OPTIMAL MASS TRANSPORTATION IN THE HEISENBERG GROUP

L. AMBROSIO AND S. RIGOT

1. Introduction

The problem of optimal transportation, raised by G. Monge at the end of the XVIII century, has received a lot of attention in recent years, due to the ever growing number of applications of the theory of optimal transportation (see for instance [14], [33], [38]). In the Monge formulation, rephrased and generalized in modern mathematical terms, we are given two Borel probability measures μ, ν in a topological space X and we minimize

$$\psi \mapsto \int_X c(x,\psi(x))\, d\mu(x)$$

among all transport maps ψ, i.e., all Borel maps $\psi : X \to X$ such that $\psi_\# \mu = \nu$ (ψ pushes μ into ν, i.e., $\nu(B) = \mu(\psi^{-1}(B))$ for all Borel set B). Here $c(x,y)$ is given and represents the cost for shipping a unit of mass from x to y ($c(x,y) = |x-y|$ in Monge's formulation).

The main difficulty in attacking the problem is due to the fact that, even when a linear structure is available, the condition $\psi_\#\mu = \nu$ is genuinely non linear with respect to ψ, and this reflects into the fact that the class of transport maps is not closed with respect to any reasonable weak topology. A major breakthrough came in 1942 with Kantorovich's approach to the problem: he proposed to consider the larger class $\Pi(\mu,\nu)$ of *transport plannings*, i.e., probability measures $\boldsymbol{\gamma}$ in $X \times X$ such that $\pi_{0\#}\boldsymbol{\gamma} = \mu$ and $\pi_{1\#}\boldsymbol{\gamma} = \nu$ (π_0, $\pi_1 : X \times X \to X$ being the canonical projections on the first and on the second factor) and to minimize

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

in the class of transport plannings. Notice that any transport map ψ induces via the map $(Id \times \psi)(x) := (x, \psi(x))$ a planning $\boldsymbol{\gamma}$ with the same cost, defined by $(Id \times \psi)_\#\mu$. Due to the linearity of the new constraint, existence of generalized solutions holds under quite general assumptions, for instance if X is a Polish space and c is lower semicontinuous and bounded from below.

Kantorovich's approach also leads to a dual formulation of the optimal transport problem of paramount importance. In fact, almost all proofs of the existence of optimal transport maps either make use of this dual formulation or use necessary and sufficient optimality conditions for optimal plannings to show that any optimal planning is induced by a transport (see [6], [7], [8], [16], [17], [26], [32], [35], [36] but the bibliography is not exhaustive).

Confining our discussion to the case when (X,d) is a metric space and $c(x,y) = d(x,y)^p$ with $p > 1$ (the case $p = 1$ is quite delicate, see [3] and the bibliography therein), the most general existence result available so far is due to R. Mc Cann [26]: it states that if (M,g) is a compact C^3 Riemannian manifold and d is the induced Riemannian distance, then there is existence and uniqueness of an optimal transport map provided the initial measure μ is

Date: 19.11.2002.

absolutely continuous with respect to vol_M, the volume measure of the manifold. Moreover the optimal transport map can be characterized by the equation

$$\psi(x) = \exp_x(-\nabla\varphi(x)) \qquad \text{for } \mu\text{-a.e. } x \in M \tag{1.1}$$

where $\exp_x$ is the Riemannian exponential and φ is a c-concave map (see §4 for terminology). The main technical difficulty to be overcame is the lack of differentiability in the large for the cost function (due to the compactness of the manifold): the clever proof uses some tools of nonsmooth analysis and the fact that, for x close to y, the geodesics are unique and the squared distance function is C^2 (here the smoothness assumption on M plays a role). This leads to semiconcavity estimates for the distance function which are a good substitute, in the situation at hand, for classical differentiability.

It would be quite interesting to extend this result to the class of stratified Carnot groups, arising in a very precise sense (see for instance [5], [20]) as limits of Riemannian manifolds. In this setting uniqueness of geodesics is lost even in the small, as the example of the Heisenberg group $\mathbb{H}^n$ (the simplest model of noncommutative Carnot group) shows.

Since no general strategy for attacking this problem seems by now available, in this paper we investigate the optimal transportation problem in the Heisenberg group $\mathbb{H}^n$, considering for simplicity only the power $p = 2$ (but see §7) and two natural distances on $\mathbb{H}^n$. The first distance we consider is the Carnot–Carathéodory distance d_c, a very natural object in connection with the Riemannian approximation mentioned above. The second distance d_g is the so-called gauge distance induced by the Korànyi norm; this distance is very natural as well, as the fundamental solution of the Kohn sublaplacian is a function of $d_g(0, x)$ (see [15]). In both cases we show existence and uniqueness of an optimal transport map, assuming that μ is absolutely continuous with respect to the Haar measure of the group (a constant multiple of the Lebesgue measure $\mathcal{L}^{2n+1}$ on $\mathbb{R}^{2n+1}$ in exponential coordinates) and a natural non degenaracy condition (see (5.2)).

Our proof works for quite different reasons, depending on the choice of the distance, and is based on the *explicit* computation of the minimal geodesics in $\mathbb{H}^n$. In the case of the distance d_c we use the fact that the unit ball has a cusp-like singularity (see Lemma 3.16), while in the case of the distance d_g we use the smoothness of the norm to derive intrinsic $BV_{\mathbb{H}}$ estimates on the Kantorovich dual potential φ, applying some results of [2] on $BV_{\mathbb{H}}$ functions on stratified groups.

The plan of the paper is the following. In §2 and §3 we recall basic facts on distances, differentiability and geodesics in $\mathbb{H}^n$. In particular we introduce suitable "spherical coordinates" on $\mathbb{H}^n$ and compute the partial derivatives of $d_c(0, \cdot)$ in this system of coordinates. In §4 we use these analitic tools to study the c-superdifferential of c-concave maps when $c(x, y)$ is either $d_c(x, y)^2/2$ or $d_g(x, y)^2/2$, very much in the spirit of [17]. In particular we show that for μ-a.e. x there is a unique y such that (x, y) belongs to the c-superdifferential, and this leads in a standard way in §5 to the existence and the uniqueness of the optimal transport map. Finally, in §6 we show that our transport map for the case $c = d_c^2/2$ arises in a very natural way as a limit of McCann's optimal transport maps relative to the approximating Riemannian manifolds M_ε. Moreover we discuss the validity of the formula (1.1) in this sub-Riemannian setting, by computing the limit of the exponential maps of M_ε and interpreting accordingly our coordinates in $\mathbb{H}^n$ as "exponential" coordinates. Finally, §7 is devoted to the discussion of extensions of our results and open problems.

2. The Heisenberg group

The Heisenberg group $\mathbb{H}^n$ is a noncommutative stratified nilpotent Lie group. As a set it can be identified with its Lie algebra $\mathbb{R}^{2n+1} \simeq \mathbb{C}^n \times \mathbb{R}$, via exponential coordinates. We denote a point in $\mathbb{H}^n$ indifferently by $x = (\xi, \eta, t) = [\zeta, t]$ 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 group law in this system of coordinates is given by

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

The center of the group is $L := \{[0, s] \in \mathbb{H}^n;\ s \in \mathbb{R}\}$. We also define L^*, L^+, L^- restricting s to belong to respectively $\mathbb{R} \setminus \{0\}$, $[0, +\infty)$, $(-\infty, 0]$.

There is a natural family of dilations $(\delta_s)_{s \in \mathbb{R}^+}$ on $\mathbb{H}^n$ defined by $\delta_s([\zeta, t]) := [s\, \zeta, s^2\, t]$. These dilations preserve the group structure.

The Haar measure of the group is given by the Lebsegue measure $\mathcal{L}^{2n+1}$, that we shall also denote sometimes by dx, on $\mathbb{H}^n \simeq \mathbb{R}^{2n+1}$. Unless otherwise stated, when saying that a property holds a.e. on $\mathbb{H}^n$, we will always mean that it holds a.e. with respect to $\mathcal{L}^{2n+1}$.

The left invariant vector fields

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

Lie generate the Lie algebra of the group, the only non trivial bracket relation being $[X_j, Y_j] = -4Z$, where

$$Z := \partial_t.$$

2.1. Metric structures.

There are several ways to define left invariant and homogeneous metrics on $\mathbb{H}^n$, i.e., metrics d such that

$$\begin{aligned} d(z \cdot x, z \cdot y) &= d(x, y), \\ d(\delta_s(x), \delta_s(y)) &= s\, d(x, y), \end{aligned}$$

for all x, y, $z \in \mathbb{H}^n$ and $s \in [0, \infty)$. We shall consider here two particular such distances, the Carnot-Carathéodory distance d_c and the gauge distance d_g.

Let us first consider the Carnot-Carathéodory distance. An absolutely continuous curve $\gamma : [0, T] \to \mathbb{R}^{2n+1}$ is said to be horizontal if there exist measurable functions $h_j : [0, T] \to \mathbb{R}$ such that

$$\dot\gamma(s) = \sum_{j=1}^{n} h_j(s) X_j(\gamma(s)) + h_{n+j}(s) Y_j(\gamma(s))$$

for a.e. $s \in [0, T]$. Equivalently,

$$\dot\gamma_{2n+1}(s) = 2 \sum_{j=1}^{n} \gamma_{n+j}(s)\dot\gamma_j(s) - \gamma_j(s)\dot\gamma_{n+j}(s) \tag{2.1}$$

for a.e. $s \in [0, T]$. We say that γ is sub-unit if moreover $\sum_{j=1}^{2n} h_j(s)^2 \leq 1$ for a.e. $s \in [0, T]$. Since $(X_1, \dots, X_n, Y_1, \dots, Y_n)$ Lie generate the Lie algebra of the group it follows from Chow's

Theorem that any two points x, y in $\mathbb{H}^n$ can be joined by a sub-unit curve. Then the Carnot-Carathéodory distance $d_c(x,y)$ between x and y is defined as

$$d_c(x,y) := \inf\{T > 0\,;\text{ there exists a sub-unit curve } \gamma : [0,T] \to \mathbb{H}^n \\ \text{such that } \gamma(0) = x,\ \gamma(T) = y\}.$$

It follows easily from the left invariance and homogeneity with respect to the dilations of the vector fields X_j and Y_j that d_c is a left invariant and homogeneous distance on $\mathbb{H}^n$. For $x \in \mathbb{H}^n$, we set

$$d_c(x) := d_c(0,x).$$

The gauge distance is induced by the following homogeneous norm

$$\|[\zeta,t]\| := \sqrt[4]{|\zeta|^4 + t^2},$$

where $|\zeta| = \sqrt{\sum_{j=1}^n |\zeta_j|^2}$. For any two points x, y in $\mathbb{H}^n$ the gauge distance $d_g(x,y)$ between x and y is given by

$$d_g(x,y) = \|x^{-1}\cdot y\|.$$

It turns out that d_g is an actual metric on $\mathbb{H}^n$ (the validity of the triangle inequality is not a trivial fact: see [23]) and one can easily check that it is left invariant and homogeneous. For $x \in \mathbb{H}^n$, we set

$$d_g(x) := d_g(0,x).$$

Both distances induce the usual euclidean topology on $\mathbb{H}^n \simeq \mathbb{R}^{2n+1}$. Moreover they turn out to be equivalent, i.e., there exists a constant $C > 0$ such that

$$C^{-1}\, d_g(x,y) \le d_c(x,y) \le C\, d_g(x,y)$$

for any x, $y \in \mathbb{H}^n$ (this is actually true for any couple of left invariant and homogeneous distances which induce the usual euclidean topology on $\mathbb{H}^n$).

2.2. **Calculus on $\mathbb{H}^n$.** We briefly recall for further reference some intrinsic notion of differentiability for real valued maps defined in $\mathbb{H}^n$ and related results. They are basically due to P. Pansu.

We say that a group homomorphism $G : \mathbb{H}^n \to \mathbb{R}$ is homogeneous if $G(\delta_s(x)) = s\,G(x)$ for all $x \in \mathbb{H}^n$ and $s > 0$. In the rest of this paragraph the letter Ω denotes an open subset of $\mathbb{H}^n$.

Definition 2.2 (Pansu differentiability). *We say that a map $f : \Omega \to \mathbb{R}$ is Pansu differentiable at $x \in \Omega$ if there exists an homogeneous group homomorphism $G : \mathbb{H}^n \to \mathbb{R}$ such that*

$$\lim_{y\to x} \frac{f(y) - f(x) - G(x^{-1}\cdot y)}{d_c(y,x)} = 0.$$

The map G is then unique and will be denoted by $D_{\mathbb{H}}f(x)$.

Let $(e_1,\dots,e_{2n+1})$ denote the usual canonical basis of $\mathbb{R}^{2n+1}$.

