kantorovich potentials.** Let  $\psi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ . We say that  $\psi$  is  $c$ -concave if  $\psi \not\equiv -\infty$  and if there exists  $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$ ,  $\varphi \not\equiv -\infty$ , such that

$$\psi(x) = \inf_{y \in X} c(x, y) - \varphi(y).$$

**Theorem 2.2.** *Let  $X$  be a Polish metric space and let  $c : X \times X \rightarrow [0, +\infty[$  be a lower semicontinuous cost function. Assume that*

$$\forall (x, y) \in X \times X, \quad c(x, y) \leq a(x) + b(y)$$

*for some  $a \in L^1(\mu)$  and  $b \in L^1(\nu)$ . Then one has*

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) d\gamma(x, y) = \max \int_X \psi(x) d\mu(x) + \int_X \psi^c(y) d\nu(y) \quad (2.2)$$

*where the above maximum is taken among all  $c$ -concave functions  $\psi$  and  $\psi^c(y) := \inf_{x \in X} c(x, y) - \psi(x)$ .*

**Definition 2.3** (Kantorovich potentials). We say that  $\psi : X \rightarrow \mathbb{R} \cup \{-\infty\}$  is a Kantorovich potential if  $\psi$  is a  $c$ -concave maximizer for the right-hand side of (2.2).

**Theorem 2.4.** *With the same assumptions as in Theorem 2.2, let  $\psi$  be a Kantorovich potential. Then  $\gamma \in \Pi(\mu, \nu)$  is an optimal transport plan if and only if*

$$c(x, y) = \psi(x) + \psi^c(y) \quad \gamma\text{-a.e. in } X \times X.$$

We will use these results for various cost functions. In the particular case which is the core of this paper and where  $c(x, y) = d(x, y)$  and  $\mu, \nu \in \mathcal{P}_c(X)$ , one can rephrase these results in terms of 1-Lipschitz Kantorovich potentials. More precisely, set

$$\text{Lip}_1(d) := \{u : X \rightarrow \mathbb{R}; |u(x) - u(y)| \leq d(x, y) \forall x, y \in X\}.$$

**Theorem 2.5.** *Let  $\mu, \nu \in \mathcal{P}_c(X)$ . Then one can find a Kantorovich potential  $u \in \text{Lip}_1(d)$  so that*

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y) d\gamma(x, y) = \int_X u(x) d\mu(x) - \int_X u(y) d\nu(y)$$

*and  $\gamma \in \Pi(\mu, \nu)$  is an optimal transport plan solution to Kantorovich transport problem (2.1) between  $\mu$  and  $\nu$  with cost  $c(x, y) = d(x, y)$  if and only if*

$$u(x) - u(y) = d(x, y) \quad \gamma\text{-a.e. in } X \times X.$$

## 2.2 Transport problem

Let  $\mu, \nu \in \mathcal{P}(X)$ . We say that a  $\mu$ -measurable map  $T : X \rightarrow X$  is a transport map between  $\mu$  and  $\nu$  if  $T_{\#}\mu = \nu$ , i.e.,  $\nu(B) = \mu(T^{-1}(B))$  for all Borel sets  $B$ .

Given  $c : X \times X \rightarrow [0, +\infty]$  a continuous cost function, we look at the transport problem between  $\mu$  and  $\nu$  with cost  $c$ :

$$\min_{T_{\#}\mu = \nu} \int_X c(x, T(x)) d\mu(x). \quad (2.3)$$

We say that a transport plan  $\gamma \in \Pi(\mu, \nu)$  is induced by a transport if there exists a  $\mu$ -measurable map  $T : X \rightarrow X$  such that  $(I \otimes T)_{\#}\mu = \gamma$  where  $(I \otimes T)(x) := (x, T(x))$ . Such a map is automatically a transport map between  $\mu$  and  $\nu$ . We also recall that if a transport plan  $\gamma$  is concentrated on a  $\gamma$ -measurable graph then  $\gamma$  is induced by a transport.

**Theorem 2.6** (Optimal transport plans versus optimal transport maps).

- (i) *Assume that  $\gamma$  is an optimal transport plan solution to Kantorovich transport problem (2.1) and that  $\gamma$  is induced by transport  $T$ . Then  $T$  is an optimal transport map solution to the transport problem (2.3).*

- (ii) Assume that any optimal transport plan solution to Kantorovich transport problem (2.1) is induced by transport. Then there exists a unique optimal transport map solution to the transport problem (2.3).

### 3 Preliminaries on $\mathbb{H}^n$

We consider the Heisenberg group  $\mathbb{H}^n$  equipped with its Carnot–Carathéodory distance. Endowed with this distance  $\mathbb{H}^n$  is a polish geodesic and non-branching metric space and a doubling metric measure space when equipped with its Haar measure.

#### 3.1 The Heisenberg group

The Heisenberg group  $\mathbb{H}^n$  is a connected, simply connected Lie group with stratified Lie algebra. We identify it with  $\mathbb{C}^n \times \mathbb{R}$  equipped with the group law

$$[\zeta, t] \cdot [\zeta', t'] := \left[ \zeta + \zeta', t + t' + 2 \sum_{j=1}^n \operatorname{Im} \zeta_j \bar{\zeta}'_j \right]$$

where  $\zeta = (\zeta_1, \dots, \zeta_n)$ ,  $\zeta' = (\zeta'_1, \dots, \zeta'_n) \in \mathbb{C}^n$  and  $t, t' \in \mathbb{R}$ . The unit element is 0 and the center of the group is

$$L := \{[0, t] \in \mathbb{H}^n; t \in \mathbb{R}\}.$$

There is a natural family of dilations  $\delta_r$  on  $\mathbb{H}^n$  defined by  $\delta_r([\zeta, t]) := [r\zeta, r^2t]$ . These dilations are group homomorphisms.

We may also identify  $\mathbb{H}^n$  with  $\mathbb{R}^{2n+1}$  via the correspondence  $[\zeta, t] = (\xi, \eta, t)$  where  $\xi = (\xi_1, \dots, \xi_n)$ ,  $\eta = (\eta_1, \dots, \eta_n) \in \mathbb{R}^n$ ,  $t \in \mathbb{R}$  and  $\zeta = (\zeta_1, \dots, \zeta_n) \in \mathbb{C}^n$  with  $\zeta_j = \xi_j + i\eta_j$ . The horizontal subbundle of the tangent bundle is defined by

$$\mathcal{H} := \operatorname{span}\{X_j; j = 1, \dots, n\} \oplus \operatorname{span}\{Y_j; j = 1, \dots, n\}$$

where the left invariant vector fields  $X_j$  and  $Y_j$  are given by

$$X_j := \partial_{\xi_j} + 2\eta_j \partial_t, \quad Y_j := \partial_{\eta_j} - 2\xi_j \partial_t.$$

Vector fields in  $\mathcal{H}$  will be called horizontal vector fields. Setting  $T := \partial_t$ , the only non-trivial bracket relations are  $[X_j, Y_j] = -4T$  hence  $\operatorname{span}\{T\} = [\mathcal{H}, \mathcal{H}]$  and the Lie algebra  $\mathcal{H}^n$  of  $\mathbb{H}^n$  admits the stratification  $\mathcal{H}^n = \mathcal{H} \oplus \operatorname{span}\{T\}$ .

The Lebesgue measure  $\mathcal{L}^{2n+1}$  on  $\mathbb{H}^n \approx \mathbb{R}^{2n+1}$  is a Haar measure of the group. It is  $(2n+2)$ -homogeneous with respect to the dilations,

$$\mathcal{L}^{2n+1}(\delta_r(A)) = r^{2n+2} \mathcal{L}^{2n+1}(A)$$

for all Borel sets  $A$  and all  $r > 0$ .

### 3.2 Carnot–Carathéodory distance

The Carnot–Carathéodory distance on  $\mathbb{H}^n$  is defined by

$$d(x, y) = \inf\{\text{length}_{g_0}(\gamma); \gamma \text{ horizontal } C^1\text{-smooth curve joining } x \text{ to } y\}, \quad (3.1)$$

where a  $C^1$ -smooth curve is said to be horizontal if, at every point, its tangent vector belongs to the horizontal subbundle of the tangent bundle and  $g_0$  is the left invariant Riemannian metric which makes  $(X_1, \dots, X_n, Y_1, \dots, Y_n, T)$  an orthonormal basis. For a general presentation of Carnot–Carathéodory spaces, see e.g. [8], [21].

The topology induced by this distance is the original (Euclidean) topology on  $\mathbb{H}^n \approx (\mathbb{R}^{2n+1}, g_0)$  and  $(\mathbb{H}^n, d)$  is a complete metric space. The distance is left invariant and 1-homogeneous with respect to the dilations,

$$d(x \cdot y, x \cdot z) = d(y, z) \quad \text{and} \quad d(\delta_r(y), \delta_r(z)) = r d(y, z)$$

for all  $x, y, z \in \mathbb{H}^n$  and all  $r > 0$ . It follows in particular that  $B(x, r) = x \cdot \delta_r(B(0, 1))$  and hence

$$\mathcal{L}^{2n+1}(B(x, r)) = c_n r^{2n+2} \quad (3.2)$$

for all  $x \in \mathbb{H}^n$ , all  $r > 0$  and where  $c_n := \mathcal{L}^{2n+1}(B(0, 1)) > 0$ . The measure  $\mathcal{L}^{2n+1}$  is in particular a doubling measure on  $(\mathbb{H}^n, d)$ . For more details about doubling metric measure spaces, see e.g. [19].

As a classical fact, endowed with its Carnot–Carathéodory distance,  $\mathbb{H}^n$  is a geodesic space, i.e., for all  $x, y \in \mathbb{H}^n$ , there exists a curve  $\sigma \in C([a, b], \mathbb{H}^n)$  such that  $\sigma(a) = x$ ,  $\sigma(b) = y$  and  $d(x, y) = l(\sigma)$  where

$$l(\sigma) = \sup_{N \in \mathbb{N}^*} \sup_{a=t_0 \leq \dots \leq t_N=b} \sum_{i=0}^{N-1} d(\sigma(t_i), \sigma(t_{i+1})).$$

Up to a reparameterization one can always assume that length minimizing curves  $\sigma$  are parameterized proportionally to arc-length, i.e.,

$$d(\sigma(s), \sigma(s')) = v (s' - s)$$

for all  $s < s' \in [a, b]$ , where  $v := d(\sigma(a), \sigma(b))/(b - a)$  is the (constant) speed of the curve. For proofs and further details we refer to e.g. [9] and [18]. As a convention we will use throughout this paper the terminology *minimal curves* to denote length minimizing curves parameterized proportionally to arc-length.



**Definition 3.1** (Minimal curves). We say that a continuous curve  $\sigma : [a, b] \rightarrow \mathbb{H}^n$  is a minimal curve if  $l(\sigma) = d(\sigma(a), \sigma(b))$  and  $\sigma$  is parameterized proportionally to arc-length.