Definition 2.3 (Derivative along vector fields). *We say that a map $f : \Omega \to \mathbb{R}$ is differentiable at $x \in \Omega$ along X_j (resp. Y_j) if the map $s \mapsto f(x\cdot\delta_s(e_j))$ (resp. $s \mapsto f(x\cdot\delta_s(e_{n+j}))$) is differentiable at $s = 0$ and we denote the corresponding derivative by $X_jf(x)$ (resp. $Y_jf(x)$). Similarly we say that f is differentiable at x along Z if $s \mapsto f(x\cdot\delta_s(e_{2n+1}))$ is differentiable at $s = 0$ and we denote this derivative by $Zf(x)$.*

Note that these notions coincide with the usual ones when f is of class C^1 on $\Omega \subset \mathbb{H}^n \simeq \mathbb{R}^{2n+1}$ in the usual euclidean sense. When f is differentiable at x along X_j and Y_j for all $j \in \{1, \dots, n\}$, we set

$$\nabla_{\mathbb{H}} f(x) := (X_1 f(x), \dots, X_n f(x), Y_1 f(x), \dots, Y_n f(x))$$

and

$$|\nabla_{\mathbb{H}} f(x)| := |(X_j f(x) + iY_j f(x))|.$$

The relationship between Pansu's differential and derivatives along the vector fields is given in the following proposition.

Proposition 2.4. *If $f : \Omega \to \mathbb{R}$ is Pansu differentiable at $x \in \Omega$ then f is differentiable at x along X_j and Y_j for all $j \in \{1, \dots, n\}$ and*

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

An extension of Rademacher's Theorem holds in the present setting, see [31]. We say that a map $f : \Omega \to \mathbb{R}$ is C-Lipschitz if

$$|f(x) - f(y)| \leq C\, d_c(x, y)$$

for all x, $y \in \Omega$. Notice that since d_c and d_g are equivalent we do not need to specify the distance when speaking about Lipschitz maps. However, when considering Lipschitz constants we shall in general refer to the Carnot-Carathéodory distance as in the previous definition.

Theorem 2.5 (Pansu differentiability theorem)**.** *Let $f : \Omega \to \mathbb{R}$ be a C-Lipschitz function. Then f is Pansu differentiable a.e. on Ω and $|\nabla_{\mathbb{H}} f(x)| \leq C$ for a.e. $x \in \Omega$.*

Finally, let us define right an left derivatives along the vector fields X_j and Y_j.

Definition 2.6. *Let $f : \Omega \to \mathbb{R}$ be a function and $x \in \Omega$. We define $X_j^+ f(x)$ (resp. $X_j^- f(x)$) as the right (resp. left) derivative of $s \mapsto f(x \cdot \delta_s(e_j))$ at $s = 0$. We define similarly $Y_j^+ f(x)$ and $Y_j^- f(x)$ as the right and left derivatives of $s \mapsto f(x \cdot \delta_s(e_{n+j}))$ at $s = 0$.*

Remark 2.7 (Chain rule). A simple computation based on the classical chain rule shows that

$$\frac{d}{dt}(\phi \circ \gamma)\bigg|_{t=s} = \sum_{j=1}^{n} X_j \phi(\gamma(s)) \dot{\gamma}_j(s) + \sum_{j=1}^{n} Y_j \phi(\gamma(s)) \dot{\gamma}_{n+j}(s)$$

provided $\gamma : (0, T) \to \mathbb{H}^n$ is differentiable at $s \in (0, T)$, ϕ is continuously differentiable in the classical sense in a neighbourhood of $\gamma(s)$ and the horizontality condition (2.1) is fulfilled at s.

3. Geodesics, CC-spherical coordinates and differentiability of the Carnot-Carathéodory distance

We now turn our attention to minimal geodesics for the distance d_c. By minimal geodesics we shall always mean in this paragraph minimal geodesics *with constant and unit speed* as defined in the usual way as follows.

Definition 3.1 (Minimal geodesics)**.** *We say that a curve $\gamma : [0, T] \to \mathbb{H}^n$ is a minimal geodesic (with constant and unit speed) if $d_c(\gamma(t), \gamma(t')) = |t - t'|$ for all t, $t' \in [0, T]$.*

It turns out that $(\mathbb{H}^n, d_c)$ is a complete and locally compact metric length space in the sense of [20, Definition 1.7], hence any two points can be joined by a minimal geodesic ([20, Theorem 1.10]). Notice that minimal geodesics are 1-Lipschitz functions and it follows from Pansu's Theorem on differentiability of Lipschitz functions between stratified and nilpotent Carnot groups (of which Theorem 2.5 is a particular case) that minimal geodesics are sub-unit curves. Hence we have:

Theorem 3.2. *For any x, $y \in \mathbb{H}^n$, there exists a sub-unit minimal geodesic joining x to y.*

In our present setting where the Heisenberg group is equipped with its Carnot-Carathéodory distance, the equations for the minimal geodesics can be explicitely computed and can be found in the literature, see [19] (see also [5], [28], [29]), with more or less complete proof. However for sake of completeness and because the precise characterization of sub-unit minimal geodesics will play at some places a quite important role here we wish in Theorem 3.4 to give a quite detailed and self-contained proof of this characterization. We set $\mathbb{S} = \{a + ib \in \mathbb{C}^n;\, |a + ib| = 1\}$. For any $a + ib \in \mathbb{S}$, $v \in \mathbb{R}$ and $r > 0$, we say that a curve $\gamma : [0, r] \to \mathbb{H}^n$ is a curve with parameter $(a + ib, v, r)$ if $\gamma(s) = (\xi(s), \eta(s), t(s))$ where

$$\begin{aligned}
\xi_j(s) &= \frac{b_j\,(1-\cos\frac{vs}{r}) + a_j\,\sin\frac{vs}{r}}{v}\,r\,,\\
\eta_j(s) &= \frac{-a_j\,(1-\cos\frac{vs}{r}) + b_j\,\sin\frac{vs}{r}}{v}\,r\,,\\
t(s) &= 2\,\frac{\frac{vs}{r} - \sin\frac{vs}{r}}{v^2}\,r^2\,,
\end{aligned} \tag{3.3}$$

when $v \neq 0$ and

$$\gamma(s) = (a_1 s, \dots, a_n s, b_1 s, \dots, b_n s, 0) \tag{3.3'}$$

when $v = 0$.

Theorem 3.4 (Parameterization of minimal geodesics). *Sub-unit minimal geodesics starting from 0 are curves γ with parameter $(a+ib, v, r)$ for some $(a+ib, v, r) \in \mathbb{S}\times[-2\pi, 2\pi]\times(0, +\infty)$. In particular curves γ with parameter $(a + ib, v, r) \in \mathbb{S} \times \mathbb{R} \times (0, +\infty)$ with $|v| > 2\pi$* are not *minimal geodesics. Conversely any curve γ with parameter $(a + ib, v, r) \in \mathbb{S} \times [-2\pi, 2\pi] \times (0, +\infty)$ is a sub-unit minimal geodesic starting from 0. Moreover we have the following more precise description.*

1. For any $x = [0, t] \in L^*$ with $t > 0$, resp. $t < 0$, sub-unit minimal geodesics from 0 to x form a one parameter family of curves, each one being obtained from one another by a rotation around the axis L. More precisely this family is composed by all curves γ with parameter $(a + ib, 2\pi, \sqrt{\pi t})$, resp. $(a + ib, -2\pi, \sqrt{\pi|t|})$, where $a + ib \in \mathbb{S}$. In particular we have

$$d_c([0, t]) = \sqrt{\pi|t|} \qquad \forall t \in \mathbb{R}. \tag{3.5}$$

2. For any $x \in \mathbb{H}^n \setminus L$ there exists a unique $(a + ib, v, r) \in \mathbb{S} \times (-2\pi, 2\pi) \times (0, +\infty)$ so that the curve γ with parameter $(a+ib, v, r)$ joins 0 to x and in particular γ is the unique sub-unit minimal geodesic from 0 to x.

Proof. Any sub-unit minimal geodesic starting from 0 must be a curve with parameter $(a+ib,v,r)$ for some $(a+ib,v,r)\in\mathbb{S}\times\mathbb{R}\times(0,+\infty)$. This follows for instance from tools of control theory, applying the so-called Pontryagin maximum principle, see for instance [5], [19], [29], and this is well established. What we would like to emphasize here is that such a curve is actually a minimal geodesic if and only if $v\in[-2\pi,2\pi]$ and that one has the description given in points 1 and 2 of the Theorem. Note that all curves γ with parameter $(a+ib,v,r)$ are sub-unit curves so that we always have $d_c(\gamma(r))\le r$ with equality if and only if γ is a minimal geodesic.

We first consider sub-unit minimal geodesics connecting 0 to L^*. Let $x=[0,t]\in L^+\setminus\{0\}$. All curves with parameter $(a+ib,2\pi,\sqrt{\pi t})$ for some $a+ib\in\mathbb{S}$ connect 0 to x hence $d_c(x)\le\sqrt{\pi t}$. Next let γ be a curve with parameter $(a+ib,v,r)\in\mathbb{S}\times\mathbb{R}\times(0,+\infty)$, assume that γ is a sub-unit minimal geodesic from 0 to x and let us show that $v=2\pi$. One has $v\neq 0$ because $t\neq 0$ and because of (3.3'). If $\gamma(s)=(\xi(s),\eta(s),t(s))$ for $s\in[0,r]$, one has $t(r)=t>0$ hence $v>0$. One also must have $\xi_j(r)=\eta_j(r)=0$, i.e.,

$$b_j\,(1-\cos v)+a_j\,\sin v=0,$$
$$b_j\,\sin v-a_j\,(1-\cos v)=0$$

for all $j\in\{1,\dots,n\}$. Since $|a+ib|=1$, the determinant of these systems must vanish, i.e., $1-\cos v=0$ and thus $v=2k\pi$ for some $k\in\mathbb{N}^*$. Then since $t(r)=t$ one must have $r=\sqrt{k\pi t}$ and by minimality of γ it follows that $d_c(x)=\sqrt{k\pi t}\le\sqrt{\pi t}$. This implies that $k=1$ hence $v=2\pi$, and thus $d_c(x)=\sqrt{\pi t}$. It also follows that any curve with parameter $(a+ib,2\pi,\sqrt{\pi t})$ for some $a+ib\in\mathbb{S}$ is actually a sub-unit minimal geodesic from 0 to x and that they are the only possible ones. This proves the point 1 in the Theorem when $x\in L^+\setminus\{0\}$. One can argue in a similar way when $x\in L^-\setminus\{0\}$.

Let us now prove that any curve γ with parameter $(a+ib,v,r)\in\mathbb{S}\times(-2\pi,2\pi)\times(0,+\infty)$ is a sub-unit minimal geodesic starting from 0. When $v\neq 0$, we simply note that $\gamma:[0,r]\to\mathbb{H}^n$ is the restriction to $[0,r]$ of the sub-unit minimal geodesic $\tilde\gamma:[0,2\pi r/v]\to\mathbb{H}^n$ with parameter $(a+ib,2\pi,2\pi r/v)$. Hence γ is still a sub-unit minimal geodesic. To deal with the case $v=0$, we note that a curve γ with parameter $(a+ib,v,r)$ connects 0 to some $x=[\zeta,0]$ if and only if $v=0$ and then such a curve is unique (and is actually the usual euclidean segment between 0 and x). Then the minimality of curves with parameter $(a+ib,0,r)$ follows from the general existence result of Theorem 3.2. We postpone the proof of the non minimality of curves with parameter $((a+ib),v,r)\in\mathbb{S}\times\mathbb{R}\times(0,+\infty)$ for some $|v|>2\pi$, which is in some sense the most delicate point here, after Lemma 3.11. This fact will actually be used only in Lemma 6.8 at the very end of this paper.

To conclude we prove now the existence and uniqueness for any $x\in\mathbb{H}^n\setminus L$ of a parameter $(a+ib,v,r)\in\mathbb{S}\times(-2\pi,2\pi)\times(0,+\infty)$ so that the associated curve γ connects 0 to x. The uniqueness of a sub-unit mimimal geodesic from 0 to x in full generality will follow from this together with the non minimality of curves with parameter $(a+ib,v,r)\in\mathbb{S}\times\mathbb{R}\times(0,+\infty)$ for some $|v|>2\pi$ to be proved later. From now on $x\in\mathbb{H}^n\setminus L$ is fixed. Let Σ denote the union of all sub-unit minimal geodesics from 0 to say $z=[0,1/\pi]$. If $x=[\zeta,t]$ with $t>0$ one can always find $\lambda>0$ so that $\delta_\lambda(x)\in\Sigma$, i.e., one can find $a+ib\in\mathbb{S}$ such that $\delta_\lambda(x)$ is on the sub-unit minimal geodesic $\tilde\gamma$ with parameter $(a+ib,2\pi,1)$ from 0 to z. Let $s_0\in(0,1)$ be such that $\tilde\gamma(s_0)=\delta_\lambda(x)$. Then the curve $\gamma:[0,s_0/\lambda]\to\mathbb{H}^n$ defined by $\gamma(s)=\delta_{1/\lambda}(\tilde\gamma(\lambda s))$ connects 0 to x and is a curve with parameter $(a+ib,2\pi s_0,s_0/\lambda)$. Since $2\pi s_0<2\pi$ we get the required existence (note that the minimality of γ also follows more intrinsically from that of $\tilde{\gamma}$ and the homogeneity of d_c with respect to the dilations). One argue in the same way, considering $z' = [0, -1/\pi]$ in place of z, when $t < 0$. Let us now prove uniqueness. By contradiction assume that one can find two curves γ and $\tilde{\gamma}$ with parameter $(a+ib, v, r) \in \mathbb{S} \times (-2\pi, 2\pi) \times (0, +\infty)$, resp. $(\tilde{a}+i\tilde{b}, \tilde{v}, \tilde{r}) \in \mathbb{S} \times (-2\pi, 2\pi) \times (0, +\infty)$, connecting 0 to x. By minimality of both curves one must have $r = d_c(x) = \tilde{r}$. If $v = 0$ then $t = 0$ by (3.3') and thus $v' = 0$, and conversely. If $v \neq 0$ and $v' \neq 0$, since $\gamma(r) = \tilde{\gamma}(r)$ one must have in particular $|\xi(r) + i\eta(r)| = |\tilde{\xi}(r) + i\tilde{\eta}(r)|$ and $t(r) = \tilde{t}(r)$ hence

$$\frac{1 - \cos v}{v^2} = \frac{1 - \cos \tilde{v}}{\tilde{v}^2} \quad \text{and} \quad \frac{v - \sin v}{v^2} = \frac{\tilde{v} - \sin \tilde{v}}{\tilde{v}^2}.$$

It follows that

$$\frac{1 - \cos v}{v - \sin v} = \frac{1 - \cos \tilde{v}}{\tilde{v} - \sin \tilde{v}}.$$

But the function $v \mapsto (1 - \cos v)/(v - \sin v)$ is strictly decreasing on $(0, 2\pi)$ hence $v = \tilde{v}$. Then once $v = v'$ and $r = r'$ one automatically gets that $a + ib = \tilde{a} + i\tilde{b}$ solving the systems given by the expressions of $\xi_j(r)$ and $\eta_j(r)$, resp. $\tilde{\xi}_j(r)$ and $\tilde{\eta}_j(r)$, whose determinants are non vanishing because $v \in (-2\pi, 2\pi)$. □

One can use this characterization of the minimal geodesics to define some kind of spherical coordinates well adapted to the Carnot-Carathéodory distance. Note that we will only use here the minimality of all curves with parameter $(a + ib, v, r) \in \mathbb{S} \times (-2\pi, 2\pi) \times (0, +\infty)$ together with the existence and uniqueness of such a curve connecting 0 to any $x \in \mathbb{H}^n \setminus L$ (but not the non minimality of curves with parameter $(a + ib, v, r) \in \mathbb{S} \times \mathbb{R} \times (0, +\infty)$ with $|v| > 2\pi$ whose proof will be given later). We set

$$\begin{aligned} \Phi : \mathbb{S} \times [-2\pi, 2\pi] \times (0, +\infty) &\to \mathbb{H}^n \\ (a + ib, v, r) &\mapsto (\xi, \eta, t) \end{aligned}$$

where

$$\begin{aligned} \xi_j &:= \frac{b_j(1 - \cos v) + a_j \sin v}{v} r, \\ \eta_j &:= \frac{-a_j(1 - \cos v) + b_j \sin v}{v} r, \\ t &:= 2\,\frac{v - \sin v}{v^2} r^2, \end{aligned} \tag{3.6}$$

if $v \neq 0$, and

$$\xi_j := a_j r, \quad \eta_j := b_j r, \quad t := 0, \tag{3.6'}$$

if $v = 0$.

It follows immediately from Theorem 3.4 that the range of Φ is $\mathbb{H}^n \setminus \{0\}$ and that

$$d_c(\Phi(a + ib, v, r)) = r \tag{3.7}$$

for all $(a + ib, v, r) \in \mathbb{S} \times [-2\pi, 2\pi] \times (0, +\infty)$.

This map Φ is not injective on $\{(a + ib, v, r);\ a + ib \in \mathbb{S},\ |v| = 2\pi,\ r > 0\}$ whose image corresponds to the center L of the group. More precisely one has

$$\begin{aligned} L^+ \setminus \{0\} &= \{\Phi(a + ib, v, r) \in \mathbb{H}^n;\ a + ib \in \mathbb{S},\ v = 2\pi,\ r > 0\}, \\ L^- \setminus \{0\} &= \{\Phi(a + ib, v, r) \in \mathbb{H}^n;\ a + ib \in \mathbb{S},\ v = -2\pi,\ r > 0\}. \end{aligned} \tag{3.8}$$

On the other hand it follows from the first part of the second point in Theorem 3.4 that Φ is a bijective map from D onto $\mathbb{H}^n \setminus L$ where

$$D = \mathbb{S} \times (-2\pi, 2\pi) \times (0, +\infty).$$

Moreover an explicit computation of the Jacobian of Φ (see [29]) shows that it does not vanish on D.

Proposition 3.9. *The map Φ is a C^1 diffeomorphism from D onto $\mathbb{H}^n \setminus L$.*

One can then consider the variables $a + ib \in \mathbb{S}$, $v \in (-2\pi, 2\pi)$ and $r > 0$ as a system of global coordinates on $\mathbb{H}^n \setminus L$ (see also §5 and §6). When r is fixed, $(a+ib, v) \in \mathbb{S} \times (-2\pi, 2\pi) \mapsto \Phi(a + ib, v, r)$ gives a parameterization of the sphere centered at the origin with radius r, at least outside of L, that's why we shall call these coordinates CC-spherical, or simply spherical, in the sequel (see also [4]).

Notice also that an elementary computation gives the rule for the computation of the inverse in these coordinates:

$$[\Phi(a + ib, v, r)]^{-1} = \Phi(\tilde{a} + i\tilde{b}, -v, r) \qquad \text{with } \tilde{b}_j + i\tilde{a}_j := -(b_j + ia_j)e^{iv}. \tag{3.10}$$

We now use these spherical coordinates to study the differentiability properties of the distance function d_c.

Lemma 3.11 (Differential of d_c). *The distance function from the origin d_c is of class C^1 on $\mathbb{H}^n \setminus L$. Moreover if $x \in \mathbb{H}^n \setminus L$ and $(a+ib, v, r) \in D$ are such that $x = \Phi(a+ib, v, r)$ we have*

$$X_j d_c(x) = b_j \sin v + a_j \cos v, \tag{3.12}$$

$$Y_j d_c(x) = b_j \cos v - a_j \sin v, \tag{3.13}$$

$$Z d_c(x) = \frac{v}{4r}. \tag{3.14}$$

Proof. First since Φ a C^1 diffeomorphism on D it follows from (3.7) that d_c is of class C^1 in the usual euclidean sense on $\mathbb{H}^n \setminus L$. Let $x \in \mathbb{H}^n \setminus L$ be fixed, $(a + ib, v, r) \in D$ be such that $\Phi(a + ib, v, r) = x$ and let $\gamma : [0, r] \to \mathbb{H}^n$ be the corresponding sub-unit minimal geodesic joining 0 to x with parameter $(a + ib, v, r)$. Since $s = d_c(\gamma(s))$ and $\gamma(s) \notin L$ for all $s \in (0, r)$, one gets by differentiation with respect to s, taking into account Remark 2.7, that

$$\begin{aligned}
1 &= \frac{d}{ds}(d_c \circ \gamma)(s) \\
&= \sum_{j=1}^{n} \left(b_j \sin\frac{vs}{r} + a_j \cos\frac{vs}{r}\right) X_j d_c(\gamma(s)) + \left(b_j \cos\frac{vs}{r} - a_j \sin\frac{vs}{r}\right) Y_j d_c(\gamma(s)) \\
&\le |a + ib| \cdot |\nabla_{\mathbb{H}}\, d_c(\gamma(s))| \le 1.
\end{aligned}$$

Hence we have equality in Cauchy-Schwartz's inequality and since $|a + ib| = 1$ we get $|\nabla_{\mathbb{H}}\, d_c(x)| = 1$ (see also [30] for the validity of this eikonal equation in a more general framework) and

$$X_j d_c(\gamma(s)) = b_j \sin\frac{vs}{r} + a_j \cos\frac{vs}{r}, \quad Y_j d_c(\gamma(s)) = b_j \cos\frac{vs}{r} - a_j \sin\frac{vs}{r},$$

for all $s \in (0, r)$. Finally (3.12) and (3.13) follow by continuity, letting $s \uparrow r$.

Let us now prove (3.14). We fix $x = [\xi + i\eta, t] \in \mathbb{H}^n \setminus L$ and $(a + ib, v, r) \in D$ such that $\Phi(a + ib, v, r) = x$ as before. Let us first assume that $v \neq 0$. For $h \in \mathbb{R}$, we set

$x \cdot [0,h] = [\xi + i\eta, t+h] = \Phi(a_h + ib_h, v_h, r_h)$ with $(a_h + ib_h, v_h, r_h) \in D$. We shall always consider h small enough so that $v_h \neq 0$. What we want to prove is that

$$d_c(x \cdot [0,h]) - d_c(x) = \frac{v}{4r} \cdot h + o(h).$$

By minimality of the associated curves with parameter $(a+ib, v, r)$, resp. $(a_h + ib_h, v_h, r_h)$, (see also (3.7)) we have $r = d_c(x)$ and $r_h = d_c(x \cdot [0,h])$ hence it will be sufficient to prove that

$$r_h - r = \frac{v}{4r} \cdot h + o(h). \tag{3.15}$$

Set $\varepsilon_h := v_h - v$. Since Φ is a C^1 diffeomorphism from D onto $\mathbb{H}^n \setminus L$, we have $\varepsilon_h = o(1)$. It follows from the expressions of ξ_j and η_j in terms of a_j, b_j and r given by the definition (3.6) of Φ that

$$|\xi + i\eta|^2 = 2 \cdot \frac{1-\cos v}{v^2} \cdot r^2 = 2 \cdot \frac{1 - \cos v_h}{v_h^2} \cdot r_h^2 \cdot$$

Then, considering now the expression of $t+h$ and t, we get

$$\begin{aligned} t + h &= 2 \cdot \frac{v_h - \sin v_h}{v_h^2} \cdot r_h^2 = 2 \cdot \frac{v_h - \sin v_h}{1 - \cos v_h} \cdot \frac{1-\cos v}{v^2} \cdot r^2 \\ &= \frac{v_h - \sin v_h}{1-\cos v_h} \cdot \frac{1-\cos v}{v - \sin v} \cdot t = \frac{f(v_h)}{f(v)} \cdot t \\ &= t + \frac{f'(v)}{f(v)} \cdot t \cdot \varepsilon_h + o(\varepsilon_h), \end{aligned}$$

where

$$f(v) := \frac{v - \sin v}{1 - \cos v}\,,$$

and it follows that

$$\varepsilon_h \underset{h \to 0}{\sim} \frac{f(v)}{f'(v)} \cdot \frac{1}{t} \cdot h.$$

Going back to r_h and setting

$$g(v) := \frac{|v|}{\sqrt{1-\cos v}}\,,$$

we get

$$\begin{aligned} r_h &= \frac{g(v_h)}{g(v)} \cdot r = r + \frac{g'(v)}{g(v)} \cdot r \cdot \varepsilon_h + o(\varepsilon_h) \\ &= r + \frac{g'(v)}{g(v)} \cdot \frac{f(v)}{f'(v)} \cdot \frac{r}{t} \cdot h + o(h). \end{aligned}$$