In general Carnot–Carathéodory spaces, issues about uniqueness and regularity of minimal curves between any two points as well as issues about the regularity of the distance function to a given point could be delicate. In the specific case of the Heisenberg group, equations of all minimal curves can be explicitly computed and exploited to overcome these difficulties. We recall below the description of minimal curves in  $\mathbb{H}^n$ , see e.g. [17], [5]. We set

$$\Omega := \{(x, y) \in \mathbb{H}^n \times \mathbb{H}^n; x^{-1} \cdot y \notin L\}. \quad (3.3)$$

**Theorem 3.2** (Minimal curves in  $\mathbb{H}^n$ ). *Minimal curves in  $\mathbb{H}^n$  are horizontal  $C^1$ -smooth curves such that the infimum in (3.1) is achieved. One has the more precise description:*

- (i) *Non-trivial minimal curves starting from 0 and parameterized on  $[0, 1]$  are all curves  $\sigma_{\chi, \varphi}$  for some  $\chi \in \mathbb{C}^n \setminus \{0\}$  and  $\varphi \in [-2\pi, 2\pi]$  where*

$$\sigma_{\chi, \varphi}(s) = \left[ i \frac{(e^{-i\varphi s} - 1)\chi}{\varphi}, 2|\chi|^2 \frac{\varphi s - \sin(\varphi s)}{\varphi^2} \right]$$

*if  $\varphi \in [-2\pi, 2\pi] \setminus \{0\}$  and*

$$\sigma_{\chi, \varphi}(s) = [\chi s, 0]$$

*if  $\varphi = 0$ . Moreover one has  $|\chi| = d(0, \sigma_{\chi, \varphi}(1))$ .*

- (ii) *For all  $(x, y) \in \Omega$ , there is a unique minimal curve  $x \cdot \sigma_{\chi, \varphi}$  between  $x$  and  $y$  for some  $\chi \in \mathbb{C}^n \setminus \{0\}$  and some  $\varphi \in (-2\pi, 2\pi)$  and one has  $|\chi| = d(x, y)$ .*
- (iii) *If  $(x, y) \notin \Omega$ ,  $x^{-1} \cdot y = [0, t]$  for some  $t \in \mathbb{R}^*$ , there are infinitely many minimal curves between  $x$  and  $y$ . These curves are all curves of the form  $x \cdot \sigma_{\chi, 2\pi}$  if  $t > 0$ ,  $x \cdot \sigma_{\chi, -2\pi}$  if  $t < 0$ , for all  $\chi \in \mathbb{C}^n$  such that  $|\chi| = \sqrt{\pi|t|}$ .*

Here and in the following,  $|\chi| = (\sum_{j=1}^n |\chi_j|^2)^{1/2}$  for  $\chi = (\chi_1, \dots, \chi_n) \in \mathbb{C}^n$ . In particular it follows from this description that  $(\mathbb{H}^n, d)$  is non-branching.

**Proposition 3.3** (Non-branching property of  $\mathbb{H}^n$ ). *The space  $(\mathbb{H}^n, d)$  is non-branching, i.e., any two minimal curves which coincide on a non-trivial interval coincide on the whole intersection of their intervals of definition.*

Equivalently for any quadruple of points  $z, x, y, y' \in \mathbb{H}^n$ , if  $z$  is a midpoint of  $x$  and  $y$  as well as a midpoint of  $x$  and  $y'$ , then  $y = y'$ .

The next lemma collects some differentiability properties of the distance function to a given point to be used later. For  $y \in \mathbb{H}^n$ , we set  $L_y := y \cdot L$ .

**Lemma 3.4.** *Let  $y \in \mathbb{H}^n$  and set  $d_y(x) := d(x, y)$ . Then the function  $d_y$  is of class  $C^\infty$  on  $\mathbb{H}^n \setminus L_y$  (equipped with the usual differential structure when identifying  $\mathbb{H}^n$  with  $\mathbb{R}^{2n+1}$ ). Moreover one has*

(i)  $|\nabla_H d_y(x)| = 1$  for all  $x \in \mathbb{H}^n \setminus L_y$  where

$$\nabla_H d_y(x) := (X_1 d_y(x) + i Y_1 d_y(x), \dots, X_n d_y(x) + i Y_n d_y(x)).$$

(ii) *If  $\nabla d_y(x) = \nabla d_{y'}(x)$  and  $d(x, y) = d(x, y')$  for some  $x \in \mathbb{H}^n \setminus (L_y \cup L_{y'})$ , then  $y = y'$ . Here  $\nabla = (\partial_{\xi_1}, \dots, \partial_{\xi_n}, \partial_{\eta_1}, \dots, \partial_{\eta_n}, \partial_t)$  denotes the classical gradient when identifying  $\mathbb{H}^n$  with  $\mathbb{R}^{2n+1}$ .*

*Proof.* Set  $\Phi(\chi, \varphi) := \sigma_{\chi, \varphi}(1)$  where  $\sigma_{\chi, \varphi}$  is given in Theorem 3.2. This map is a  $C^\infty$ -diffeomorphism from  $\mathbb{C}^n \setminus \{0\} \times (-2\pi, 2\pi)$  onto  $\mathbb{H}^n \setminus L$  (see e.g. [5], [20], [22]). If  $x = \Phi(\chi, \varphi) \in \mathbb{H}^n \setminus L$  with  $(\chi, \varphi) \in \mathbb{C}^n \setminus \{0\} \times (-2\pi, 2\pi)$ , one has  $d_0(x) = |\chi|$  and

$$\nabla_H d_0(x) = \frac{\chi}{|\chi|} e^{-i\varphi} \quad \text{and} \quad \partial_t d_0(x) = \frac{\varphi}{4|\chi|},$$

see [5, Lemma 3.11]. Next, by left invariance, we have  $d_y(x) = d_0(y^{-1} \cdot x)$ ,  $\nabla_H d_y(x) = \nabla_H d_0(y^{-1} \cdot x)$  and  $\partial_t d_y(x) = \partial_t d_0(y^{-1} \cdot x)$  if  $x \in \mathbb{H}^n \setminus L_y$  and the lemma follows easily.  $\square$

### 3.3 Interpolation between measures

The notion of interpolation constructed from a transport plan between any two measures will be one of the key notion to be used later. To define it in our geometrical context, we first fix a measurable selection of minimal curves, i.e., a Borel map  $S : \mathbb{H}^n \times \mathbb{H}^n \rightarrow C([0, 1], \mathbb{H}^n)$  such that for all  $x, y \in \mathbb{H}^n$ ,  $S(x, y)$  is a minimal curve joining  $x$  and  $y$ . The existence of such a measurable recipe to join any two points in  $\mathbb{H}^n$  by a minimal curve follows from general theorems about measurable selections, see e.g. [27, Chapter 7]. Next we set  $e_t(\sigma) := \sigma(t)$  for all  $\sigma \in C([0, 1], \mathbb{H}^n)$  and  $t \in [0, 1]$ . In particular  $e_t(S(x, y))$  denotes the point lying at distance  $t d(x, y)$  from  $x$  on the selected minimal curve  $S(x, y)$  between  $x$  and  $y$ .

**Definition 3.5.** Let  $\mu, \nu \in \mathcal{P}(\mathbb{H}^n)$  and let  $\gamma \in \Pi(\mu, \nu)$ . The interpolations between  $\mu$  and  $\nu$  constructed from  $\gamma$  are defined as the family  $((e_t \circ S)_{\#}\gamma)_{t \in [0,1]}$  of Borel probability measures on  $\mathbb{H}^n$ .

Note that these interpolations depend a priori on the measurable selection  $S$  of minimal curves. This is actually not a serious issue for our purposes. We will moreover always consider interpolations constructed from transport plans that are concentrated on the set  $\Omega$  on which  $S(x, y)$  is nothing but the unique minimal curve between  $x$  and  $y$ . Note also for further reference that  $S|_{\Omega}$  is continuous.

### 3.4 Intrinsic differentiability

Intrinsic differentiability properties of real-valued Lipschitz functions on  $\mathbb{H}^n$ , namely a Rademacher's type theorem, will be useful when considering 1-Lipschitz Kantorovich potentials. This theorem is a particular case of a more general result due to P. Pansu. We say that a group homomorphism  $g : \mathbb{H}^n \rightarrow \mathbb{R}$  is homogeneous if  $g(\delta_r(x)) = r g(x)$  for all  $x \in \mathbb{H}^n$  and all  $r > 0$ .

**Definition 3.6.** We say that a map  $f : \mathbb{H}^n \rightarrow \mathbb{R}$  is Pansu-differentiable at  $x \in \mathbb{H}^n$  if there exists a homogeneous group homomorphism  $g : \mathbb{H}^n \rightarrow \mathbb{R}$  such that

$$\lim_{y \rightarrow x} \frac{f(y) - f(x) - g(x^{-1} \cdot y)}{d(y, x)} = 0.$$

The map  $g$  is then unique and will be denoted by  $D_H f(x)$ .

If  $f : \mathbb{H}^n \rightarrow \mathbb{R}$  is Pansu-differentiable at  $x \in \mathbb{H}^n$  then the maps  $s \mapsto f(x \cdot \delta_s[e_j, 0])$ , resp.  $s \mapsto f(x \cdot \delta_s[e_{n+j}, 0])$ , are differentiable at  $s = 0$  and if we denote the corresponding derivatives by  $X_j f(x)$ , resp.  $Y_j f(x)$ , then

$$D_H f(x)(\xi, \eta, t) = \sum_{j=1}^n \xi_j X_j f(x) + \eta_j Y_j f(x).$$

Here  $e_j = (\delta_1^j, \dots, \delta_n^j) \in \mathbb{C}^n$  and  $e_{n+j} = (i\delta_1^j, \dots, i\delta_n^j) \in \mathbb{C}^n$ . Using similar notations as in the classical smooth case, we then set  $\nabla_H f(x) := (X_1 f(x) + iY_1 f(x), \dots, X_n f(x) + iY_n f(x))$ .

**Theorem 3.7** (Pansu-differentiability theorem). [23] *Let  $f : (\mathbb{H}^n, d) \rightarrow \mathbb{R}$  be a  $C$ -Lipschitz function. Then, for  $\mathcal{L}^{2n+1}$ -a.e.  $x \in \mathbb{H}^n$ , the function  $f$  is Pansu-differentiable at  $x$  and  $|\nabla_H f(x)| \leq C$ .*

The next lemma will be used to prove that any optimal transport plan is concentrated on the set  $\Omega$ .

**Lemma 3.8.** *Let  $u \in \text{Lip}_1(d)$ ,  $x \in \mathbb{H}^n$  be such that  $u$  is Pansu-differentiable at  $x$  with  $|\nabla_H u(x)| \leq 1$  and let  $y \in \mathbb{H}^n$  be such that  $u(x) - u(y) = d(x, y)$ . Then there exists a unique minimal curve between  $x$  and  $y$ .*

*Proof.* Let  $\sigma : [0, 1] \rightarrow \mathbb{H}^n$  be a minimal curve between  $x$  and  $y$ . Then  $\sigma$  is a horizontal  $C^1$ -smooth curve and if  $\sigma(t) = (\sigma_1(t), \dots, \sigma_{2n+1}(t)) \in \mathbb{H}^n \approx \mathbb{R}^{2n+1}$ , one has for all  $t \in [0, 1]$ ,

$$\dot{\sigma}(t) = \sum_{j=1}^n \dot{\sigma}_j(t) X_j(\sigma(t)) + \dot{\sigma}_{n+j}(t) Y_j(\sigma(t))$$

and  $|\dot{\sigma}_H(t)| = d(x, y)$  where  $\dot{\sigma}_H(t) := (\dot{\sigma}_1(t) + i \dot{\sigma}_{n+1}(t), \dots, \dot{\sigma}_n(t) + i \dot{\sigma}_{2n}(t)) \in \mathbb{C}^n$ . On the other hand, one has

$$u(x) - u(\sigma(t)) = d(x, \sigma(t)) = t d(x, y)$$

for all  $t \in [0, 1]$ . Differentiating this equality with respect to  $t$ , we get

$$\sum_{j=1}^n \dot{\sigma}_j(0) X_j u(x) + \dot{\sigma}_{n+j}(0) Y_j u(x) = \frac{d}{dt} u(\sigma(t)) \Big|_{t=0} = -d(x, y).$$

All together, it follows that

$$\begin{aligned} d(x, y) &= \left| \sum_{j=1}^n \dot{\sigma}_j(0) X_j u(x) + \dot{\sigma}_{n+j}(0) Y_j u(x) \right| \\ &\leq |\nabla_H u(x)| |\dot{\sigma}_H(0)| \leq d(x, y). \end{aligned}$$

In particular, there is equality in all the previous inequalities which implies in turn that  $\dot{\sigma}_H(0) = -d(x, y) \nabla_H u(x)$ . On the other hand one knows from Theorem 3.2 that  $\sigma = x \cdot \sigma_{\chi, \varphi}$  for some  $\chi \in \mathbb{C}^n \setminus \{0\}$  and  $\varphi \in [-2\pi, 2\pi]$ . In particular one has  $\dot{\sigma}_H(0) = \chi$ . It follows that  $\chi = -d(x, y) \nabla_H u(x)$  is uniquely determined hence there is a unique minimal curve joining  $x$  and  $y$  according once again to the description given in Theorem 3.2.  $\square$

#### 4 Properties of $\Pi_1(\mu, \nu)$ and $\Pi_2(\mu, \nu)$

Let  $\mu, \nu \in \mathcal{P}_c(\mathbb{H}^n)$  be fixed. We denote by  $\Pi_1(\mu, \nu)$  the set of optimal transport plans solution to Kantorovich transport problem (2.1) between  $\mu$  and  $\nu$  with cost  $c(x, y) = d(x, y)$ .