We have

$$g'(v) = \frac{2(1-\cos v) - v\sin v}{2\,v\,(1-\cos v)} \cdot g(v)\,,$$
$$f'(v) = \frac{2(1-\cos v) - v\sin v}{(1-\cos v)(v-\sin v)} \cdot f(v)\,,$$
$$\frac{r}{t} = \frac{v^2}{2\,(v-\sin v)} \cdot \frac{1}{r}$$

and thus

$$r_h = r + \frac{v}{4r} \cdot h + o(h)$$

as wanted. When $v = 0$, computations are similar and even simpler. In that case we have $x = [\xi + i\eta, 0] = \Phi(a+ib, 0, r)$ with $\xi_j = a_j r$ and $\eta_j = b_j r$. If $h \neq 0$ and $[\xi + i\eta, h] = \Phi(a_h + ib_h, v_h, r_h)$ then $v_h \neq 0$, $v_h = o(1)$ and

$$|\xi + i\eta| = r = \frac{\sqrt{2(1-\cos v_h)}}{|v_h|} \cdot r_h\,,$$

hence $r_h = r + o(v_h)$ and the conclusion will follow if we prove that, up to a multiplicative constant, v_h and h are equivalent. We have

$$h = 2 \cdot \frac{v_h - \sin v_h}{v_h^2} \cdot r_h^2 = \frac{v_h - \sin v_h}{1-\cos v_h} \cdot r^2 = \frac{r^2}{3} \cdot v_h + o(v_h)\,,$$

hence

$$v_h \underset{h\to 0}{\sim} \frac{3}{r^2} \cdot h$$

as wanted. □

We return now to Theorem 3.4 and prove the non minimality of curves with parameter $(a+ib, v, r)$ for some $|v| > 2\pi$ with $v \neq 0 \,\mathrm{mod}(2\pi)$. By contradiction assume that there exists such a curve γ with is a sub-unit minimal geodesic. For simplicity assume that $v > 0$ so that $v > 2\pi$ and set $x = \gamma(r) \in \mathbb{H}^n \backslash L$. By the open mapping theorem one can find a neighbourhood U of x and a neighbourhood V of $(a+ib, v, r)$ and solutions of the geodesic equations (not necessarily minimal), i.e., curves with parameter $(\tilde{a} + i\tilde{b}, \tilde{v}, \tilde{r}) \in V$, whose end points cover U. In particular for $x_h = x \cdot [0, h]$ with h small enough one can find $(a_h + ib_h, v_h, r_h)$ such that the curve γ_h with parameter $(a_h + ib_h, v_h, r_h)$ connects 0 to x_h and such that $v_h - v = o(1)$. Then, arguing exactly in the same way as for the proof of (3.15) (which does not use the minimality of γ_h), one gets

$$r_h - r = \frac{v}{4r} \cdot h + o(h).$$

By minimality of γ we have $d_c(x) = r$ and we always have $d_c(x_h) \leq r_h$. It follows

$$d_c(x_h) - d_c(x) \leq \frac{v}{4r} \cdot h + o(h)$$

Then, taking $h < 0$ and passing to the limit when $h \uparrow 0$, we get $rZd_c(x) \geq v/4 > \pi/2$. On the other hand we already know from (3.14) that $rZd_c(x) = v'/4$ where $(a' + ib', v', r') \in D$ are the CC-spherical coordinates of x, hence $rZd_c(x) < \pi/2$, which gives the required contradiction. One can argue in a similar way when $v = 0$ or $v < 0$ and this concludes the proof of Theorem 3.4.

We now go back to the differentiability properties of the distance function d_c and turn our attention to the situation along L^* where this function turns out to be singular. The following Lemma gives a precise quantification of this in terms of the right and left derivatives of d_c along the vector field X_j.

Lemma 3.16. *For any $x \in L^*$ and $j \in \{1,\dots,n\}$ we have*

$$X_j^+ d_c(x) = -1, \qquad X_j^- d_c(x) = 1.$$

Proof. Let $x = [0,t] \in L^*$ and $j \in \{1,\dots,n\}$ be fixed. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(s) = d_c(x \cdot \delta_s(e_j))$. We have $f'(s) = X_j d_c(x \cdot \delta_s(e_j))$ whenever at least one of the two quantities exists. Thus since $x \cdot \delta_s(e_j) \notin L$ whenever $s \neq 0$ it follows from Lemma 3.11 that f is differentiable on $\mathbb{R}^*$ and, for any $s \neq 0$,

$$f'(s) = X_j d_c(x \cdot \delta_s(e_j)) = b_j(s) \sin v_s + a_j(s) \cos v_s$$

where $(a(s) + ib(s), v_s, r_s) \in D$ are the spherical coordinates of $x \cdot \delta_s(e_j)$. Then in terms of the function f the Lemma will follow if we prove that

$$\lim_{s \to 0^+} f'(s) = -1, \qquad \lim_{s \to 0^-} f'(s) = 1.$$

If $x \cdot \delta_s(e_j) = [\xi(s) + i\eta(s), t]$, we have $\xi_l(s) = \eta_l(s) = 0$ for all $l \neq j$, hence, according to (3.6),

$$b_l(s)(1 - \cos v_s) + a_l(s) \sin v_s = 0\,,$$
$$b_l(s) \sin v_s - a_l(s)(1 - \cos v_s) = 0\,.$$

Note that $v_s \neq 0$ because otherwise we would have $t = 0$ (see (3.6')) and $|v_s| \neq 2\pi$ if $s \neq 0$ because then $x \cdot \delta_s(e_j) \notin L$. Then it follows from the previous system that $a_l(s) = b_l(s) = 0$ for all $l \neq j$. Next we get $|a(s) + ib(s)| = |a_j(s) + ib_j(s)| = 1$ hence one can find $u_s \in [0, 2\pi)$ such that $a_j(s) = \sin u_s$ and $b_j(s) = \cos u_s$. Then we have

$$f'(s) = \sin(u_s + v_s)$$

and we will get the conclusion as soon as we prove that

$$\lim_{s \to 0} v_s = \begin{cases} 2\pi^- & \text{if } t > 0 \\ -2\pi^+ & \text{if } t < 0 \end{cases} \tag{3.17}$$

and

$$\lim_{s \to 0^+} u_s = 3\pi/2, \tag{3.18}$$

$$\lim_{s \to 0^-} u_s = \pi/2. \tag{3.19}$$

We first prove (3.17). Since $|v_s| < 2\pi$ it is sufficient to prove that the only possible limit for v_s is 2π if $t > 0$ and -2π if $t < 0$. Thus let $(s_k)_{k \in \mathbb{N}}$ be a sequence converging to 0, set $(a_k + ib_k, v_k, r_k) := (a(s_k) + ib(s_k), v_{s_k}, r_{s_k})$, $u_k := u_{s_k}$ and let us assume that v_k converges to some $v \in [-2\pi, 2\pi]$. We have $r_k = d_c(x \cdot \delta_{s_k}(e_j))$ by (3.7) hence r_k converges to $d_c(x)$. Since u_k is bounded one can also assume that, at least up to a subsequence, u_k converges to some u. Then, setting $a = \sin u\, \tilde{e}_j$ and $b = \cos u\, \tilde{e}_j$, where $\tilde{e}_j$ denotes the j-th vector of the canonical basis of $\mathbb{R}^n$, we have since Φ is continuous,

$$x = \lim_{k \to +\infty} x \cdot \delta_{s_k}(e_j) = \lim_{k \to +\infty} \Phi(a_k + ib_k, v_k, r_k) = \Phi(a + ib, v, d_c(x))\,.$$

Then, according to (3.8), we get $v = 2\pi$ if $x \in L^+ \setminus \{0\}$, that is, if $t > 0$, and $v = -2\pi$ if $t < 0$ as wanted.

Let us now prove (3.18). Since u_s is bounded it is also sufficient here to prove that the only possible limit for u_s when s goes to 0^+ is $3\pi/2$. Thus we consider a sequence $(s_k)_{k\in\mathbb{N}}$ of positive numbers that goes to 0 and, with the same notations as above, we assume that u_k converges to some $u \in [0, 2\pi]$. As before we have $v_k \neq 0$ and, since $\eta_j(s_k) = 0$, we get from (3.6) that

$$-\sin u_k(1 - \cos v_k) + \cos u_k \sin v_k = 0\,.$$

We also have $|v_k| < 2\pi$ hence one can rewrite the previous equation as

$$\tan u_k = \frac{\sin v_k}{1 - \cos v_k}\,\cdot$$

Let us now assume that $t > 0$. Then it follows from (3.17) that

$$\lim_{k\to+\infty} \tan u_k = -\infty$$

therefore either $u = 3\pi/2^+$ or $\pi/2^+$. On the other hand we have $s_k > 0$ by assumption and $v_k > 0$ for k large enough since v_k converges to 2π. Then, setting

$$\begin{aligned}\alpha_k &:= \cos u_k(1 - \cos v_k) + \sin u_k \sin v_k \\ &= \cos u_k - \cos(u_k + v_k) \\ &= 2\sin(u_k + \frac{v_k}{2}) \sin\frac{v_k}{2}\end{aligned} \tag{3.20}$$

and recalling that $\xi_j(s_k) = s_k$, we get from (3.6) that $\alpha_k = s_k v_k / r_k > 0$ if k is large enough. Now, if $u = \pi/2^+$, the first sin in the factorization of α_k converges to -1 and the second one converges to 0^+. This gives a contradiction hence $u = 3\pi/2$ as wanted. When $t < 0$ then $\tan u_k$ goes to $+\infty$ because now v_k converges to $-2\pi^+$, thus either $u = 3\pi/2^-$ or $\pi/2^-$ and one excludes the case $u = \pi/2^-$ by considering as before the sign of α_k.

We argue in a similar way to prove (3.19). We consider a sequence $(s_k)_{k\in\mathbb{N}}$ of negative numbers that goes to 0 and, with the same notations as before, we assume that u_k converges to some $u \in [0, 2\pi]$. If $t > 0$ then $\tan u_k$ tends to $-\infty$ as before hence either $u = \pi/2^+$ or $3\pi/2^+$. We have $s_k < 0$ by assumption and $v_k > 0$ if k is large enough because v_k converges to $2\pi^-$. Hence $\alpha_k = s_k v_k / r_k$ must be negative. On the other hand, if $u = 3\pi/2^+$ and $v_k \to 2\pi^-$, we get from (3.20) that $\alpha_k > 0$ for k large enough. This gives the required contradiction and $u = \pi/2$ as required. Finally one argue in the same way to get the conclusion when $t < 0$. □

Remark 3.21. One can prove in the same way that

$$Y_j^+ d_c(x) = -1, \qquad Y_j^- d_c(x) = 1$$

for all $x \in L^*$ and $j \in \{1, \ldots, n\}$.

4. Superdifferential and c-concave maps

Let $c : \mathbb{H}^n \times \mathbb{H}^n \to \mathbb{R}$ be a function. We first recall the general definitions of c-concavity and c-superdifferential (see [34], [35], [36]).

Definition 4.1 (c-concavity). *We say that a function $\varphi : \mathbb{H}^n \to \mathbb{R}$ is c-concave if*

$$\varphi(x) = \inf_{y \in Y} c(x,y) + \phi(y) \qquad \forall x \in \mathbb{H}^n \tag{4.2}$$

for some suitable nonempty set $Y \subset \mathbb{H}^n$ and $\phi : Y \to \mathbb{R} \cup \{+\infty\}$, $\phi \not\equiv +\infty$.

Definition 4.3 (c-superdifferential). *Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a function. The c-superdifferential $\partial_c\varphi(x)$ of φ at $x \in \mathbb{H}^n$ is defined by*

$$\partial_c\varphi(x) = \{y \in \mathbb{H}^n;\ c(x,y) - \varphi(x) \le c(z,y) - \varphi(z) \text{ for all } z \in \mathbb{H}^n\}\,.$$

The c-superdifferential $\partial_c\varphi$ of φ is the graph of the multivalued map $x \mapsto \partial_c\varphi(x)$, i.e.,

$$\partial_c\varphi = \{(x,y) \in \mathbb{H}^n \times \mathbb{H}^n;\ y \in \partial_c\varphi(x)\}\,.$$

We shall more specifically consider these notions with a function c given by $c(x,y) = d_c(x,y)^2/2$ or $c(x,y) = d_g(x,y)^2/2$, the main result of this section being that in both cases the c-superdifferential of a locally Lipschitz c-concave map is, up to a null set, a graph. Recall that $\pi_0 : \mathbb{H}^n \times \mathbb{H}^n \to \mathbb{H}^n$ denote the first canonical projection.

Theorem 4.4. *Let $c(x,y) = d_c(x,y)^2/2$ or $c(x,y) = d_g(x,y)^2/2$ and let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a c-concave and locally Lipschitz map. Then there exists a measurable graph Γ such that $\mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi \setminus \Gamma)) = 0$.*

By measurable graph we mean a set Γ of the form

$$\Gamma = \{(x,\psi(x)) \in \mathbb{H}^n \times \mathbb{H}^n;\ x \in S\}$$

for some measurable set S and some measurable map $\psi : S \to \mathbb{H}^n$. We also refer to Lemma 4.8 and Lemma 4.11 for explicit expressions of the corresponding graph in Theorem 4.4.

The proof relies strongly on the differentiability properties of the functions d_c^2 and d_g^2 and those of the related c-concave maps. Note that in the previous statement φ is locally Lipschitz hence $X_j\varphi$ and $Y_j\varphi$ exist a.e. on $\mathbb{H}^n$, by Pansu's Theorem. Let us also stress right now that in both cases any c-concave φ with Y in (4.2) bounded is locally Lipschitz. More generally this is true whenever $c(x,y)$ is of the form $d(x,y)^2/2$ for some distance d.