We first prove some geometric properties of optimal transport plans. These properties follow from the behavior of minimal curves in  $(\mathbb{H}^n, d)$ . In the next lemma, we prove that any optimal transport plan is concentrated on the set  $\Omega$  (see (3.3)) of pair of points that are connected by a unique minimal curve.

**Lemma 4.1.** *Let  $\gamma \in \Pi_1(\mu, \nu)$  and assume that  $\mu \ll \mathcal{L}^{2n+1}$ . Then for  $\gamma$ -a.e.  $(x, y)$ , there exists a unique minimal curve between  $x$  and  $y$ .*

*Proof.* Let  $u \in \text{Lip}_1(d)$  be a Kantorovich potential associated to Kantorovich transport problem (2.1) between  $\mu$  and  $\nu$  with cost  $c(x, y) = d(x, y)$  (see Section 2 and Theorem 2.5 there). Since  $u \in \text{Lip}_1(d)$ , we know from Theorem 3.7 that for  $\mathcal{L}^{2n+1}$ -a.e., and hence  $\mu$ -a.e.,  $x \in \mathbb{H}^n$ ,  $u$  is Pansu-differentiable at  $x$  with  $|\nabla_H u(x)| \leq 1$ . Then the conclusion follows from Lemma 3.8 since  $u(x) - u(y) = d(x, y)$  for  $\gamma$ -a.e.  $(x, y)$  (see Theorem 2.5).  $\square$

The next lemma says that minimal curves used by an optimal transport plan cannot bifurcate. It follows essentially from the non-branching property of  $(\mathbb{H}^n, d)$ .

**Lemma 4.2.** *Let  $\gamma \in \Pi_1(\mu, \nu)$ . Then  $\gamma$  is concentrated on a set  $\Gamma$  such that the following holds. For all  $(x, y) \in \Gamma$  and  $(x', y') \in \Gamma$  such that  $x \neq y$  and  $x \neq x'$ , if  $x'$  lies on a minimal curve between  $x$  and  $y$  then all points  $x, x', y$  and  $y'$  lie on the same minimal curve. More precisely, there exists a minimal curve  $\sigma : [a, b] \rightarrow \mathbb{H}^n$  such that  $x = \sigma(a)$ ,  $y = \sigma(t)$  for some  $t \in (a, b]$ ,  $x' = \sigma(s)$  for some  $s \in (a, t]$  and  $y' = \sigma(t')$  for some  $t' \in [s, b]$ .*

*Proof.* Let  $(x, y) \in \mathbb{H}^n \times \mathbb{H}^n$  and  $(x', y') \in \mathbb{H}^n \times \mathbb{H}^n$  such that  $x \neq y$  and  $x' \neq x$ . Assume that  $x' \in \sigma((0, d(x, y)))$  where  $\sigma : [0, d(x, y)] \rightarrow \mathbb{H}^n$  is a unit-speed minimal curve between  $x$  and  $y$ . Let  $\sigma'$  be a unit-speed minimal curve between  $x'$  and  $y'$  parameterized on  $[d(x, x'), d(x, x') + d(x', y')]$ . Assume moreover that

$$d(x, y) + d(x', y') \leq d(x, y') + d(x', y).$$

Recall that this holds true for  $\gamma$ -a.e.  $(x, y)$  and  $(x', y')$  by Theorem 2.1. Then the curve  $\tilde{\sigma} : [0, d(x, x') + d(x', y')] \rightarrow \mathbb{H}^n$  which coincides with  $\sigma$  on  $[0, d(x, x')]$

and  $\sigma'$  on  $[d(x, x'), d(x, x') + d(x', y')]$  is a length minimizing curve between  $x$  and  $y'$ . Indeed, otherwise we would have

$$\begin{aligned} d(x, y') &< l(\tilde{\sigma}) = l(\sigma|_{[0, d(x, x')])} + l(\sigma'|_{[d(x, x'), d(x, x') + d(x', y')]) \\ &= d(x, x') + d(x', y'). \end{aligned}$$

Since  $x'$  lies on a minimal curve between  $x$  and  $y$ , we have  $d(x, x') + d(x', y) = d(x, y)$  and we get

$$d(x, y') + d(x', y) < d(x, y) + d(x', y')$$

which gives a contradiction. It follows that  $\sigma$  and  $\tilde{\sigma}$  are unit-speed minimal curves that coincide on the non-trivial interval  $[0, d(x, x')]$ . Since  $\mathbb{H}^n$  is non-branching (see Proposition 3.3), this implies that  $\sigma$  and  $\tilde{\sigma}$  are sub-arcs of the same minimal curve, namely  $\sigma$  if  $d(x, y') \leq d(x, y)$  and  $\tilde{\sigma}$  otherwise, on which all points  $x, x', y$  and  $y'$  lie. And the conclusion follows.  $\square$

We denote by  $\Pi_2(\mu, \nu)$  the set of transport plans solution to the secondary variational problem:

$$\min_{\gamma \in \Pi_1(\mu, \nu)} \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\gamma(x, y).$$

Optimal transport plans selected through the variational approximations to be introduced in Section 5 will be solution to this secondary variational problem. The next lemma gives a one-dimensional monotonicity condition along minimal curves used by optimal transport plans in  $\Pi_2(\mu, \nu)$ . This follows essentially from a constrained version of  $d^2$ -cyclical monotonicity.

**Lemma 4.3.** *Let  $\gamma \in \Pi_2(\mu, \nu)$ . Then  $\gamma$  is concentrated on a set  $\Gamma$  such that the following holds. For all  $(x, y) \in \Gamma$  and  $(x', y') \in \Gamma$  such that  $x \neq y$  and  $x \neq x'$ , if  $x'$  lies on a minimal curve between  $x$  and  $y$  then all points  $x, x', y$  and  $y'$  lie on the same minimal curve ordered in that way.*

In other words, there exists a minimal curve  $\sigma : [a, b] \rightarrow \mathbb{H}^n$  such that  $\sigma(a) = x$ ,  $\sigma(t) = y$  for some  $t \in (a, b]$ ,  $\sigma(s) = x'$  for some  $s \in (a, t]$  and  $\sigma(t') = y'$  for some  $t' \in [t, b]$ .

*Proof.* First, as a classical fact, one can rephrase the secondary variational problem as a classical Kantorovich transport problem (2.1) between  $\mu$  and  $\nu$  with cost  $c(x, y) = \beta(x, y)$  with

$$\beta(x, y) = \begin{cases} d(x, y)^2 & \text{if } u(x) - u(y) = d(x, y), \\ +\infty & \text{otherwise,} \end{cases}$$

where  $u \in \text{Lip}_1(d)$  is a Kantorovich potential associated to Kantorovich transport problem (2.1) between  $\mu$  and  $\nu$  with cost  $c(x, y) = d(x, y)$  (see Section 2 and Theorem 2.5 there). Since  $\beta$  is lower semicontinuous and

$$\int_{\mathbb{H}^n \times \mathbb{H}^n} \beta(x, y) d\gamma(x, y) < +\infty$$

for all  $\gamma \in \Pi_2(\mu, \nu)$ , it follows from Theorem 2.1 that any  $\gamma \in \Pi_2(\mu, \nu)$  is concentrated on a  $\beta$ -cyclically monotone set. So, taking into account the fact that  $\Pi_2(\mu, \nu) \subset \Pi_1(\mu, \nu)$ , we know that  $\gamma \in \Pi_2(\mu, \nu)$  is concentrated on a set  $\Gamma$  such that

$$u(x) - u(y) = d(x, y)$$

for all  $(x, y) \in \Gamma$ ,

$$\beta(x, y) + \beta(x', y') \leq \beta(x, y') + \beta(x', y)$$

for all  $(x, y) \in \Gamma$  and  $(x', y') \in \Gamma$  and the conclusion of Lemma 4.2 holds.

Then let  $(x, y) \in \Gamma$  and  $(x', y') \in \Gamma$  be as in the statement. By Lemma 4.2, the conclusion will follow if we show that  $d(x', y) \leq d(x', y')$ . First we check that  $\beta(x', y) = d(x', y)^2$  and  $\beta(x, y') = d(x, y')^2$ . We have

$$u(x) \leq u(x') + d(x, x') \leq u(y) + d(x', y) + d(x, x') = d(x, y) + u(y) = u(x)$$

hence all these inequalities are equalities. In particular, we get that  $u(x') = u(y) + d(x', y)$  hence  $\beta(x', y) = d(x', y)^2$ . We also get that

$$u(x) = d(x, x') + u(x') = d(x, x') + u(y') + d(x', y') = u(y') + d(x, y')$$

hence  $\beta(x, y') = d(x, y')^2$ . If  $d(x', y') < d(x', y)$ , we get

$$\begin{aligned} & \beta(x, y') + \beta(x', y) - \beta(x', y') - \beta(x, y) \\ &= d(x, y')^2 + d(x', y)^2 - d(x', y')^2 - d(x, y)^2 \\ &= (d(x, x') + d(x', y'))^2 + d(x', y)^2 - d(x', y')^2 - (d(x, x') + d(x', y))^2 \\ &= 2d(x, x')(d(x', y') - d(x', y)) < 0 \end{aligned}$$

which gives a contradiction. □

## 5 Variational approximations

We introduce variational approximations in the spirit of [4] (see also [11], [3]) by rephrasing in our geometrical context the variational approximations considered

recently in [24]. This approximation procedure will be used to select optimal transport plans that will be eventually proved to be induced by transport maps.

Let  $\mu, \nu \in \mathcal{P}_c(\mathbb{H}^n)$  be fixed. Let  $K$  be a compact subset of  $\mathbb{H}^n$  such that  $\text{spt } \mu \cup \text{spt } \nu \subset K$  and set

$$\Pi := \{\gamma \in \mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n); (\pi_1)_\# \gamma = \mu, \text{spt } (\pi_2)_\# \gamma \subset K\}.$$

For  $\varepsilon > 0$  fixed and  $\gamma \in \Pi$ , we set

$$\begin{aligned} C_\varepsilon(\gamma) := & \frac{1}{\varepsilon} W_1((\pi_2)_\# \gamma, \nu) + \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\gamma(x, y) \\ & + \varepsilon \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\gamma(x, y) + \varepsilon^{6n+8} \text{card}(\text{spt } (\pi_2)_\# \gamma) \end{aligned}$$

and consider the family of minimization problems:

$$\min\{C_\varepsilon(\gamma); \gamma \in \Pi\}. \quad (P_\varepsilon)$$

Here  $W_1$  denotes the 1-Wasserstein distance defined for any two probability measures  $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{H}^n)$  by

$$W_1(\mu_1, \mu_2) := \min_{\gamma \in \Pi(\mu_1, \mu_2)} \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\gamma(x, y).$$

First we note that  $(P_\varepsilon)$  always admits solutions.

**Theorem 5.1.** *For any  $\varepsilon > 0$ , the problem  $(P_\varepsilon)$  admits at least one solution and  $\min\{C_\varepsilon(\gamma); \gamma \in \Pi\} < +\infty$ .*

*Proof.* First note that since  $K$  is compact,  $C_\varepsilon(\gamma) < +\infty$  for any  $\gamma \in \Pi$  such that  $(\pi_2)_\# \gamma$  is finitely atomic. Next the existence of solutions to  $(P_\varepsilon)$  follows from the weak compactness of  $\Pi$ , the lower semicontinuity of the three first terms to be minimized and the Kuratowski convergence of the supports of weakly converging probability measures (see [2, Chapter 5]).  $\square$

Next, weak limits of solutions to  $(P_\varepsilon)$  are optimal transport plans that are solutions to the secondary variational problem introduced in Section 4 to which we refer for the definition of  $\Pi_2(\mu, \nu)$ . Modulo minor modifications due to our geometrical context, this can be proved with the same arguments as those given in [24].