Lemma 4.5. *Let d be a distance on $\mathbb{H}^n$ and set $c(x,y) = d(x,y)^2/2$. If φ is a c-concave map with Y in* (4.2) *bounded with respect to d then φ is locally Lipschitz with respect to d. Moreover, if $Lip_K(\varphi)$ denotes the Lipschitz constant of φ on the compact set K, we have $Lip_K(\varphi) \le \sup_{y\in Y, z \in K} d(z,y)$.*

Proof. This is a consequence of the following simple computation,

$$\begin{aligned} 2|c(x,y) - c(x',y)| &= |d(x,y)^2 - d(x',y)^2| \\ &\le d(x,y)\,|d(x,y) - d(x',y)| + d(x',y)\,|d(x,y) - d(x',y)| \\ &\le 2\,d(x,x') \sup_{y\in Y, z\in K} d(z,y) \end{aligned}$$

whenever $y \in Y$ and x, $x' \in K$ for some bounded set K. Hence φ can be locally represented as the infimum of a family of uniformly Lipschitz functions and thus is itself locally Lipschitz. The bound on the local Lipschitz constant for φ follows easily from the previous sequence of inequalities. □

We now consider the derivative of c-concave maps along the vector field Z. In what follows we shall often denote by c_y the function defined by $c_y(x) = c(x,y)$.

Lemma 4.6. *Let $c(x,y) = d_c(x,y)^2/2$ and let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a c-concave map. Then $Z\varphi$ exists and $|Z\varphi| \leq \pi/2$ a.e. in $\mathbb{H}^n$. If $c(x,y) = d_g(x,y)^2/2$ the same statement holds with $1/2$ in place of $\pi/2$.*

Proof. First let us assume that $c(x,y) = d_c(x,y)^2/2$ and let φ be a c-concave map. We prove first that, for any $\zeta \in \mathbb{C}^n$, the function $\varphi_\zeta : \mathbb{R} \to \mathbb{R}$ defined by $\varphi_\zeta(t) = \varphi([\zeta,t])$ is $\frac{\pi}{2}$-Lipschitz on $\mathbb{R}$ equipped with the usual euclidean distance. To see this we first fix $z \in \mathbb{H}^n$ and set $f_z(t) = 2c_z([0,t]) = d_c(z^{-1}\cdot[0,t])^2$ for $t \in \mathbb{R}$. Using the left invariance of the vector field Z, Lemma 3.11 and more specifically (3.14), we get that $f_z'(t)$ exists whenever $z^{-1}\cdot[0,t] \notin L$, and, if $(a+ib,v,r) \in D$ denote the spherical coordinates of $z^{-1}\cdot[0,t]$,

$$f_z'(t) = 2Zc_z([0,t]) = 2\,d_c(z^{-1}\cdot[0,t])\,Zd_c(z^{-1}\cdot[0,t]) = v/2$$

hence $|f_z'(t)| < \pi$. Note that $z^{-1}\cdot[0,t] \in L$ for some $t \in \mathbb{R}$ is equivalent to $z \in L$ and then $z^{-1}\cdot[0,t] \in L$ for all $t \in \mathbb{R}$. Hence we have just shown that if $z \notin L$ then f_z is π-Lipschitz. If $z \in L$, $z = [0,t_z]$ for some $t_z \in \mathbb{R}$, it follows from (3.5) that $f_z(t) = d_c([0,t-t_z])^2 = \pi|t-t_z|$ hence $|f_z'(t)| = \pi$ for any $t \neq t_z$. Next, since φ is c-concave, we have

$$\varphi_\zeta(t) = \inf_{y\in Y} c_y([\zeta,t]) + \phi(y) = \inf_{y\in Y} \frac{1}{2} f_{x\cdot y}(t) + \phi(y)$$

for some suitable $Y \subset \mathbb{H}^n$ and $\phi : Y \to \mathbb{R}\cup\{+\infty\}$ and with $x = [\zeta,0]^{-1}$. In other words φ_ζ is the infimum of a family of functions that are $\frac{\pi}{2}$-Lipschitz and hence is itself $\frac{\pi}{2}$-Lipschitz as claimed. In particular φ_ζ is differentiable a.e. on $\mathbb{R}$ for all $\zeta \in \mathbb{C}^n$. Finally, since $Z\varphi([\zeta,t]) = \varphi_\zeta'(t)$, the Lemma follows from Fubini's Theorem.

When $c(x,y) = d_g(x,y)^2/2$ the proof follows the same lines. With the same notations as before, if $z = [\zeta,t_z]$, a direct computation gives

$$f_z'(t) = \frac{t-t_z}{\sqrt{|\zeta|^4 + (t-t_z)^2}}$$

whenever $z^{-1}\cdot[0,t] \neq 0$. Hence if $z \notin L$, that is, $\zeta \neq 0$, we have $|f_z'(t)| < 1$ and f_z is 1-Lipschitz. If $z \in L$, i.e., $\zeta = 0$, then f_z is continuous on $\mathbb{R}$ and differentiable on $\mathbb{R}^*$ with $|f_z'(t)| = 1$ for any $t \neq t_z$ hence f_z is also 1-Lipschitz. And one concludes as in the first case. □

The general strategy to prove Theorem 4.4 in both cases of the Carnot-Carathéodory and the gauge distance is similar. However, due to the different behaviour of d_c^2 and d_g^2 along the center L of the group, d_c^2 being singular there whereas $\nabla_{\mathbb{H}}\, d_g^2$ vanishes, we shall technically get the final conclusion for quite different reasons.

4.1. Superdifferential of c-concave maps when $\mathbf{c(x,y) = d_c(x,y)^2/2}$.

In this paragraph we set $c(x,y) = d_c(x,y)^2/2$ and prove Theorem 4.4 in that case. First we consider a differentiability point x for a map φ along X_j and we prove that due to the singularity of the distance function d_c on L^* one can exclude the case where $\partial_c\varphi(x)$ meets $x\cdot L^*$.

Lemma 4.7. *Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ and $x \in \mathbb{H}^n$ be such that $X_j\varphi(x)$ exists for some $j \in \{1,\dots,n\}$. Then $\partial_c\varphi(x) \cap x\cdot L^* = \emptyset$.*

Proof. We let $\varphi : \mathbb{H}^n \to \mathbb{R}$ and $x \in \mathbb{H}^n$ be as in the statement and we assume by contradiction that one can find $y \in \partial_c\varphi(x) \cap x \cdot L^*$. By definition of the c-superdifferential we have

$$\varphi(x \cdot \delta_s(e_j)) - \varphi(x) \leq c_y(x \cdot \delta_s(e_j)) - c_y(x)$$

for all $s \in \mathbb{R}$. If $s > 0$, dividing by s and recalling the definition of c_y, we get

$$\frac{\varphi(x \cdot \delta_s(e_j)) - \varphi(x)}{s} \leq \frac{d_c(y^{-1} \cdot x \cdot \delta_s(e_j))^2 - d_c(y^{-1} \cdot x)^2}{2\,s}.$$

The left hand side goes to $X_j\varphi(x)$ when s goes to 0. We have $y^{-1} \cdot x \in L^*$ by assumption and it follows from Lemma 3.16 that the right hand side goes to $d_c(y^{-1} \cdot x) X_j^+ d_c(y^{-1} \cdot x) = -d_c(y^{-1} \cdot x) < 0$. Hence we get $X_j\varphi(x) < 0$. On the other hand if $s < 0$ we have

$$\frac{\varphi(x \cdot \delta_s(e_j)) - \varphi(x)}{s} \geq \frac{d_c(y^{-1} \cdot x \cdot \delta_s(e_j))^2 - d_c(y^{-1} \cdot x)^2}{2\,s}.$$

Now the right hand side goes to $d_c(y^{-1} \cdot x) X_j^- d_c(y^{-1} \cdot x) = d_c(y^{-1} \cdot x) > 0$ once again by Lemma 3.16 and hence $X_j\varphi(x) > 0$ which gives a contradiction and concludes the proof. ☐

Next we show that the c-superdifferential of any map at a differentiability point along X_j, Y_j and Z is a singleton.

Lemma 4.8. *Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ and let $x \in \pi_0(\partial_c\varphi)$ be such that $X_j\varphi(x)$, $Y_j\varphi(x)$ exist for all $j \in \{1, \dots, n\}$ and $Z\varphi(x)$ exists. Then $\partial_c\varphi(x) = \{y\}$ is a singleton, and if $y \neq x$ the point $x^{-1} \cdot y$ has spherical coordinates*

$$a_j := \frac{-X_j\varphi(x)}{|\nabla_{\mathbb{H}}\varphi(x)|}, \quad b_j := \frac{-Y_j\varphi(x)}{|\nabla_{\mathbb{H}}\varphi(x)|}, \quad v := -4Z\varphi(x), \quad r := |\nabla_{\mathbb{H}}\varphi(x)|. \tag{4.9}$$

Proof. Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ and $x \in \pi_0(\partial_c\varphi)$ be as in the statement. We already know from Lemma 4.7 that $\partial_c\varphi(x) \cap x \cdot L^* = \emptyset$. Next let us prove that $\partial_c\varphi(x) \setminus x \cdot L$ is a singleton whenever it is non empty. Thus assume that one can find $y \in \partial_c\varphi(x) \setminus x \cdot L$. Recalling that $c_y(x) = d_c(y^{-1} \cdot x)^2/2$, it follows from the left invariance of the vector fields X_j, Y_j and Z and from Lemma 3.11 that $X_jc_y(x)$, $Y_jc_y(x)$ and $Zc_y(x)$ exist. Moreover, according to (3.12), (3.13) and (3.14), we have

$$\begin{aligned} X_jc_y(x) &= d_c(y^{-1} \cdot x)\, X_jd_c(y^{-1} \cdot x) = r\,(b_j \sin v + a_j \cos v), \\ Y_jc_y(x) &= d_c(y^{-1} \cdot x)\, Y_jd_c(y^{-1} \cdot x) = r(b_j \cos v - a_j \sin v), \\ Zc_y(x) &= d_c(y^{-1} \cdot x)\, Zd_c(y^{-1} \cdot x) = \frac{v}{4}, \end{aligned}$$

where $(a + ib, v, r) \in D$ are the spherical coordinates of $y^{-1} \cdot x$. On the other hand, since $X_j\varphi(x)$, $Y_j\varphi(x)$ and $Z\varphi(x)$ also exist, it follows from the minimality of $x' \mapsto c_y(x') - \varphi(x')$ at $x' = x$ that $X_j\varphi(x) = X_jc_y(x)$, $Y_j\varphi(x) = Y_jc_y(x)$, $Z\varphi(x) = Zc_y(x)$, that is,

$$\begin{aligned} r(b_j \sin v + a_j \cos v) &= X_j\varphi(x), \\ r(b_j \cos v - a_j \sin v) &= Y_j\varphi(x), \\ v &= 4\,Z\varphi(x). \end{aligned} \tag{4.10}$$

Recalling that $a + ib \in \mathbb{S}$ this implies $r = |\nabla_{\mathbb{H}}\varphi(x)|$, and then, with v and r fixed, (a_j, b_j) is uniquely determined by the two first equations in (4.10). Since $y = x \cdot \Phi(a + ib, v, r)^{-1}$ we obtain (4.9) taking into account (3.10). Let us note that this also implies that $\nabla_{\mathbb{H}}\varphi(x) \neq 0$

whenever $\partial_c\varphi(x) \setminus x \cdot L \neq \emptyset$ since here $r = d_c(y^{-1} \cdot x) > 0$. Finally to conclude the proof it remains to show that one cannot have $x \in \partial_c\varphi(x)$ together with $\partial_c\varphi(x) \setminus x \cdot L \neq \emptyset$. Note that

$$2X_j c_x(x) = \lim_{s\to 0} d_c(\delta_s(e_j))^2/s = \lim_{s\to 0} s = 0$$

by homogeneity. Similarly we have $Y_j c_x(x) = 0$. Hence if $x \in \partial_c\varphi(x)$ we have $\nabla_{\mathbb{H}}\, \varphi(x) = \nabla_{\mathbb{H}}\, c_x(x) = 0$. This contradicts the preceding remark for the case where $\partial_c\varphi(x) \setminus x \cdot L \neq \emptyset$ and concludes the proof. □

Theorem 4.4 follows now quite easily.

Proof of Theorem 4.4. Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a c-concave and locally Lipschitz map. It follows from Theorem 2.5 and Proposition 2.4 that $X_j\varphi$ and $Y_j\varphi$ exist a.e. on $\mathbb{H}^n$ and from Lemma 4.6 that $Z\varphi$ also exists a.e. on $\mathbb{H}^n$. Thus, setting

$$S = \{x \in \pi_0(\partial_c\varphi);\ X_j\varphi(x), Y_j\varphi(x) \text{ exist for all } j \in \{1, \dots, n\}, Z\varphi(x) \text{ exists}\}\,,$$

we have $\mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi) \setminus S) = 0$. According to Lemma 4.8, $\partial_c\varphi(x)$ is a singleton for any $x \in S$. Moreover denoting by $\psi(x)$ the unique point in $\partial_c\varphi(x)$ for $x \in S$ it follows from (4.9) that the map ψ is measurable. Finally let Γ be the graph of ψ. Since $\partial_c\varphi \cap \pi_0^{-1}(S) = \Gamma$ we have $\pi_0(\partial_c\varphi \setminus \Gamma) = \pi_0(\partial_c\varphi \setminus \pi_0^{-1}(S)) = \pi_0(\partial_c\varphi) \setminus S$ hence $\mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi \setminus \Gamma)) = 0$ as wanted. □

4.2. **Superdifferential of c-concave maps when $\mathbf{c(x,y) = d_g(x,y)^2/2}$.**
We now assume that $c(x,y) = d_g(x,y)^2/2$ and prove Theorem 4.4 in that case.

Lemma 4.11. *Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ and let $x \in \pi_0(\partial_c\varphi)$ be such that $X_j\varphi(x)$, $Y_j\varphi(x)$ exist for all $j \in \{1, \dots, n\}$ and $Z\varphi(x)$ exists. Then we have*

(i) $$\text{either } \partial_c\varphi(x) \subset \mathbb{H}^n \setminus x \cdot L\,,$$

and then $\partial_c\varphi(x) = \{y\}$ is a singleton with $y^{-1} \cdot x = (\xi, \eta, t)$ where

$$\begin{aligned}
\xi_j &= \sqrt{1 - 4|Z\varphi(x)|^2}\, X_j\varphi(x) - 2Z\varphi(x)\, Y_j\varphi(x)\,,\\
\eta_j &= 2Z\varphi(x)\, X_j\varphi(x) + \sqrt{1 - 4|Z\varphi(x)|^2}\, Y_j\varphi(x)\,,\\
|\xi + i\eta| &= |\nabla_{\mathbb{H}}\, \varphi(x)|\,, \qquad t = \frac{2|\nabla_{\mathbb{H}}\, \varphi(x)|^2}{\sqrt{1 - 4|Z\varphi(x)|^2}}\, Z\varphi(x),
\end{aligned}$$

(ii) $$\text{or } \partial_c\varphi(x) \subset x \cdot L.$$

Moreover, when (ii) holds and $\partial_c\varphi(x) \setminus \{x\} \neq \emptyset$, we have

$$\nabla_{\mathbb{H}}\, \varphi(x) = 0 \ \text{ and } \ |Z\varphi(x)| = \frac{1}{2}. \tag{4.12}$$

Proof. First of all let us compute the derivatives of the function $c_y : \mathbb{H}^n \to \mathbb{R}$, $c_y(x) := d_g(y^{-1} \cdot x)^2/2$. If $y \in \mathbb{H}^n$, $y \neq x$, we have

$$X_j c_y(x) = \frac{(\xi_j\rho^2 + \eta_j t)}{\sqrt{\rho^4 + t^2}}\,, \quad Y_j c_y(x) = \frac{(\eta_j\rho^2 - \xi_j t)}{\sqrt{\rho^4 + t^2}}\,, \quad Zc_y(x) = \frac{t}{2\sqrt{\rho^4 + t^2}}\,,$$

whenever $y^{-1} \cdot x = (\xi, \eta, t)$ and $\rho = |\xi + i\eta|$. In particular

$$|\nabla_{\mathbb{H}}\, c_y(x)| = \rho.$$

When $y = x$, using the homogeneity of the gauge distance, we get

$$2X_j c_x(x) = \lim_{s\to 0} d_g(\delta_s(e_j))^2/s = \lim_{s\to 0} s = 0 \qquad \forall j \in \{1,\dots,n\}$$

and similarly $Y_j c_x(x) = 0$ for all $j \in \{1,\dots,n\}$, hence

$$\nabla_{\mathbb{H}}\, c_x(x) = 0.$$

Note that $Zc_x(x)$ does not exist but we have $Z^+c_x(x) = 1/2$ and $Z^-c_x(x) = -1/2$. Next since $X_j\varphi(x)$, $Y_j\varphi(x)$ and $Z\varphi(x)$ exist by assumption, the minimality of $x' \mapsto c_y(x') - \varphi(x')$ at $x' = x$ when $y \in \partial_c\varphi(x)$ implies that for such a pair $(x,y) \in \partial_c\varphi$ we have $X_j c_y(x) = X_j\varphi(x)$ and $Y_j c_y(x) = Y_j\varphi(x)$ and also $Zc_y(x) = Z\varphi(x)$ whenever $y \neq x$.

Let us now prove the alternative of the statement. We argue by contradiction and first assume that one can find $y \in \partial_c\varphi(x) \setminus x\cdot L$ and $y' \in \partial_c\varphi(x) \cap x \cdot L$. If $y^{-1}\cdot x = (\xi,\eta,t)$ and $\rho = |\xi + i\eta|$, we have $\rho \neq 0$ because $y^{-1}\cdot x \notin L$ and it follows that

$$|\nabla_{\mathbb{H}}\,\varphi(x)| = |\nabla_{\mathbb{H}}\, c_y(x)| = \rho > 0.$$

On the other hand if $y'^{-1}\cdot x = (\xi',\eta',t')$ and $\rho' = |\xi' + i\eta'|$, then $\rho' = 0$ because now $y'^{-1}\cdot x \in L$ and the same argument gives

$$\nabla_{\mathbb{H}}\,\varphi(x) = 0$$

and thus the required contradiction. Note that if $y' \neq x$ then $t' \neq 0$, $Z\varphi(x) = Zc_{y'}(x) = t'/2|t'|$ and (4.12) follows. Next it remains to show that $\partial_c\varphi(x)\setminus x\cdot L$ is a singleton whenever it is not empty. With the same notation as above where $y \in \partial_c\varphi(x)\setminus x\cdot L$, $y^{-1}\cdot x = (\xi,\eta,t)$ and $\rho = |\xi+i\eta|$, we have already proved that $\rho = |\nabla_{\mathbb{H}}\,\varphi(x)|$. Then, thanks to the expression of $Zc_y(x)$ and recalling that $Z\varphi(x) = Zc_y(x)$ because $y \neq x$, we get that $|Z\varphi(x)| < 1/2$ and

$$t = \frac{2|\nabla_{\mathbb{H}}\,\varphi(x)|^2}{\sqrt{1-4|Z\varphi(x)|^2}}\, Z\varphi(x).$$

Finally, once $\rho > 0$ and t are given, (ξ_j,η_j) is the unique solution of the 2×2 system

$$\begin{aligned}
\frac{\rho^2}{\sqrt{\rho^4+t^2}}\,\xi_j + \frac{t}{\sqrt{\rho^4+t^2}}\,\eta_j &= X_j\varphi(x)\,,\\
-\frac{t}{\sqrt{\rho^4+t^2}}\,\xi_j + \frac{\rho^2}{\sqrt{\rho^4+t^2}}\,\eta_j &= Y_j\varphi(x).
\end{aligned}$$

Hence $y = x\cdot(\xi,\eta,t)^{-1}$ is uniquely determined as claimed and one gets the explicit form of $\xi + i\eta$ taking into account the expression of ρ and t and solving these systems. □

Now to prove that $\partial_c\varphi$ is, up to a null set, a measurable graph when φ is a locally Lipschitz c-concave function, it will be sufficient to rule out the case where (ii) happens together with $\partial_c\varphi(x)\setminus\{x\} \neq \emptyset$. This will be an easy consequence of the following Lemma.

Lemma 4.13. *Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a c-concave and locally Lipschitz map. Then, for all $j \in \{1,\dots,n\}$, we have*

$$\mathcal{L}^{2n+1}(\{x \in \mathbb{H}^n;\, X_j\varphi(x) = Y_j\varphi(x) = 0,\, Z\varphi(x) \neq 0\}) = 0.$$

Proof. Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a c-concave and locally Lipschitz map, fix $j \in \{1, \ldots, n\}$ and set

$$S := \{x \in \mathbb{H}^n;\ X_j\varphi(x) = Y_j\varphi(x) = 0,\ Z\varphi(x) \neq 0\}.$$

First recall that $X_j\varphi$ and $Y_j\varphi$ exist a.e. by Theorem 2.5 and Proposition 2.4, $Z\varphi$ also exists a.e. by Lemma 4.6 and $X_j\varphi$, $Y_j\varphi$ and $Z\varphi \in L^\infty_{loc}(\mathbb{H}^n)$. Next it follows from [2, Theorem 4.3] that $X_j\varphi$ and $Y_j\varphi \in BV_{\mathbb{H},loc}$ (we refer to [2] and the references therein for the definition and more details about the intrinsic BV space $BV_{\mathbb{H}}$). Then it follows from the differentiability properties of $BV_{\mathbb{H}}$ functions proved in [2] (in particular Theorem 2.2 there) that $X_j\varphi$ and $Y_j\varphi$ are approximately differentiable a.e. in $\mathbb{H}^n$ with zero approximate differential a.e. on S. Then, once again by [2, Theorem 2.2], we get that the measure $X_jY_j\varphi - Y_jX_j\varphi$ has zero absolutely continuous part with respect to $\mathcal{L}^{2n+1}$ on S. On the other hand this measure coincides with the measure $-4Z\varphi\, d\mathcal{L}^{2n+1}$ which has non vanishing absolutely continous part on S. Hence $\mathcal{L}^{2n+1}(S) = 0$. □

We have now all the ingredients to prove Theorem 4.4 in a similar way as in the case of the Carnot-Carathéodory distance.

Proof of Theorem 4.4. Let $\varphi : \mathbb{H}^n \to \mathbb{R}$ be a c-concave and locally Lipschitz map. If

$$S = \{x \in \pi_0(\partial_c\varphi);\ X_j\varphi(x), Y_j\varphi(x) \text{ exist for all } j \in \{1, \ldots, n\}, Z\varphi(x) \text{ exists}\}\,,$$

then $\mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi) \setminus S) = 0$ for the same reasons as in the case of the Carnot-Carathéodory distance (φ is locally Lipschitz and $Z\varphi$ exists a.e. by Lemma 4.6). Next it follows from Lemma 4.13 and (4.12) that either case (i) in Lemma 4.11 or $\partial_c\varphi(x) = \{x\}$ hold for a.e. $x \in \pi_0(\partial_c\varphi)$. Let us denote by $\hat{S}$ the set of such points and by $\psi(x)$ the unique point in $\partial_c\varphi(x)$ when $x \in \hat{S}$. The map ψ is measurable because of the formulas in (i) of Lemma 4.11. Finally if Γ denotes the graph of ψ over $\hat{S}$, as in the case where $c(x, y) = d_c(x, y)^2/2$, we have $\mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi \setminus \Gamma)) = \mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi) \setminus \hat{S}) = \mathcal{L}^{2n+1}(\pi_0(\partial_c\varphi) \setminus S) = 0$. □

5. Optimal transportation

We consider in this section the problem of optimal transportation in $\mathbb{H}^n$ associated to the cost functions $c(x, y) = d_c(x, y)^2/2$ or $c(x, y) = d_g(x, y)^2/2$.

Before stating the main result of this paper we introduce a new notation, useful to make the connection with Mc Cann's existence theorem more clear. We interpret the CC-spherical coordinates introduced in §3 as some kind of generalized exponential coordinates on $\mathbb{H}^n$. For any $A + iB \in \mathbb{C}^n$ and $w \in [-\pi/2, \pi/2]$ we set

$$\exp_{\mathbb{H}}(A + iB, w) := \Phi\left(\frac{A + iB}{|A + iB|}, 4\,w, |A + iB|\right)$$

where Φ is the defining function for the CC-spherical coordinates given in (3.6) and (3.6'), with the convention that $\exp_{\mathbb{H}}(A + iB, w) = 0$ whenever $A + iB = 0$. Recall that it follows from Proposition 3.9 that $\exp_{\mathbb{H}}$ is a C^1 diffeomorphism from $(\mathbb{C}^n \setminus \{0\}) \times (-\pi/2, \pi/2)$ onto $\mathbb{H}^n \setminus L$ and that $\exp_{\mathbb{H}}(A + iB, w) \in L^+ \setminus \{0\}$, resp. $L^- \setminus \{0\}$, if and only if $w = \pi/2$, resp. $w = -\pi/2$. This "exponential" map should not be confused with the canonical exponential map of the theory of Lie groups (mapping the Lie algebra onto the Lie group). Actually in this paper we always tacitly identified elements of the group and elements of the Lie algebra, so that the canonical exponential map never appeared in a explicit form.