**Lemma 5.2.** *Let  $\varepsilon_k$  be a sequence converging to 0 and  $\gamma_{\varepsilon_k}$  a sequence of solutions to  $(P_{\varepsilon_k})$  which is weakly converging to some  $\gamma \in \mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$ . Then  $\gamma \in \Pi_2(\mu, \nu)$ .*



*Proof.* First we note that for any  $m \geq 1$ , one can find a finite set  $F_m \subset K$  such that  $\text{card } F_m \leq C m^{2n+2}$  for some constant  $C > 0$  which depends only on  $n$  and  $\text{diam } K$  and a Borel map  $p_m : K \rightarrow F_m$  such that

$$d(p_m(x), x) < 1/m$$

for all  $x \in K$ . Indeed choose  $x_1 \in K$ . For  $i \geq 2$ , choose by induction  $x_i \in K \setminus \bigcup_{j < i} B(x_j, 1/m)$  as long as  $K \setminus \bigcup_{j < i} B(x_j, 1/m) \neq \emptyset$ . Let  $F_m$  denote the set of all these points. The balls  $B(x_i, 1/(2m))$  are mutually disjoint and  $\bigcup_i B(x_i, 1/(2m)) \subset B(x_1, \text{diam } K + 1)$ . Remembering (3.2), it follows that

$$\begin{aligned} c_n(2m)^{-2n-2} \text{card } F &\leq \mathcal{L}^{2n+1} \left( \bigcup_i B(x_i, 1/(2m)) \right) \\ &\leq \mathcal{L}^{2n+1} (B(x_1, \text{diam } K + 1)) = c_n(\text{diam } K + 1)^{2n+2} \end{aligned}$$

for any finite subset  $F \subset F_m$ , hence  $F_m$  is a finite set with  $\text{card } F_m \leq C m^{2n+2}$  where  $C$  depends only on  $n$  and  $\text{diam } K$ . Next, by construction, for any  $x \in K$ , there exists a unique  $x_i \in F_m$  such that  $x \in B(x_i, 1/m) \setminus \bigcup_{j < i} B(x_j, 1/m)$  and we then set  $p_m(x) := x_i$ .

The proof of the lemma can now be completed following the same arguments as those in [24]. For sake of completeness, we sketch these arguments below. Let  $\gamma_{\varepsilon_k}$  and  $\gamma$  be as in the statement. For  $m \geq 1$ , one sets  $\nu_m := (p_m)_\# \nu$ . Note that by construction of  $F_m$  and  $p_m$ , one has  $\text{card}(\text{spt } \nu_m) \leq C m^{2n+2}$  and  $W_1(\nu_m, \nu) \leq 1/m$ . To check that  $\gamma \in \Pi(\mu, \nu)$ , one takes some  $\gamma_m \in \Pi(\mu, \nu_m) \subset \Pi$  and uses the optimality of  $\gamma_{\varepsilon_k}$  which implies

$$\begin{aligned} W_1((\pi_2)_\# \gamma_{\varepsilon_k}, \nu) &\leq \varepsilon_k C_{\varepsilon_k}(\gamma_m) \\ &\leq \frac{1}{m} + \varepsilon_k \text{diam}(K) + \varepsilon_k^2 \text{diam}(K)^2 + C \varepsilon_k^{6n+9} m^{2n+2}. \end{aligned}$$

Then one lets  $\varepsilon_k \rightarrow 0$  with  $m \geq 1$  fixed and then  $m \rightarrow +\infty$  to get that  $(\pi_2)_\# \gamma_{\varepsilon_k}$  converges weakly to  $\nu$ . Since it also converges weakly to  $(\pi_2)_\# \gamma$ , it follows that  $\gamma \in \Pi(\mu, \nu)$ .

To check that  $\gamma \in \Pi_1(\mu, \nu)$ , one takes  $\bar{\gamma} \in \Pi_1(\mu, \nu)$  and sets  $\bar{\gamma}_m := (\text{Id}, p_m)_\# \bar{\gamma} \in \Pi(\mu, \nu_m)$  where  $(\text{Id}, p_m)(x, y) = (x, p_m(y))$ . By optimality of  $\gamma_{\varepsilon_k}$ , one has

$$\begin{aligned} \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\gamma_{\varepsilon_k}(x, y) &\leq C_{\varepsilon_k}(\bar{\gamma}_m) \\ &\leq \frac{1}{m \varepsilon_k} + \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\bar{\gamma}_m(x, y) + \varepsilon_k \text{diam}(K)^2 + C \varepsilon_k^{6n+8} m^{2n+2}. \end{aligned}$$

Choosing  $m$  of the order of  $\varepsilon_k^{-2}$  and letting  $\varepsilon_k \rightarrow 0$ , one gets

$$\int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\gamma(x, y) \leq \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\bar{\gamma}(x, y)$$

hence  $\gamma \in \Pi_1(\mu, \nu)$ .

Finally, to check that  $\gamma \in \Pi_2(\mu, \nu)$ , one uses once again the optimality of  $\gamma_{\varepsilon_k}$  in the following way,

$$\begin{aligned} W_1((\pi_2)_\# \gamma_{\varepsilon_k}, \nu) + \varepsilon_k \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\gamma_{\varepsilon_k}(x, y) \\ + \varepsilon_k^2 \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\gamma_{\varepsilon_k}(x, y) \leq \varepsilon_k C_{\varepsilon_k}(\bar{\gamma}_m). \end{aligned}$$

On the other hand, one has

$$\begin{aligned} W_1(\mu, \nu) &\leq W_1(\mu, (\pi_2)_\# \gamma_{\varepsilon_k}) + W_1((\pi_2)_\# \gamma_{\varepsilon_k}, \nu) \\ &\leq \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\gamma_{\varepsilon_k}(x, y) + W_1((\pi_2)_\# \gamma_{\varepsilon_k}, \nu) \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\bar{\gamma}_m(x, y) &= \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, p_m(y)) d\bar{\gamma}(x, y) \\ &\leq \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y) d\bar{\gamma}(x, y) + \int_{\mathbb{H}^n \times \mathbb{H}^n} d(y, p_m(y)) d\bar{\gamma}(x, y) \\ &\leq W_1(\mu, \nu) + \frac{1}{m}. \end{aligned}$$

It follows that

$$\begin{aligned} (1 - \varepsilon_k) W_1((\pi_2)_\# \gamma_{\varepsilon_k}, \nu) + \varepsilon_k W_1(\mu, \nu) + \varepsilon_k^2 \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\gamma_{\varepsilon_k} \\ \leq \frac{1}{m} + \frac{\varepsilon_k}{m} + \varepsilon_k W_1(\mu, \nu) + \varepsilon_k^2 \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\bar{\gamma}_m + C \varepsilon_k^{6n+9} m^{2n+2} \end{aligned}$$

and hence

$$\begin{aligned} \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\gamma_{\varepsilon_k}(x, y) \\ \leq \frac{1}{m \varepsilon_k^2} + \frac{1}{m \varepsilon_k} + \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\bar{\gamma}_m(x, y) + C \varepsilon_k^{6n+7} m^{2n+2} \end{aligned}$$

provided  $\varepsilon_k \leq 1$ . Choosing  $m$  of the order of  $\varepsilon_k^{-3}$  and letting  $\varepsilon_k \rightarrow 0$ , one gets

$$\int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\gamma(x, y) \leq \int_{\mathbb{H}^n \times \mathbb{H}^n} d(x, y)^2 d\bar{\gamma}(x, y).$$

Since  $\bar{\gamma} \in \Pi_1(\mu, \nu)$  was arbitrary, it follows that  $\gamma \in \Pi_2(\mu, \nu)$ .  $\square$

The rest of this section is devoted to the study of the solutions to  $(P_\varepsilon)$ . We fix  $\varepsilon > 0$  and set

$$c_\varepsilon(x, y) := d(x, y) + \varepsilon d(x, y)^2.$$

We first recall the following classical fact.

**Lemma 5.3.** *Let  $\gamma_\varepsilon$  be a solution to  $(P_\varepsilon)$ . Then for any Borel set  $U \subset \mathbb{H}^n \times \mathbb{H}^n$ ,  $(\pi_2)_\#(\gamma_\varepsilon|_U)$  is finitely atomic and  $\gamma_\varepsilon|_U$  is a solution to Kantorovich transport problem (2.1) between  $(\pi_1)_\#(\gamma_\varepsilon|_U)$  and  $(\pi_2)_\#(\gamma_\varepsilon|_U)$  with cost  $c_\varepsilon$ .*

*Proof.* The fact that  $(\pi_2)_\#(\gamma_\varepsilon|_U)$  is finitely atomic obviously follows from the fact that  $C_\varepsilon(\gamma_\varepsilon) = \min\{C_\varepsilon(\gamma); \gamma \in \Pi\} < +\infty$ . Next it is also immediate that  $\gamma_\varepsilon$  is a solution to Kantorovich transport problem (2.1) between  $\mu$  and  $(\pi_2)_\#(\gamma_\varepsilon)$  with cost  $c_\varepsilon$ . Then as a classical fact, the claim follows from the linearity of the functional to be minimized with respect to the transport plan. If  $\gamma \in \Pi((\pi_1)_\#(\gamma_\varepsilon|_U), (\pi_2)_\#(\gamma_\varepsilon|_U))$ , one indeed simply compares  $C_\varepsilon(\gamma_\varepsilon)$  with  $C_\varepsilon(\hat{\gamma})$  where  $\hat{\gamma} = \gamma_\varepsilon|_{(\mathbb{H}^n \times \mathbb{H}^n) \setminus U} + \gamma \in \Pi(\mu, (\pi_2)_\#(\gamma_\varepsilon))$  to get the conclusion.  $\square$

Next in this section we consider interpolations between two measures  $\bar{\mu}, \bar{\nu} \in \mathcal{P}_c(\mathbb{H}^n)$  that are constructed from a transport plan solution to Kantorovich transport problem (2.1) between these two measures with cost  $c_\varepsilon$ . We prove absolute continuity and, more importantly,  $L^\infty$ -estimates on the density with respect to  $\mathcal{L}^{2n+1}$  of these interpolations whenever  $\bar{\mu} \ll \mathcal{L}^{2n+1}$  and  $\bar{\nu}$  is finitely atomic, see Proposition 5.7. We divide the arguments into several steps. First we prove that any solution to this Kantorovich transport problem is induced by a transport.

**Theorem 5.4.** *Let  $\bar{\mu}, \bar{\nu} \in \mathcal{P}_c(\mathbb{H}^n)$  be fixed. Assume that  $\bar{\mu} \ll \mathcal{L}^{2n+1}$  and that  $\bar{\nu}$  is finitely atomic. Then any solution to Kantorovich transport problem (2.1) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$  is induced by a transport. In particular there exists a unique optimal transport map solution to the transport problem (2.3) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$ .*

*Proof.* Let  $\psi$  be a Kantorovich potential for Kantorovich transport problem (2.1) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$  given by Theorem 2.2. Let  $\{y_i\}_{i=1}^k$  denote the

atoms of  $\bar{\nu}$ . We prove that for  $\mathcal{L}^{2n+1}$ -a.e.  $x \in \mathbb{H}^n$ , there is at most one point  $y_i$  for some  $i \in \{1, \dots, k\}$  such that

$$\psi(x) + \psi^c(y_i) = c_\varepsilon(x, y_i).$$

Since  $\bar{\mu} \ll \mathcal{L}^{2n+1}$ , it will follow that any transport plan solution to Kantorovich transport problem (2.1) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$  is concentrated on a  $\bar{\mu}$ -measurable graph and hence induced by a transport. This implies in turn existence and uniqueness of the optimal transport map solution to the transport problem (2.3) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$  (see Theorem 2.6).