A motivation for such a terminology and notation comes from the following relation between CC-spherical coordinates and velocities of sub-unit minimal geodesics in $(\mathbb{H}^n, d_c)$. If $x =$

$\Phi(a+ib,v,r) \in \mathbb{H}^n \setminus L$ with $(a+ib,v,r) \in D$ and if $\gamma : [0,r] \to \mathbb{H}^n$ denotes the sub-unit minimal geodesic in $(\mathbb{H}^n, d_c)$ joining 0 to x (whose equation is given by (3.3) or (3.3')), then the curve $l : [0,1] \to \mathbb{H}^n$ defined by $l(s) = \gamma(sr)$ satisfies

$$\begin{aligned} \dot{l}_j(s) &= v\, l_{n+j}(s) + a_j\, r\,, \\ \dot{l}_{n+j}(s) &= v\, l_j(s) + b_j\, r\,, \\ \dot{l}_{2n+1}(s) &= 2\sum_{j=1}^{n} l_{n+j}(s)\, \dot{l}_j(s) - l_j(s)\, \dot{l}_{n+j}(s). \end{aligned}$$

Hence CC-spherical coordinates $(a+ib,v,r) \in D$ and $\mathbb{H}$-exponential coordinates $(A+iB,w) \in \mathbb{C}^n \setminus \{0\} \times (-\pi/2, \pi/2)$ are related to the velocity of l through the relation $\dot{l}_j(0) = a_j\, r = A_j$, $\dot{l}_{n+j}(0) = b_j\, r = B_j$ for all $j \in \{1, \dots, n\}$ and the fact that l is an horizontal curve. Moreover it turns out, as we shall see in Lemma 6.8, that one can also recover these $\mathbb{H}$-exponential coordinates as the limit of canonical approximating Riemannian exponential coordinates.

Theorem 5.1 (Optimal transport for $c = d_c^2/2$). *Let μ and ν be two Borel probability measures on $\mathbb{H}^n$. Assume that μ is absolutely continuous with respect to $\mathcal{L}^{2n+1}$ and that*

$$\int_{\mathbb{H}^n} d_c(x)^2\, d\mu(x) + \int_{\mathbb{H}^n} d_c(y)^2\, d\nu(y) < +\infty. \tag{5.2}$$

Then there exists a unique optimal transport plan from μ to ν, and this plan is induced by a map ψ. If $\operatorname{spt}\nu$ *is compact, ψ is given by*

$$\psi(x) := x \cdot \exp_{\mathbb{H}}(-X\varphi(x) - iY\varphi(x), -Z\varphi(x)) \qquad \text{for } \mu\text{-a.e. } x \in \mathbb{H}^n \tag{5.3}$$

for some c-concave and locally Lipschitz map φ.

Conversely, if ψ is representable as in (5.3) *for some map φ such that*

$$X\varphi(x),\ Y\varphi(x),\ Z\varphi(x) \text{ exist} \quad \text{and} \quad \varphi(x) = \min_{y\in\mathbb{H}^n} c(x,y) - \varphi^c(y) \quad \text{for } \mu\text{-a.e. } x \in \mathbb{H}^n \tag{5.4}$$

and if

$$\int_{\mathbb{H}^n} d_c(x)^2 + d_c(\psi(x))^2\, d\mu(x) < +\infty \tag{5.5}$$

then ψ is the optimal transport map between μ and $\nu := \psi_\#\mu$.

Here $\varphi^c(y) := \inf_{x\in\mathbb{H}^n} c(x,y) - \varphi(x)$ is the so-called c-transform of φ. Notice that the second condition in (5.4) is slightly stronger than c-concavity, since it requires the infimum to be attained. But this will be necessary in view of (5.6).

Proof. We first prove that any optimal planning $\boldsymbol{\gamma}$ from μ to ν is concentrated on a graph, so that $\boldsymbol{\gamma} = (Id \times \psi)_\#\mu$ for some admissible, and then optimal, map ψ. As c is continuous and (5.2) holds, the duality theory for the optimal transport problem (see for instance [3] and also [33], [17], [38]) imply the existence of a *Kantorovich potential* φ, i.e., a c-concave map $\varphi : \mathbb{H}^n \to \mathbb{R}$ such that $\boldsymbol{\gamma}$ is optimal if and only if

$$c(x,y) = \varphi(x) + \varphi^c(y) \qquad \boldsymbol{\gamma}\text{-a.e.} \tag{5.6}$$

Notice also that the inequality $\varphi(x)+\varphi^c(y) \le c(x,y)$ is always satisfied, so that $c(x',y)-\varphi(x')$ attains its minimum, equal to $\varphi^c(y)$, at $x' = x$ for $\boldsymbol{\gamma}$-a.e. $(x,y) \in \mathbb{H}^n \times \mathbb{H}^n$. It follows that any optimal $\boldsymbol{\gamma}$ is concentrated on the c-superdifferential $\partial_c\varphi$ of φ. Note also that since the first marginal of $\boldsymbol{\gamma}$ is μ we also have $\partial_c\varphi(x) \neq \emptyset$ for μ-a.e. x. Finally, we recall that when $\operatorname{spt}\nu$ is bounded then, by construction, φ can represented as in (4.2) with Y bounded, and therefore is locally Lipschitz in $\mathbb{H}^n$.

For any integer $R > 0$ we define

$$\varphi_R(x) := \inf_{y \in B_R(0)} c(x, y) - \varphi^c(y)$$

and notice that φ_R is c-concave, locally Lipschitz according to Lemma 4.5, c-concave and $\varphi_R \geq \varphi$. One can apply Theorem 4.4 to obtain that for $\mathcal{L}^{2n+1}$-a.e. (and then μ-a.e.) $x \in \pi_0(\partial_c \varphi_R)$ the c-superdifferential $\partial_c \varphi_R(x)$ at x is a singleton. This implies that $\partial_c \varphi(x)$ is also a singleton. Indeed assume that one can find y and $y' \in \partial_c \varphi(x)$ and take R large enough so that $R > \max(d_c(y), d_c(y'))$. Then, since $y \in B_R(0)$, we have for all $x' \in \mathbb{H}^n$ that $\varphi_R(x') \leq c(x', y) - \varphi^c(y)$. It follows

$$c(x, y) - \varphi_R(x) \leq c(x, y) - \varphi(x) = \varphi^c(y) \leq c(x', y) - \varphi_R(x')$$

hence $y \in \partial_c \varphi_R(x)$. Similarly we have $y' \in \partial_c \varphi_R(x)$ hence $y = y'$ and $\partial_c \varphi(x)$ is a singleton for μ-a.e. x, i.e., any optimal $\boldsymbol{\gamma}$ is concentrated on a graph as claimed.

Uniqueness follow from the linearity of the Kantorovich formulation together with the representation of any optimal planning as measures induced by a transport map. Let ψ and ψ' be two optimal transport maps from μ to ν. The plannings $\boldsymbol{\gamma} = (Id \times \psi)_{\#}\mu$ and $\boldsymbol{\gamma}' = (Id \times \psi')_{\#}\mu$ are both optimal. By linearity $\boldsymbol{\gamma}'' = (\boldsymbol{\gamma} + \boldsymbol{\gamma}')/2$ is still optimal and thus is induced by a transport map ψ'', that is, $\boldsymbol{\gamma}'' = (Id \times \psi'')_{\#}\mu$. But this is possible only if $\psi'' = \psi = \psi'$ μ-a.e. on $\mathbb{H}^n$.

Now we show the representation (5.3). As $\operatorname{spt} \nu \subset B_R(0)$ for some $R > 0$, it follows that φ is c-concave and locally Lipschitz, therefore we can apply Theorem 4.4 directly to φ. Then the explicit formula for y follows by (4.9) and by the definition of $\exp_{\mathbb{H}}$.

Now we show that the representation (5.3), together with (5.4) and (5.5), implies minimality of ψ. Again by (5.6) the optimality of $(Id \times \psi)_{\#}\mu$ would be proved if we show that $c(x, \psi(x)) = \varphi(x) + \varphi^c(\psi(x))$ for μ-a.e. x. By (5.4) we can consider with no loss of generality points x where $X\varphi(x)$, $Y\varphi(x)$ and $Z\varphi(x)$ exist. Moreover, we can also assume that there exists y such that $\varphi(x) = c(x, y) - \varphi^c(y)$. It follows that

$$c(x', y) - \varphi(x') \geq \varphi^c(y) = c(x, y) - \varphi(x)$$

for all $x' \in \mathbb{H}^n$ hence $y \in \partial_c \varphi(x)$ and Theorem 4.4 together with (4.9) and the definition of $\exp_{\mathbb{H}}$ yield $y = \psi(x)$. □

Example 5.7 (Optimality of right translations). Let μ be a Borel probability measure on $\mathbb{H}^n$, absolutely continuous with respect to $\mathcal{L}^{2n+1}$ and with compact support. Then the right translations ψ of the form $\psi(x) := x \cdot x_0$ for some $x_0 \in \mathbb{H}^n$ are optimal transport maps from μ to $\nu := \psi_{\#}\mu$ relative to the cost $c(x, y) := d_c(x, y)^2/2$ if and only if $x_0 = [\zeta_0, 0]$ for some $\zeta_0 \in \mathbb{C}^n$. Indeed if $x_0 = 0$ this is obvious. On the other hand we know that if a map ψ is an optimal transport then $\psi(x) \notin x \cdot L^*$ for μ-a.e. $x \in \mathbb{H}^n$ (remember Lemma 4.7) hence right translations associated to any $x_0 \in L^*$ cannot be optimal. Next note that since μ has compact support, (5.5) holds for any right translation $\psi(x) = x \cdot x_0$. For simplicity let us work from now on in $\mathbb{H}^1$ (arguments are similar in $\mathbb{H}^n$). Let $x_0 = (\xi_0, \eta_0, t_0) \in \mathbb{H}^1 \setminus L$ be fixed and assume that the associated right translation is optimal. It follows from the representation (5.3) that one must have $x_0 = \exp_{\mathbb{H}}(-X_1\varphi(x) - iY_1\varphi(x), Z\varphi(x))$ for some c-concave and locally Lipschitz map φ. One can also assume with no loss of generality that $\varphi(0) = 0$. Then it is not hard to see, using the commutator relations, that this implies that

$t_0 = 0$ and $\varphi(\xi, \eta, t) = -\xi_0\xi - \eta_0\eta$. Conversely let $x_0 = (\xi_0, \eta_0, 0) \in \mathbb{H}^1 \setminus \{0\}$ be fixed, set $\varphi(\xi, \eta, t) = -\xi_0\xi - \eta_0\eta$ and let us prove that φ satisfies (5.4) so that the associated right translation $\psi(x) = x \cdot x_0 = x \cdot \exp_{\mathbb{H}}(-X_1\varphi(x) - iY_1\varphi(x), Z\varphi(x))$ is optimal according to Theorem 5.1. Since $X_1\varphi$, $Y_1\varphi$ and $Z\varphi$ exist everywhere, the conclusion will follow if we prove that

$$\varphi(x) = \min_{y \in \mathbb{H}^1} c(x, y) - \frac{\varphi(x_0)}{2} + \varphi(y). \tag{5.8}$$

Thus let $x = (\xi_x, \eta_x, t_x)$ be fixed and set $f(y) := c(x, y) - \varphi(x_0)/2 + \varphi(y)$. We first prove that f attains its infimum on $\mathbb{H}^1$. On one hand, remembering that $d_c \geq C_0\, d_g$ for some absolute constant $C_0 > 0$, we have for all $(\xi, \eta, t) \in \mathbb{H}^1$

$$\begin{aligned} f(\xi, \eta, t) &\geq C\, d_g(x, (\xi, \eta, t))^2 - \frac{\varphi(x_0)}{2} + \varphi(\xi, \eta, t) \\ &\geq 2^{-1} C\, (|\xi - \xi_x|^2 + |\eta - \eta_x|^2 + |t - t_x - 2(\xi\eta_x - \eta\xi_x)|) - \frac{\varphi(x_0)}{2} - \xi_0\xi - \eta_0\eta \\ &\geq C_1\xi^2 - C_1'|\xi| + C_2\eta^2 - C_2'|\eta| + C_3|t| - C_4 \end{aligned}$$

for some suitable positive constants which depends only on x and x_0. It follows that one can find $M > 0$ large enough so that $f(y) > \varphi(x)$ for all $y = (\xi, \eta, t)$ such that $\max(|\xi|, |\eta|, |t|) \geq M$. On the other hand we have $f(x \cdot x_0) = \varphi(x)$. Then since f is continuous and hence reaches its minimum on the compact set $\{(\xi, \eta, t) \in \mathbb{H}^1;\ \max(|\xi|, |\eta|, |t|) \leq M\}$, it follows immediately that f reaches its minimum on $\mathbb{H}^1$ as well. Finally let $y_0 \in \mathbb{H}^1$ be such that f is minimal at y_0 and let us prove that $y_0 = x \cdot x_0$ so that (5.8) holds. First since $f(x) > \varphi(x) = f(x \cdot x_0)$ we have $y_0 \neq x$. Next recall that $y \mapsto c_x(y) := c(x, y)$ is differentiable along Z on $\mathbb{H}^n \setminus \{x\}$ and along X_1 and Y_1 on $\mathbb{H}^n \setminus x \cdot L$ (see (3.5) and Lemma 3.11). If $y_0 \in x \cdot L \setminus \{x\}$ then we get by minimality of f at y_0 that $Zc_x(y_0) = Zf(y_0) = 0$. On the other hand it follows from (3.5) that $|Zc_x(y_0)| = \pi/2$ which gives a contradiction. Hence $y_0 \notin x \cdot L$. Then setting $x^{-1} \cdot y_0 = \exp_{\mathbb{H}}(A + iB, w)$ with $(A + iB, w) \in \mathbb{C} \setminus \{0\} \times (-\pi/2, \pi/2)$ and using once again the minimality of f at y_0 together with (3.12), (3.13) and (3.14), we get $w = Zc_x(y_0) = Zf(y_0) = 0$ hence $w = 0$ and $A = X_1c_x(y_0) = -X_1\varphi(y_0) = \xi_0$, $B = Y_1c_x(y_0) = -Y_1\varphi(y_0) = \eta_0$ hence $y_0 = x \cdot x_0$ as claimed.