For  $i \neq j$ , set  $h_{ij}(x) := c_\varepsilon(x, y_i) - c_\varepsilon(x, y_j) + \psi^c(y_j) - \psi^c(y_i)$ . It follows from Lemma 3.4 that  $h_{ij}$  is of class  $C^\infty$  on the open set  $\mathbb{H}^n \setminus (L_{y_i} \cup L_{y_j})$  with  $\nabla h_{ij} \neq 0$ . Indeed assume on the contrary that  $\nabla h_{ij}(x) = 0$  for some  $x \in \mathbb{H}^n \setminus (L_{y_i} \cup L_{y_j})$ . Then, differentiating along the horizontal vector fields  $X_j$  and  $Y_j$ , we would have

$$\nabla_H d_{y_i}(x) (1 + 2\varepsilon d_{y_i}(x)) = \nabla_H d_{y_j}(x) (1 + 2\varepsilon d_{y_j}(x)).$$

Since  $|\nabla_H d_{y_i}(x)| = |\nabla_H d_{y_j}(x)|$  (see Lemma 3.4(i)), this would imply that  $d_{y_i}(x) = d_{y_j}(x)$  and in turn that  $\nabla_H d_{y_i}(x) = \nabla_H d_{y_j}(x)$ . Since we also have by assumption  $\partial_t d_{y_i}(x) = \partial_t d_{y_j}(x)$ , Lemma 3.4(ii) would give  $y_i = y_j$ . It follows that the set  $\{x \in \mathbb{H}^n \setminus (L_{y_i} \cup L_{y_j}); h_{ij}(x) = 0\}$  is a  $C^\infty$ -smooth submanifold of dimension  $2n$  in  $\mathbb{R}^{2n+1}$  and hence has Lebesgue measure 0. Since  $\mathcal{L}^{2n+1}(L_{y_i}) = 0$ , it follows that

$$\mathcal{L}^{2n+1}\left(\bigcup_{i \neq j} \{x \in \mathbb{H}^n; h_{ij}(x) = 0\}\right) = 0$$

and the claim follows.  $\square$

If  $T : \mathbb{H}^n \rightarrow \mathbb{H}^n$ , we set  $T_t = e_t \circ S \circ (I \otimes T)$ , i.e.,  $T_t(x)$  is the point lying at distance  $t d(x, T(x))$  from  $x$  on the selected minimal curve  $S(x, T(x))$  between  $x$  and  $T(x)$  (see Subsection 3.3 for the definition of  $S$  and  $e_t$ ).

**Proposition 5.5.** [27, Chapter 7] *Let  $\bar{\mu}, \bar{\nu} \in \mathcal{P}_c(\mathbb{H}^n)$  be fixed such that  $\bar{\mu} \ll \mathcal{L}^{2n+1}$  and  $\bar{\nu}$  is finitely atomic. Let  $T^\varepsilon$  be the optimal transport map solution to the transport problem (2.3) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$ . Then there exists a  $\bar{\mu}$ -measurable set  $A$  such that  $\bar{\mu}(A) = 1$  and such that for each  $t \in [0, 1)$ ,  $T_t^\varepsilon|_A$  is injective.*

The cost  $c_\varepsilon$  can be recovered as coming from a so-called coercive Lagrangian action. Since  $(\mathbb{H}^n, d)$  is non-branching, the proposition essentially follows from

[27, Chapter 7, Theorem 7.30]. However one does not need the full strength of the theory developed in [27, Chapter 7] to get the conclusion of Proposition 5.5 and we sketch below the arguments for the reader's convenience.

*Proof.* Let  $0 \leq s < t \leq 1$  and  $x, y \in \mathbb{H}^n$ . Set

$$c_\varepsilon^{s,t}(x, y) = d(x, y) + \varepsilon \frac{d(x, y)^2}{t - s}.$$

The space  $(\mathbb{H}^n, d)$  being a geodesic space and  $u \mapsto u + \varepsilon u^2$  being strictly increasing and strictly convex on  $[0, +\infty)$ , one has

$$c_\varepsilon(x, y) \leq c_\varepsilon^{0,t}(x, z) + c_\varepsilon^{t,1}(z, y)$$

for all  $x, y, z \in \mathbb{H}^n$  and  $t \in (0, 1)$ , with equality if and only if any curve in  $C([0, 1], \mathbb{H}^n)$  obtained by concatenation of a minimal curve  $\sigma_{x,z} : [0, t] \rightarrow \mathbb{H}^n$  between  $x$  and  $z$  and a minimal curve  $\sigma_{z,y} : [t, 1] \rightarrow \mathbb{H}^n$  between  $z$  and  $y$  is a minimal curve between  $x$  and  $y$ .

On the other hand, by  $c_\varepsilon$ -cyclical monotonicity, one knows that there exists a  $\bar{\mu}$ -measurable set  $A$  such that  $\bar{\mu}(A) = 1$  and

$$c_\varepsilon(x, T^\varepsilon(x)) + c_\varepsilon(\tilde{x}, T^\varepsilon(\tilde{x})) \leq c_\varepsilon(x, T^\varepsilon(\tilde{x})) + c_\varepsilon(\tilde{x}, T^\varepsilon(x))$$

for all  $x, \tilde{x} \in A$  (see Theorem 2.1).

Now let  $t \in (0, 1)$  be fixed and let  $x, \tilde{x} \in A$ . Assume that  $T_t^\varepsilon(x) = T_t^\varepsilon(\tilde{x})$ . Then

$$c_\varepsilon(x, T^\varepsilon(\tilde{x})) \leq c_\varepsilon^{0,t}(x, z) + c_\varepsilon^{t,1}(z, T^\varepsilon(\tilde{x})), \quad (5.1)$$

and similarly,

$$c_\varepsilon(\tilde{x}, T^\varepsilon(x)) \leq c_\varepsilon^{0,t}(\tilde{x}, z) + c_\varepsilon^{t,1}(z, T^\varepsilon(x))$$

where  $z = T_t^\varepsilon(x) = T_t^\varepsilon(\tilde{x})$ . It follows that

$$\begin{aligned} & c_\varepsilon(x, T^\varepsilon(x)) + c_\varepsilon(\tilde{x}, T^\varepsilon(\tilde{x})) \\ & \leq c_\varepsilon(x, T^\varepsilon(\tilde{x})) + c_\varepsilon(\tilde{x}, T^\varepsilon(x)) \\ & \leq c_\varepsilon^{0,t}(x, T_t^\varepsilon(x)) + c_\varepsilon^{t,1}(T_t^\varepsilon(x), T^\varepsilon(x)) + c_\varepsilon^{0,t}(\tilde{x}, T_t^\varepsilon(\tilde{x})) + c_\varepsilon^{t,1}(T_t^\varepsilon(\tilde{x}), T^\varepsilon(\tilde{x})) \\ & = c_\varepsilon(x, T^\varepsilon(x)) + c_\varepsilon(\tilde{x}, T^\varepsilon(\tilde{x})). \end{aligned}$$

Hence equality has to hold in all these inequalities. In particular equality holds in (5.1). It follows that the curve obtained by concatenation of the minimal curve  $s \in [0, t] \mapsto e_s(S(x, T^\varepsilon(x)))$  between  $x$  and  $z$  with the minimal curve  $s \in [t, 1] \mapsto e_s(S(\tilde{x}, T^\varepsilon(\tilde{x})))$  between  $z$  and  $T^\varepsilon(\tilde{x})$  is a minimal curve. Since this curve coincides with the minimal curve  $\sigma : s \in [0, 1] \mapsto e_s(S(x, T^\varepsilon(x)))$  on the non-trivial

interval  $[0, t]$  and  $\mathbb{H}^n$  is non-branching (see Proposition 3.3), we get that it coincides with  $\sigma$  on the whole interval  $[0, 1]$ . Similarly, it coincides with the minimal curve  $\tilde{\sigma} : s \in [0, 1] \mapsto e_s(S(\tilde{x}, T^s(\tilde{x})))$  on the whole interval  $[0, 1]$ . Hence  $\sigma = \tilde{\sigma}$  and in particular  $x = \sigma(0) = \tilde{\sigma}(0) = \tilde{x}$ .  $\square$

We turn now to the main estimate that will lead to Proposition 5.7.

**Proposition 5.6.** *Let  $\bar{\mu} \in \mathcal{P}(\mathbb{H}^n)$ ,  $\bar{\mu} \ll \mathcal{L}^{2n+1}$ ,  $\bar{\mu} = \rho \mathcal{L}^{2n+1}$ , and  $T : \mathbb{H}^n \rightarrow \mathbb{H}^n$  be a  $\bar{\mu}$ -measurable map such that  $T_{\#}\bar{\mu}$  is finitely atomic. Let  $t \in (0, 1)$  and set  $\bar{\mu}_t := T_{t\#}\bar{\mu}$ . Assume that there exists a  $\bar{\mu}$ -measurable set  $A$  such that  $\bar{\mu}(A) = 1$  and  $T_t|_A$  is injective. Then  $\bar{\mu}_t \ll \mathcal{L}^{2n+1}$ ,  $\bar{\mu}_t = \rho_t \mathcal{L}^{2n+1}$  with*

$$\rho_t \leq \frac{\mathbb{1}_{T_t(A)}}{(1-t)^{2n+3}} \rho \circ T_t^{-1}|_{T_t(A)} \quad \mathcal{L}^{2n+1}\text{-a.e.}$$

Arguments for the proof of this proposition can be found in [16, Section 3] even-though not explicitly stated in the same way in that paper. They rely on the following estimate:

$$\mathcal{L}^{2n+1}(E) \leq \frac{1}{(1-t)^{2n+3}} \mathcal{L}^{2n+1}((e_t \circ S)(E, y))$$

for any  $y \in \mathbb{H}^n$  and  $E \subset \mathbb{H}^n$  which is proved in [20, Section 2] and which roughly means that  $(\mathbb{H}^n, d, \mathcal{L}^{2n+1})$  satisfies a so-called Measure Contraction Property. We detail the proof below for the reader's convenience.

*Proof.* Let  $\{y_i\}_{i=1}^k$  denote the atoms of  $T_{\#}\bar{\mu}$ . Set  $A_i = T^{-1}(\{y_i\}) \cap A$  and  $\hat{A} = \bigcup_i A_i$ . The sets  $A_i$  are mutually disjoint and  $\bar{\mu}(\hat{A}) = 1$  by hypothesis. For any  $x \in A_i$ ,  $T_t(x) = (e_t \circ S)(x, y_i)$ , hence

$$\mathcal{L}^{2n+1}(E) \leq \frac{1}{(1-t)^{2n+3}} \mathcal{L}^{2n+1}(T_t(E))$$

for any  $E \subset A_i$ . Next if  $E \subset \hat{A}$ , writing  $E = \bigcup_i (E \cap A_i)$  where the sets  $A_i$  are mutually disjoint and remembering that  $T_t$  is injective on  $\hat{A} \subset A$  by hypothesis, one gets

$$\mathcal{L}^{2n+1}(E) \leq \frac{1}{(1-t)^{2n+3}} \mathcal{L}^{2n+1}(T_t(E)).$$

It follows that for any  $F \subset \mathbb{H}^n$ ,

$$\mathcal{L}^{2n+1}(T_t^{-1}(F) \cap \hat{A}) \leq \frac{1}{(1-t)^{2n+3}} \mathcal{L}^{2n+1}(F \cap T_t(\hat{A})).$$

Assume that  $F \subset \mathbb{H}^n$  is such that  $\mathcal{L}^{2n+1}(F) = 0$ . We get  $\mathcal{L}^{2n+1}(T_t^{-1}(F) \cap \hat{A}) = 0$  from the previous inequality. On the other hand  $\bar{\mu}_t(F) = \bar{\mu}(T_t^{-1}(F)) = \bar{\mu}(T_t^{-1}(F) \cap \hat{A})$  hence  $\bar{\mu}_t(F) = 0$  since  $\bar{\mu} \ll \mathcal{L}^{2n+1}$  and it follows that  $\bar{\mu}_t \ll \mathcal{L}^{2n+1}$ .

Next, to prove the estimate on the density of  $\bar{\mu}_t$  with respect to  $\mathcal{L}^{2n+1}$ , we note that the inequality above implies that

$$\int_{\hat{A}} g(T_t) d\mathcal{L}^{2n+1} \leq \frac{1}{(1-t)^{2n+3}} \int_{T_t(\hat{A})} g d\mathcal{L}^{2n+1}$$

for any non-negative measurable map  $g : \mathbb{H}^n \rightarrow [0, +\infty]$ . Let  $h : \mathbb{H}^n \rightarrow [0, +\infty]$  be a non-negative measurable map and set

$$g(x) = \mathbb{1}_{T_t(\hat{A})}(x) h(x) \rho(T_t^{-1} \lfloor_{T_t(\hat{A})}(x)).$$

Then

$$\begin{aligned} \int_{\hat{A}} g(T_t(x)) d\mathcal{L}^{2n+1}(x) \\ \leq \frac{1}{(1-t)^{2n+3}} \int_{T_t(\hat{A})} h(x) \rho(T_t^{-1} \lfloor_{T_t(\hat{A})}(x)) d\mathcal{L}^{2n+1}(x). \end{aligned}$$

On the other hand

$$\int_{\hat{A}} g(T_t) d\mathcal{L}^{2n+1} = \int_{\hat{A}} h(T_t) \rho d\mathcal{L}^{2n+1} = \int_{\mathbb{H}^n} h(T_t) d\bar{\mu} = \int_{\mathbb{H}^n} h d\bar{\mu}_t,$$

hence

$$\int_{\mathbb{H}^n} h(x) d\bar{\mu}_t(x) \leq \frac{1}{(1-t)^{2n+3}} \int_{T_t(\hat{A})} h(x) \rho(T_t^{-1} \lfloor_{T_t(\hat{A})}(x)) d\mathcal{L}^{2n+1}(x).$$

Remembering that  $\hat{A} \subset A$ , this concludes the proof.  $\square$

Finally, combining Theorem 5.4, Propositions 5.5 and 5.6, we get the following proposition which gives the absolute continuity of the interpolations together with an  $L^\infty$ -estimate on their density. Note that if  $\gamma_\varepsilon$  is the transport plan solution to Kantorovich transport problem (2.1) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$  and  $T^\varepsilon$  the optimal transport map solution to the transport problem (2.3) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$ , which hence induces  $\gamma_\varepsilon$ , then  $(e_t \circ S)_\# \gamma_\varepsilon = T_t^\varepsilon \# \bar{\mu}$ .

**Proposition 5.7.** *Let  $\bar{\mu}, \bar{\nu} \in \mathcal{P}_c(\mathbb{H}^n)$  be fixed. Assume that  $\bar{\mu} \ll \mathcal{L}^{2n+1}$  with  $\bar{\mu} = \rho \mathcal{L}^{2n+1}$  and  $\bar{\nu}$  is finitely atomic. Let  $\gamma_\varepsilon$  be the transport plan solution*

to Kantorovich transport problem (2.1) between  $\bar{\mu}$  and  $\bar{\nu}$  with cost  $c_\varepsilon$ . Then for any  $t \in [0, 1)$ , the interpolation  $(e_t \circ S)_\# \gamma_\varepsilon$  is absolutely continuous with respect to  $\mathcal{L}^{2n+1}$ ,  $(e_t \circ S)_\# \gamma_\varepsilon = \rho_t^\varepsilon \mathcal{L}^{2n+1}$ , and one has

$$\|\rho_t^\varepsilon\|_{L^\infty} \leq \frac{1}{(1-t)^{2n+3}} \|\rho\|_{L^\infty}.$$

## 6 Properties of measures $\gamma \in \mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$ with $(\pi_1)_\# \gamma \ll \mathcal{L}^{2n+1}$

This section is independent of the transport problem. We state some properties of measures in  $\mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$  with first marginal absolutely continuous with respect to  $\mathcal{L}^{2n+1}$ . These properties are essential steps in the strategy adopted here to solve Monge's transport problem. They are the exact counterpart in the framework of  $(\mathbb{H}^n, d, \mathcal{L}^{2n+1})$  of similar properties proved in [13] in  $\mathbb{R}^n$ . These properties hold actually true in more general settings, for instance in any separable doubling metric measure space.

We first recall some facts about Lebesgue points of Borel functions and density of absolutely continuous measures. Since the measure  $\mathcal{L}^{2n+1}$  is a doubling measure on  $(\mathbb{H}^n, d)$ , see (3.2), if  $\rho : \mathbb{H}^n \rightarrow [0, +\infty]$  is a  $\mathcal{L}^{2n+1}$ -locally summable Borel function then for  $\mathcal{L}^{2n+1}$ -a.e.  $x \in \mathbb{H}^n$ , one has

$$\lim_{r \rightarrow 0} \frac{1}{\mathcal{L}^{2n+1}(B(x, r))} \int_{B(x, r)} |\rho(y) - \rho(x)| d\mathcal{L}^{2n+1}(y) = 0, \quad (6.1)$$

see e.g. [19]. A point  $x \in \mathbb{H}^n$  where (6.1) holds is called a Lebesgue point of  $\rho$  and we denote by  $\text{Leb } \rho$  the set of all Lebesgue points of  $\rho$ .

In the rest of this paper, and especially in the lemma to follow, it will be technically convenient to consider the density  $\rho$  of an absolutely continuous measure  $\mu \ll \mathcal{L}^{2n+1}$  as a  $\mathcal{L}^{2n+1}$ -summable Borel function, i.e., a function well-defined everywhere, so that one can speak about its Lebesgue points and its value at any arbitrary point without any ambiguity. If  $\mu \in \mathcal{P}(\mathbb{H}^n)$ , we set

$$\rho(x) := \limsup_{r \rightarrow 0} \frac{\mu(B(x, r))}{\mathcal{L}^{2n+1}(B(x, r))}. \quad (6.2)$$

This map  $\rho : \mathbb{H}^n \rightarrow [0, +\infty]$  is a Borel map and if  $\mu \ll \mathcal{L}^{2n+1}$  then  $\mu = \rho \mathcal{L}^{2n+1}$ . By a slight abuse of terminology, when speaking about the density of an absolutely continuous measure  $\mu \in \mathcal{P}(\mathbb{H}^n)$  with respect to  $\mathcal{L}^{2n+1}$ , we will thus always refer in the following to the Borel function  $\rho$  defined above.

The next lemma will be an essential ingredient in the proof of Lemma 7.1.



**Lemma 6.1.** *Let  $\gamma \in \mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$  be such that  $(\pi_1)_\# \gamma \ll \mathcal{L}^{2n+1}$ . Then  $\gamma$  is concentrated on a set  $\Gamma$  such that, for all  $(x, y) \in \Gamma$  and all  $r > 0$ , there exist  $y' \in \mathbb{H}^n$  and  $r' > 0$  such that*

- (i)  $y \in B(y', r') \subset \subset B(y, r)$ ,
- (ii)  $x \in \text{Leb } \rho$  and  $\rho(x) < +\infty$ ,
- (iii)  $x \in \text{Leb } \rho'$  and  $\rho'(x) > 0$ ,

where  $\rho$  denotes the density of  $(\pi_1)_\# \gamma$  and  $\rho'$  the density of  $(\pi_1)_\# \gamma|_{(\mathbb{H}^n \times B(y', r'))}$  with respect to  $\mathcal{L}^{2n+1}$ .

*Proof.* Let  $(y_m)_{m \geq 1}$  be a dense sequence in  $\mathbb{H}^n$ . For each  $m, k \in \mathbb{N}^*$ , set  $\gamma_{m,k} := \gamma|_{(\mathbb{H}^n \times B(y_m, r_k))}$  where  $r_k := 1/k$ . Let  $\rho_{m,k}$  denote the density of  $(\pi_1)_\# \gamma_{m,k}$  with respect to  $\mathcal{L}^{2n+1}$ . Set  $A_{m,k} := \mathbb{H}^n \setminus (\text{Leb } \rho \cap \text{Leb } \rho_{m,k} \cap \{\rho < +\infty\})$ . We have  $\mathcal{L}^{2n+1}(A_{m,k}) = 0$ . Since  $(\pi_1)_\# \gamma \ll \mathcal{L}^{2n+1}$ , it follows that  $\gamma(A_{m,k} \times B(y_m, r_k)) \leq (\pi_1)_\# \gamma(A_{m,k}) = 0$ . Next

$$\gamma(\{\rho_{m,k} = 0\} \times B(y_m, r_k)) = (\pi_1)_\# \gamma_{m,k}(\{\rho_{m,k} = 0\}) = 0.$$

It follows that  $\gamma(D_{m,k}) = 0$  for all  $m, k \in \mathbb{N}^*$  where

$$D_{m,k} := [\mathbb{H}^n \setminus (\text{Leb } \rho \cap \text{Leb } \rho_{m,k} \cap \{\rho < +\infty\} \cap \{\rho_{m,k} > 0\})] \times B(y_m, r_k)$$

hence  $\gamma(\bigcup_{m,k} D_{m,k}) = 0$  and  $\gamma$  is concentrated on  $\mathbb{H}^n \setminus \bigcup_{m,k} D_{m,k}$ . Then the conclusion follows noting that for each  $(x, y) \in \mathbb{H}^n \times \mathbb{H}^n$  and  $r > 0$ , one can find  $m, k \in \mathbb{N}^*$  such that  $y \in B(y_m, r_k) \subset \subset B(y, r)$ .  $\square$

We say that  $x \in E$  is a Lebesgue point of a Borel set  $E$  if  $x \in \text{Leb } \mathbb{I}_E$ , i.e., if  $x \in E$  and

$$\lim_{r \rightarrow 0} \frac{\mathcal{L}^{2n+1}(E \cap B(x, r))}{\mathcal{L}^{2n+1}(B(x, r))} = 1,$$

and we denote by  $\text{Leb } E := \text{Leb } \mathbb{I}_E$  the set of all Lebesgue points of  $E$ . Note that  $\mathcal{L}^{2n+1}(E \setminus \text{Leb } E) = 0$ .

The next lemma together with Lemma 7.1 and Lemma 4.3 is one of the key ingredients of the proof of Theorem 8.1 and eventually of the existence of a solution to Monge's transport problem. It can be recovered as a consequence of Lemma 6.1. However, for sake of clarity, we state and prove it independently.

**Lemma 6.2.** *Let  $\gamma \in \mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$  be such that  $(\pi_1)_\# \gamma \ll \mathcal{L}^{2n+1}$ . Assume that  $\gamma$  is concentrated on a  $\sigma$ -compact set  $\Gamma$ . For  $y \in \mathbb{H}^n$  and  $r > 0$ , set*

$$\Gamma^{-1}(B(y, r)) = \pi_1(\Gamma \cap (\mathbb{H}^n \times B(y, r))).$$

Then  $\Gamma^{-1}(B(y, r))$  is a Borel set and  $\gamma$  is concentrated on a set  $\Gamma' \subset \Gamma$  such that for all  $(x, y) \in \Gamma'$  and all  $r > 0$ ,  $x \in \text{Leb } \Gamma^{-1}(B(y, r))$ .

*Proof.* Since  $\Gamma$  is  $\sigma$ -compact,  $\Gamma^{-1}(B(y, r))$  is also  $\sigma$ -compact hence a Borel set. Set  $A := \{(x, y) \in \Gamma; x \notin \text{Leb } \Gamma^{-1}(B(y, r)) \text{ for some } r > 0\}$  and let us show that  $\gamma(A) = 0$ . For each  $k \in \mathbb{N}^*$ , consider a countable covering of  $\mathbb{H}^n$  by balls  $(B(y_i^k, r_k))_{i \geq 1}$  of radius  $r_k := 1/(2k)$ . If  $(x, y) \in \Gamma$  and  $x \notin \text{Leb } \Gamma^{-1}(B(y, r))$  then for any  $k \geq 1/r$  and  $y_i^k$  such that  $d(y_i^k, y) < r_k$ , one has  $x \in \Gamma^{-1}(B(y_i^k, r_k)) \setminus \text{Leb } \Gamma^{-1}(B(y_i^k, r_k))$ . It follows that

$$\pi_1(A) \subset \bigcup_{k \geq 1} \bigcup_{i \geq 1} \Gamma^{-1}(B(y_i^k, r_k)) \setminus \text{Leb } \Gamma^{-1}(B(y_i^k, r_k)).$$

The set on the right-hand side has  $\mathcal{L}^{2n+1}$ -measure 0. Since  $(\pi_1)_\# \gamma \ll \mathcal{L}^{2n+1}$ , it follows that  $\gamma(A) \leq (\pi_1)_\# \gamma(\pi_1(A)) = 0$ .  $\square$

## 7 Lower density of the transport set

We consider optimal transport plans in  $\Pi_1(\mu, \nu)$  that are obtained as weak limit of solutions to the variational approximations introduced in Section 5. We prove that if  $\gamma$  is such a transport plan then it is concentrated on a set  $\Gamma$  whose related transport set has positive lower density at each point  $x \in \pi_1(\Gamma)$  for some suitable notion of lower density. As already mentioned, this is one of the main ingredient in the proof of Theorem 8.1. Following the notion of transport set introduced in e.g. [4], we define in our geometrical context the transport set related to a set  $\Gamma \subset \mathbb{H}^n \times \mathbb{H}^n$  as

$$T(\Gamma) := \{(e_t \circ S)(x, y); (x, y) \in \Gamma, t \in (0, 1)\}.$$

Recall that  $S$  is a measurable selection of minimal curves and that  $(e_t \circ S)(x, y)$  denotes the point at distance  $t d(x, y)$  from  $x$  on the selected minimal curve  $S(x, y)$  between  $x$  and  $y$ , see Subsection 3.3.

**Lemma 7.1.** *Let  $\gamma \in \Pi(\mu, \nu)$  be obtained as a weak limit of solutions to  $(P_{\varepsilon_k})$  for some sequence  $\varepsilon_k$  converging to 0. Then  $\gamma$  is concentrated on a set  $\Gamma$  such that for all  $(x, y) \in \Gamma$  such that  $x \neq y$  and all  $r > 0$ , we have*

$$\liminf_{\delta \downarrow 0} \frac{\mathcal{L}^{2n+1}(T(\Gamma \cap [B(x, \frac{\delta}{2}) \times B(y, r)]) \cap B(x, \delta))}{\mathcal{L}^{2n+1}(B(x, \delta))} > 0.$$

The proof below follows the line of the proof of the similar property in [13]. In our context it requires however some technical refinement.

*Proof.* We consider the set  $\Gamma$  obtained by Lemma 6.1,  $(x, y) \in \Gamma$  with  $x \neq y$  and  $r > 0$ . Then let  $y' \in \mathbb{H}^n$  and  $r' > 0$  be given by Lemma 6.1 so that Lemma 6.1(i), (ii) and (iii) hold. Using the same notations as in this lemma, we set

$$G := \left\{ z \in \mathbb{H}^n; \frac{1}{2}\rho'(x) \leq \rho'(z) \text{ and } \rho(z) \leq 2\rho(x) \right\}.$$

Then  $G$  is a Borel set. We have  $0 < \rho'(x) \leq \rho(x)$  (remember the convention about densities of absolutely continuous measure, see (6.2)). Since  $x \in \text{Leb } \rho \cap \text{Leb } \rho'$ , see Lemma 6.1(ii) and (iii), it follows that  $x \in \text{Leb } G$ .

Fix  $\delta > 0$  such that  $\delta < d(x, y) + r$  and

$$\frac{1}{2}\mathcal{L}^{2n+1}(B(x, s)) \leq \mathcal{L}^{2n+1}(G \cap B(x, s)) \quad (7.1)$$

for all  $s \in (0, \delta)$  and fix  $t > 0$  such that  $4t(d(x, y) + r) < \delta$ .

We set  $G_\delta := G \cap B(x, \frac{\delta}{2})$ ,  $A_\delta := G_\delta \times B(y', r')$  and  $\gamma_\delta := \gamma|_{A_\delta}$ . We shall prove that

$$\frac{\rho'(x)}{4} \mathcal{L}^{2n+1}\left(B\left(x, \frac{\delta}{2}\right)\right) \leq (e_t \circ S)_\# \gamma_\delta(B(x, \delta)) \quad (7.2)$$

and

$$\begin{aligned} & (e_t \circ S)_\# \gamma_\delta(B(x, \delta)) \\ & \leq 2^{2n+4} \rho(x) \mathcal{L}^{2n+1}\left(T\left(\Gamma \cap \left[B\left(x, \frac{\delta}{2}\right) \times B(y, r)\right]\right) \cap B(x, \delta)\right). \end{aligned} \quad (7.3)$$

Then (7.2) and (7.3) will yield

$$\begin{aligned} & 2^{-(2n+6)} \frac{\rho'(x)}{\rho(x)} \mathcal{L}^{2n+1}\left(B\left(x, \frac{\delta}{2}\right)\right) \\ & \leq \mathcal{L}^{2n+1}\left(T\left(\Gamma \cap \left[B\left(x, \frac{\delta}{2}\right) \times B(y, r)\right]\right) \cap B(x, \delta)\right) \end{aligned}$$

for any  $\delta > 0$  small enough which completes the proof.

To prove (7.2), we note that  $(\pi_1)_\# \gamma_\delta \ll \mathcal{L}^{2n+1}$  with density bounded below by  $\frac{1}{2}\rho'(x) \mathcal{L}^{2n+1}$ -a.e. on  $G_\delta$ . Together with (7.1), it follows that

$$\frac{\rho'(x)}{4} \mathcal{L}^{2n+1}\left(B\left(x, \frac{\delta}{2}\right)\right) \leq (\pi_1)_\# \gamma_\delta\left(B\left(x, \frac{\delta}{2}\right)\right).$$

Next, by choice of  $\delta$  and  $t$ , we have  $(e_t \circ S)(z, w) \in B(x, \delta)$  for all  $z \in B(x, \frac{\delta}{2})$  and  $w \in B(y, r)$ , hence

$$B\left(x, \frac{\delta}{2}\right) \times B(y', r') \subset B\left(x, \frac{\delta}{2}\right) \times B(y, r) \subset (e_t \circ S)^{-1}(B(x, \delta))$$

and it follows that

$$(\pi_1)_\# \gamma_\delta \left( B\left(x, \frac{\delta}{2}\right) \right) = \gamma_\delta \left( B\left(x, \frac{\delta}{2}\right) \times B(y', r') \right) \leq (e_t \circ S)_\# \gamma_\delta (B(x, \delta))$$

and this completes the proof of (7.2).

We prove now (7.3). By hypothesis,  $\gamma$  is a weak limit of solutions  $\gamma_k$  to  $(P_{\varepsilon_k})$  for some sequence  $\varepsilon_k$  converging to 0. For each fixed  $k \in \mathbb{N}$ , we apply Lemma 5.3 with  $U = G_\delta \times \mathbb{H}^n$  and Proposition 5.7 with  $\bar{\mu} = (\pi_1)_\#(\gamma_k|_U)$  and  $\bar{\nu} = (\pi_2)_\#(\gamma_k|_U)$ . Taking into account the fact that  $(\pi_1)_\#(\gamma_k|_U) = \mu|_{G_\delta}$ , we get that  $(e_t \circ S)_\#(\gamma_k|_{G_\delta \times \mathbb{H}^n}) \ll \mathcal{L}^{2n+1}$  with density in  $L^\infty$  and whose  $L^\infty$ -norm is bounded by

$$\frac{1}{(1-t)^{2n+3}} \|\rho|_{G_\delta}\|_{L^\infty} \leq 2^{2n+4} \rho(x). \quad (7.4)$$

Next we check that  $(e_t \circ S)_\#(\gamma_k|_{G_\delta \times \mathbb{H}^n})$  converges weakly to  $(e_t \circ S)_\#(\gamma|_{G_\delta \times \mathbb{H}^n})$ . First it follows from Lemma 7.2 (to be proved below) that  $\gamma_k|_{G_\delta \times \mathbb{H}^n}$  converges weakly to  $\gamma|_{G_\delta \times \mathbb{H}^n}$ . Then, noting that  $\gamma$  and each  $\gamma_k$  are concentrated on  $\Omega$  and that  $e_t \circ S$  is continuous on  $\Omega$ , the claim follows from Lemma 7.3 (to be proved below) applied with  $\bar{\gamma} = \gamma|_{G_\delta \times \mathbb{H}^n}$ ,  $\bar{\gamma}_k = \gamma_k|_{G_\delta \times \mathbb{H}^n}$ ,  $B = \Omega$  and  $f = \varphi \circ e_t \circ S$  where  $\varphi \in C_b(\mathbb{H}^n)$ . The fact that  $\gamma$  is concentrated on  $\Omega$  follows from Lemma 4.1. To check that  $\gamma_k$  is concentrated on  $\Omega$ , denote by  $\{y_i^k\}_i$  the finite set of the atoms of  $(\pi_2)_\# \gamma_k$ . We have that  $\gamma_k$  is concentrated on  $\mathbb{H}^n \times \{y_i^k\}_i$ . On the other hand  $\gamma_k(L_{y_i^k} \times \{y_i^k\}) \leq \gamma_k(L_{y_i^k} \times \mathbb{H}^n) = \mu(L_{y_i^k}) = 0$  since  $\mu \ll \mathcal{L}^{2n+1}$ . It follows that  $\gamma_k$  is concentrated on  $\bigcup_i [(\mathbb{H}^n \setminus L_{y_i^k}) \times \{y_i^k\}] \subset \Omega$ . Then, taking into account (7.4), we get

$$\left| \int_{\mathbb{H}^n} \varphi \, d(e_t \circ S)_\#(\gamma|_{G_\delta \times \mathbb{H}^n}) \right| \leq 2^{2n+4} \rho(x) \|\varphi\|_{L^1}$$

for every  $\varphi \in C_b(\mathbb{H}^n)$ . It follows that  $(e_t \circ S)_\#(\gamma|_{G_\delta \times \mathbb{H}^n})$  is in  $(L^1)'$  with density in  $L^\infty$  and whose  $L^\infty$ -norm is bounded by  $2^{2n+4} \rho(x)$ . Since  $(e_t \circ S)_\# \gamma_\delta \leq (e_t \circ S)_\#(\gamma|_{G_\delta \times \mathbb{H}^n})$ , the same holds true for  $(e_t \circ S)_\# \gamma_\delta$ . Finally we note that  $\gamma_\delta$  being concentrated on  $\Gamma \cap [B(x, \frac{\delta}{2}) \times B(y', r')] \subset \Gamma \cap [B(x, \frac{\delta}{2}) \times B(y, r)]$ , the measure  $(e_t \circ S)_\# \gamma_\delta$  is concentrated on  $T(\Gamma \cap [B(x, \frac{\delta}{2}) \times B(y', r')]) \subset T(\Gamma \cap [B(x, \frac{\delta}{2}) \times B(y, r)])$ . All together we get

$$\begin{aligned} & (e_t \circ S)_\# \gamma_\delta (B(x, \delta)) \\ &= (e_t \circ S)_\# \gamma_\delta \left( T\left(\Gamma \cap \left[B\left(x, \frac{\delta}{2}\right) \times B(y, r)\right]\right) \cap B(x, \delta) \right) \\ &\leq 2^{2n+4} \rho(x) \mathcal{L}^{2n+1} \left( T\left(\Gamma \cap \left[B\left(x, \frac{\delta}{2}\right) \times B(y, r)\right]\right) \cap B(x, \delta) \right) \end{aligned}$$

which proves (7.3).  $\square$

**Lemma 7.2.** *Let  $X$  be a separable and locally compact Hausdorff metric space in which every open set is  $\sigma$ -compact. Let  $(\gamma_k)_k$  be a sequence in  $\mathcal{P}(X \times X)$  which converges weakly to some  $\gamma \in \mathcal{P}(X \times X)$  and such that  $(\pi_1)_\# \gamma_k = (\pi_1)_\# \gamma$  for every  $k \in \mathbb{N}$ . Then for any Borel set  $G \subset X$ , the sequence  $(\gamma_k|_{G \times X})_k$  converges weakly to  $\gamma|_{G \times X}$ .*

*Proof.* We have to prove that for any  $\varphi \in C_b(X)$ ,

$$\lim_{k \rightarrow +\infty} \int_{X \times X} \mathbb{1}_G(x) \varphi(x, y) d\gamma_k(x, y) = \int_{X \times X} \mathbb{1}_G(x) \varphi(x, y) d\gamma(x, y).$$

It follows from Lusin's theorem that for any  $\varepsilon > 0$  there exists a closed set  $F_\varepsilon$  such that  $\mathbb{1}_G|_{F_\varepsilon}$  is continuous and  $(\pi_1)_\# \gamma(X \setminus F_\varepsilon) < \varepsilon$ . As a consequence, for every  $\varepsilon > 0$ , the restriction of  $(x, y) \mapsto \mathbb{1}_G(x) \varphi(x, y)$  to  $F_\varepsilon \times X$  is continuous and

$$\limsup_{k \rightarrow +\infty} \gamma_k((X \setminus F_\varepsilon) \times X) = (\pi_1)_\# \gamma(X \setminus F_\varepsilon) < \varepsilon.$$

Then since  $(x, y) \mapsto |\mathbb{1}_G(x) \varphi(x, y)|$  is bounded and hence uniformly integrable with respect to  $(\gamma_k)_k$ , the claim follows from [2, Proposition 5.1.10].  $\square$

**Lemma 7.3.** *Let  $X$  be a separable metric space and  $(\bar{\gamma}_k)_k$  be a sequence in  $\mathcal{P}(X)$  which converges weakly to some  $\bar{\gamma} \in \mathcal{P}(X)$ . Let  $f : X \rightarrow \mathbb{R}$  be a measurable and bounded function which is continuous in  $B$  for some Borel set  $B \subset X$  such that  $\bar{\gamma}_k(X \setminus B) = 0$  for every  $k \in \mathbb{N}$  and  $\bar{\gamma}(X \setminus B) = 0$ , then*

$$\lim_{k \rightarrow \infty} \int_X f d\gamma_k = \int_X f d\gamma.$$

*Proof.* Let  $\bar{f}$  and  $\tilde{f}$  be respectively the lower and upper semicontinuous envelope of  $f$ . We have  $\bar{f} = f = \tilde{f}$  on  $B$  and hence  $\gamma$ -a.e. and  $\gamma_k$ -a.e. for every  $k \in \mathbb{N}$ . It follows that

$$\begin{aligned} \int_X f d\gamma &= \int_X \bar{f} d\gamma \leq \liminf_{k \rightarrow \infty} \int_X \bar{f} d\gamma_k = \liminf_{k \rightarrow \infty} \int_X f d\gamma_k \\ &\leq \limsup_{k \rightarrow \infty} \int_X f d\gamma_k = \limsup_{k \rightarrow \infty} \int_X \tilde{f} d\gamma_k \leq \int_X \tilde{f} d\gamma = \int_X f d\gamma \end{aligned}$$

which proves the claim.  $\square$

## 8 Solution to Monge's problem

We prove that optimal transport plans in  $\Pi_1(\mu, \nu)$  that are obtained as weak limit of solutions of the variational approximations introduced in Section 5 are induced by a transport, hence giving a solution to Monge's transport problem as stated in Theorem 1.1. Note that due to the fact that  $\Pi$  is relatively compact in  $\mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$ , such optimal transport plans do exist.

**Theorem 8.1.** *Let  $\varepsilon_k$  be a sequence converging to 0 and  $\gamma_{\varepsilon_k}$  a sequence of solutions to  $(P_{\varepsilon_k})$  which is weakly converging to some  $\gamma \in \mathcal{P}(\mathbb{H}^n \times \mathbb{H}^n)$ . Then  $\gamma$  is concentrated on a  $\mu$ -measurable graph and hence induced by a transport.*

*Proof.* First we know from Lemma 5.2 that  $\gamma \in \Pi_2(\mu, \nu)$ . From the previous sections and using inner regularity of Borel probability measures, one can then find  $\sigma$ -compact sets  $\Gamma$  and  $\Gamma'$  such that  $\Gamma' \subset \Gamma \subset \Omega$  and the conclusions of Lemma 4.3, Lemma 6.2 and Lemma 7.1 hold. We prove here that for any  $x \in \pi_1(\Gamma')$  there is a unique  $y \in \mathbb{H}^n$  such that  $(x, y) \in \Gamma'$ .

By contradiction, assume that one can find  $x_0 \in \pi_1(\Gamma')$  and  $(x_0, y_0) \in \Gamma$ ,  $(x_0, y_1) \in \Gamma$  with  $y_0 \neq y_1$ . Without loss of generality one can assume that  $d(x_0, y_0) \leq d(x_0, y_1)$  and  $x_0 \neq y_1$ . Then, by Lemma 6.2 and Lemma 7.1, for all  $r > 0$  and for all  $\delta > 0$  small enough, one can find  $x' \in B(x_0, \delta) \cap \Gamma^{-1}(B(y_0, r)) \cap T(\Gamma \cap [B(x_0, \frac{\delta}{2}) \times B(y_1, r)])$ . It follows that one can find  $y' \in B(y_0, r)$  such that  $(x', y') \in \Gamma$  and  $(x, y) \in \Gamma \cap (B(x_0, \frac{\delta}{2}) \times B(y_1, r))$  such that  $x \neq y$ ,  $x' \neq x$  and  $x'$  lie on the minimal curve between  $x$  and  $y$ . Then it follows from Lemma 4.3 that  $x, x', y$  and  $y'$  lie on the same minimal curve ordered in that way.

Assume first that  $d(x_0, y_0) < d(x_0, y_1)$ . We know from Lemma 4.3 that  $d(x, y) \leq d(x, y')$ . On the other hand, we have

$$\begin{aligned}
 d(x, y') &\leq d(x, x_0) + d(x_0, y_0) + d(y_0, y') \\
 &\leq d(x_0, y_0) + \frac{\delta}{2} + r \\
 &= d(x_0, y_1) + d(x_0, y_0) - d(x_0, y_1) + \frac{\delta}{2} + r \\
 &\leq d(x_0, x) + d(x, y) + d(y, y_1) + d(x_0, y_0) - d(x_0, y_1) + \frac{\delta}{2} + r \\
 &\leq d(x, y) + d(x_0, y_0) - d(x_0, y_1) + \delta + 2r.
 \end{aligned}$$

It follows that  $d(x, y') < d(x, y)$  provided we take  $r > 0$  and  $\delta > 0$  small enough which gives a contradiction. If  $d(x_0, y_0) = d(x_0, y_1)$ , we have

$$\begin{aligned}
 d(x_0, y_1) &\leq d(x_0, x) + d(x, y) + d(y, y_1) \\
 &= d(x_0, x) + d(x, y') - d(y', y) + d(y, y_1) \\
 &\leq d(x, y') - d(y', y) + \frac{\delta}{2} + r, \\
 d(x, y') &\leq d(x, x_0) + d(x_0, y_0) + d(y_0, y') \leq d(x_0, y_0) + \frac{\delta}{2} + r, \\
 d(y', y) &\geq d(y_0, y_1) - d(y_0, y') - d(y_1, y) \geq d(y_0, y_1) - 2r,
 \end{aligned}$$

hence,

$$d(x_0, y_1) \leq d(x_0, y_0) - d(y_0, y_1) + 4r + \delta.$$

It follows that  $d(x_0, y_1) < d(x_0, y_0)$  provided we take  $r > 0$  and  $\delta > 0$  small enough which gives also a contradiction.  $\square$

## 9 Extension to more general metric measure spaces

First we note that a major part of intermediate steps in the strategy adopted in the present paper can be naturally extended to Polish and non-branching geodesic spaces equipped with a reference measure for which the Lebesgue's differentiation theorem holds.

Next our choice of approximating costs  $c_\varepsilon$  in the approximation procedure is not the only possible one. This choice could in particular be adapted to fit other contexts (for instance concerning the relevant properties of solutions to the transport problem associated to the approximating cost).

Finally the Measure Contraction Property is here technically very convenient. We note however that this property is unnecessarily too strong for what is actually needed in the proof about the lower density of the transport set. Much local and weaker versions about the behavior of the measure of sets transported along minimal curves are indeed sufficient as clearly shows up from the proof.

This approach can in particular be adapted to give an alternative proof of the existence of solutions to Monge's transport problem in the Riemannian setting without using Sudakov's type arguments.

For the reasons listed above it is furthermore very likely that the present strategy could be adapted and extended to other geodesic metric spaces.

## Bibliography

- [1] L. Ambrosio, *Lecture notes on optimal transport problems*. Mathematical aspects of evolving interfaces (Funchal, 2000), 1–52, Lecture Notes in Math., 1812, Springer, Berlin, 2003.
- [2] L. Ambrosio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005.
- [3] L. Ambrosio, B. Kirchheim, A. Pratelli, *Existence of optimal transport maps for crystalline norms*. Duke Math. J. 125 (2004), no. 2, 207–241.
- [4] L. Ambrosio, A. Pratelli, *Existence and stability results in the  $L^1$  theory of optimal transportation*. Optimal transportation and applications (Martina Franca, 2001), 123–160, Lecture Notes in Math., 1813, Springer, Berlin, 2003.
- [5] L. Ambrosio, S. Rigot, *Optimal mass transportation in the Heisenberg group*. J. Funct. Anal. 208 (2004), no. 2, 261–301.
- [6] G. Anzellotti, S. Baldo, *Asymptotic development by  $\Gamma$ -convergence*. Appl. Math. Optim. 27 (1993), no. 2, 105–123.
- [7] S. Bianchini, F. Cavalletti, *The Monge Problem For Distance Cost In Geodesic Spaces*, (SISSA Preprint 50/2009/M).
- [8] Sub-Riemannian geometry. Edited by André Bellaïche and Jean-Jacques Risler. Progress in Mathematics, 144. Birkhäuser Verlag, Basel, 1996.
- [9] D. Burago, Y. Burago, S. Ivanov, *A course in metric geometry*. Graduate Studies in Mathematics, 33. American Mathematical Society, Providence, RI, 2001.
- [10] L. Caravenna, *A proof of Sudakov theorem with strictly convex norms*. [www.springerlink.com/content/0025-5874/](http://www.springerlink.com/content/0025-5874/).
- [11] L. Caffarelli, M. Feldman, R. J. McCann, *Constructing optimal maps for Monge’s transport problem as a limit of strictly convex costs*. J. Amer. Math. Soc. 15 (2002), no. 1, 1–26.
- [12] T. Champion, L. De Pascale, *The Monge problem for strictly convex norm in  $\mathbb{R}^d$* , J. Eur. Math. Soc. 12 (2010), no. 6, 1355–1369.
- [13] T. Champion, L. De Pascale, *The Monge problem in  $\mathbb{R}^d$* . to appear on Duke Mathematical Journal.
- [14] L. C. Evans, W. Gangbo, *Differential equations methods for the Monge–Kantorovich mass transfer problem*. Mem. Amer. Math. Soc. 137 (1999), no. 653.
- [15] M. Feldman, R. J. McCann, *Monge’s transport problem on a Riemannian manifold*. Trans. Amer. Math. Soc. 354 (2002), no. 4, 1667–1697.
- [16] A. Figalli, N. Juillet, *Absolute continuity of Wasserstein geodesics in the Heisenberg group*. J. Funct. Anal. 255 (2008), no. 1, 133–141.



- [17] B. Gaveau, *Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents*. Acta Math. 139 (1977), no. 1–2, 95–153.
- [18] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*. Progress in Mathematics, 152. Birkhäuser Boston, Inc., Boston, MA, 1999.
- [19] J. Heinonen, *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.
- [20] N. Juillet, *Geometric Inequalities and Generalized Ricci Bounds in the Heisenberg Group*. Int Math Res Notices, 13 (2009), 2347–2373.
- [21] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*. Mathematical Surveys and Monographs, 91. American Mathematical Society, Providence, RI, 2002.
- [22] R. Monti, *Some properties of Carnot–Carathéodory balls in the Heisenberg group*. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 11 (2000), no. 3, 155–167 (2001).
- [23] P. Pansu, *Métriques de Carnot–Carathéodory et quasiisométries des espaces symétriques de rang un*. Ann. of Math. (2) 129 (1989), no. 1, 1–60.
- [24] F. Santambrogio, *Absolute continuity and summability of transport densities: simpler proofs and new estimates*. Calculus of Variations and Partial Differential Equations 36 (2009), no. 3, 343–354.
- [25] V. N. Sudakov, *Geometric problems in the theory of infinite-dimensional probability distributions*. Proc. Steklov Inst. Math. 1979, no. 2, i–v, 1–178.
- [26] N. S. Trudinger, X. J. Wang, *On the Monge mass transfer problem*. Calc. Var. Partial Differential Equations 13 (2001), no. 1, 19–31.
- [27] C. Villani, *Optimal transport. Old and new*. Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, Berlin, 2009.

Received February 18, 2010; revised July 27, 2010.

### Author information

Luigi De Pascale, Università di Pisa, Dipartimento di Matematica Applicata, Via Buonarroti 1/c, 56127 Pisa, Italy.  
E-mail: depascal@dm.unipi.it

Severine Rigot, Université de Nice Sophia-Antipolis, Laboratoire J.-A. Dieudonné, CNRS-UMR 6621, Parc Valrose, 06108 Nice cedex 02, France.  
E-mail: rigot@unice.fr