Remark 5.9. The representation (5.3) holds even when $\operatorname{spt}\nu$ is not compact, provided differentials are replaced by *approximate* differentials. We recall that φ is said to be approximately differentiable at x if there exists $\hat{\varphi}$ differentiable at x such that $\{\varphi \neq \hat{\varphi}\}$ has 0 density at x. The approximate differential is by definition the differential of $\hat{\varphi}$ at x. As the proof of the theorem shows, μ-a.e. point x is of density 1 for $\{\varphi = \varphi_R\}$ for some $R > 0$ and a differentiability point of φ_R, therefore φ is approximately differentiable μ-a.e.

In the next theorem we consider the case when $c = d_g^2/2$. Its proof is identical to the proof of Theorem 5.1, being based on the general duality theory for the optimal transport problem and on Theorem 4.4 together with Lemma 4.11.

Theorem 5.10 (Optimal transport for $c = d_g^2/2$)**.** *Let μ and ν be two Borel probability measures on $\mathbb{H}^n$. Assume that μ is absolutely continuous with respect to $\mathcal{L}^{2n+1}$ and that* (5.2) *holds. Then there exists a unique optimal transport plan from μ to ν, and this plan is induced by a map ψ. If* $\operatorname{spt}\nu$ *is compact, there exists a c-concave and locally Lipschitz function φ such* *that the coordinates (ξ,η,t) of $(\psi(x))^{-1}\cdot x$ are given for μ-a.e. $x\in\mathbb{H}^n$ such that $\psi(x)\neq x$ by*

$$\begin{aligned}\xi_j &= \sqrt{1-4|Z\varphi(x)|^2}\,X_j\varphi(x) - 2Z\varphi(x)\,Y_j\varphi(x)\,,\\ \eta_j &= 2Z\varphi(x)\,X_j\varphi(x) + \sqrt{1-4|Z\varphi(x)|^2}\,Y_j\varphi(x)\,,\\ |\xi+i\eta| &= |\nabla_{\mathbb{H}}\varphi(x)|\,,\qquad t=\frac{2|\nabla_{\mathbb{H}}\varphi(x)|^2}{\sqrt{1-4|Z\varphi(x)|^2}}\,Z\varphi(x)\,.\end{aligned}$$

Conversely, if ψ is representable as above for some c-concave map φ which satisfies (5.4) *and if*

$$\int_{\mathbb{H}^n} d_g(x)^2 + d_g(\psi(x))^2\,d\mu(x) < +\infty$$

then ψ is the optimal transport map between μ and $\nu := \psi_{\#}\mu$.

6. The Riemannian approximation

To fix the ideas we shall denote by M the set $\mathbb{R}^{2n+1}$ equipped with the usual euclidean topology and with its trivial structure of manifold. Note that as a topological space and as a manifold, M and $\mathbb{H}^n$ coincide. For any $\varepsilon>0$, we denote by M_ε the manifold M equipped with the Riemannian structure where the metric tensor at the point x has

$$X_1,\dots,X_n\,,\quad Y_1,\dots,Y_n\,,\quad Z_\varepsilon := \varepsilon\,Z$$

as an orthonormal basis. We denote by d_ε the induced Riemannian distance,

$$d_\varepsilon(x,y) := \inf\int_0^1 |\dot{l}(s)|_\varepsilon\,ds\,,$$

where the infimum is taken amoung all piecewise C^1 curves $l:[0,1]\to M_\varepsilon$ such that $l(0)=x$, $l(1)=y$, and where

$$|\dot{l}(s)|_\varepsilon = \sqrt{\sum_{j=1}^n \dot{l}_j(s)^2 + \dot{l}_{n+j}(s)^2 + \dot{l}_\varepsilon(s)^2}$$

whenever

$$\dot{l}(s) = \sum_{j=1}^n \dot{l}_j(s)X_j(l(s)) + \dot{l}_{n+j}(s)Y_j(l(s)) + \dot{l}_\varepsilon(s)Z_\varepsilon.$$

More generally, given a piecewise C^1 curve $l:[0,1]\to M_\varepsilon$, we denote by $(\dot{l}_1(s),\dots,\dot{l}_{2n}(s),\dot{l}_\varepsilon(s))$ the coordinates of $\dot{l}(s)\in T_{l(s)}M_\varepsilon$ in the basis $X_j(l(s))$, $Y_j(l(s))$, Z_ε and by $(l_1(s),\dots,l_{2n+1}(s))$ the usual cartesian coordinates of $l(s)$, i.e., when expressed in the canonical basis of $M_\varepsilon\simeq\mathbb{R}^{2n+1}$. Simple computations show that $\dot{l}_j(s)$ coincides for all $j\in\{1,\dots,2n\}$ with the j-th usual cartesian coordinate of the derivative $\dot{l}(s)$ of $l(s)$ and that its $(2n+1)$-th cartesian coordinate $\dot{l}_{2n+1}(s)$ is related to $(\dot{l}_1(s),\dots,\dot{l}_{2n}(s),\dot{l}_\varepsilon(s))$ through the relation,

$$\dot{l}_\varepsilon(s) = \frac{1}{\varepsilon}\,(\dot{l}_{2n+1}(s) - 2\sum_{j=1}^n l_{n+j}(s)\,\dot{l}_j(s) - l_j(s)\,\dot{l}_{n+j}(s)).$$

Note that the distance d_ε is left invariant with respect to the translations on $\mathbb{H}^n$, each vector field X_j, Y_j, Z_ε being itself left invariant. On the other hand it is not hard to see that

$$d_\varepsilon(x,y) \le d_{\varepsilon'}(x,y) \le d_c(x,y) \tag{6.1}$$

for all x, $y \in M$ and $0 < \varepsilon' \le \varepsilon$. It turns out that $(\mathbb{H}^n, d_c)$ arises as the limit of the approximating Riemannian manifolds M_ε in the sense that

$$d_c(x,y) = \sup_{\varepsilon>0} d_\varepsilon(x,y)$$

for all x, $y \in \mathbb{H}^n$ (see [21]).

Going back to the mass transportation problem, we fix for the rest of this section two Borel probability measures μ and ν on M with μ absolutely continuous with respect to $\mathcal{L}^{2n+1}$ and satisfying (5.2). We set

$$c_\varepsilon(x,y) := \frac{1}{2} d_\varepsilon(x,y)^2 \quad \text{and} \quad c(x,y) := \frac{1}{2} d_c(x,y)^2.$$

By (5.2) we get

$$\int_M d_\varepsilon(x)^2 \, d\mu(x) + \int_M d_\varepsilon(y)^2 \, d\nu(y) < +\infty \, ,$$

and it follows from Mc Cann's results ([26]) that there exists a unique optimal transport map ψ_ε from μ to ν relative to the cost c_ε. We prove first that one can recover our optimal transport map ψ relative to the cost c as the limit of these transports ψ_ε.

Theorem 6.2. *Let μ and ν be two Borel probability measures with μ absolutely continuous with respect to $\mathcal{L}^{2n+1}$ and satisfying* (5.2)*. Then the optimal transport maps $\psi_\varepsilon : M_\varepsilon \to M_\varepsilon$ from μ to ν relative to the cost c_ε converge in μ-measure as $\varepsilon \to 0$ to the optimal transport $\psi : \mathbb{H}^n \to \mathbb{H}^n$ from μ to ν relative to the cost c.*

Proof. We consider $\Pi(\mu,\nu)$, the class of all transport plannings from μ to ν, equipped with the restriction of any distance inducing the weak-$\star$ topology in the duality with $C_b(M \times M)$, the set of all continuous and bounded functions on $M \times M$. Recall that for this topology $\Pi(\mu,\nu)$ is compact. For $\boldsymbol{\gamma} \in \Pi(\mu,\nu)$ and $\varepsilon > 0$, we set

$$F_\varepsilon(\boldsymbol{\gamma}) := \int_{M\times M} c_\varepsilon(x,y) \, d\boldsymbol{\gamma} \quad \text{and} \quad F(\boldsymbol{\gamma}) := \int_{M\times M} c(x,y) \, d\boldsymbol{\gamma}$$

and we first prove that any limit point of optimal transport plannings from μ to ν relative to the cost c_ε, i.e., minimizers of F_ε on $\Pi(\mu,\nu)$, is an optimal transport planning from μ to ν relative to the cost c. This will follow from the Γ-convergence in $\Pi(\mu,\nu)$ of F_ε to F.

We prove first the lim inf inequality in the definition of Γ-convergence. We take $\boldsymbol{\gamma} \in \Pi(\mu,\nu)$ and a sequence $(\boldsymbol{\gamma}_\varepsilon)$ in $\Pi(\mu,\nu)$ which converges to $\boldsymbol{\gamma}$. Let $\delta > 0$ be fixed, $U \subset M \times M$ be a bounded open set and $\varphi \in C_c(M \times M)$ be such that $0 \le \varphi \le 1$ and $\varphi \equiv 1$ on U. As $c_\varepsilon \ge c_\delta$ whenever $\varepsilon \le \delta$ we have

$$\liminf_{\varepsilon\to 0} F_\varepsilon(\boldsymbol{\gamma}_\varepsilon) \ge \liminf_{\varepsilon\to 0} \int_{M\times M} \varphi(x,y) \, c_\delta(x,y) \, d\boldsymbol{\gamma}_\varepsilon \ge \int_U c_\delta(x,y) \, d\boldsymbol{\gamma}.$$

Letting $U \uparrow M \times M$ and $\delta \downarrow 0$ we get

$$\liminf_{\varepsilon\to 0} F_\varepsilon(\boldsymbol{\gamma}_\varepsilon) \ge F(\boldsymbol{\gamma}).$$

Next by the monotone convergence theorem it is obvious that $\lim_{\varepsilon\to 0} F_\varepsilon(\boldsymbol{\gamma}) = F(\boldsymbol{\gamma})$ for all $\boldsymbol{\gamma} \in \Pi(\mu,\nu)$. This proves the Γ-convergence of F_ε to F and it follows from the compactness of $\Pi(\mu,\nu)$ that any limit point of minimizers for F_ε is a minimizer for F (see for instance [11]). On the other hand one knows from Theorem 5.1 that there is a unique minimizer $\boldsymbol{\gamma}$ for F given by $\boldsymbol{\gamma} = (Id \times \psi)_{\#}\mu$, where ψ is the unique optimal transport map from μ to ν relative to the cost c. In particular it follows that $\boldsymbol{\gamma}_\varepsilon = (Id \times \psi_\varepsilon)_{\#}\mu$ converges to $\boldsymbol{\gamma}$.

Let us now prove that this implies the convergence of ψ_ε to ψ in μ-measure, i.e., that for all $\delta > 0$,

$$\lim_{\varepsilon\to 0} \mu(\{x \in M\,;\ d_c(\psi_\varepsilon(x),\psi(x)) \geq \delta\}) = 0.$$

We argue here as in [3]. Let $\delta > 0$ be fixed and fix also some $\eta > 0$. Let $\tilde{\psi} \in C_c(M)$ be such that

$$\mu(\{x \in M\,;\ \psi(x) \neq \tilde{\psi}(x)\}) \leq \eta/4.$$

Take a compact set $K \subset \{x \in M\,;\ \psi(x) = \tilde{\psi}(x)\}$ such that $\mu(M \setminus K) \leq \eta/2$, $U \subset M$ open, containing K and such that $\mu(U \setminus K) \leq \eta/2$, and finally $\varphi \in C_c(M)$, $0 \leq \varphi \leq 1$, such that $\varphi \equiv 1$ on K and $\operatorname{spt}\varphi \subset U$. Let $g \in C(M)$, $0 \leq g \leq 1$, be such that $g(0) = 0$ and $g(x) = 1$ if $d_c(0,x) \geq \delta$ and set $f(x,y) = \varphi(x)\ g(y^{-1}\cdot\tilde{\psi}(x))$. Then $f \in C_b(M\times M)$ and

$$\begin{aligned}
\mu(\{x \in M\,;\ d_c(\psi_\varepsilon(x),\psi(x)) \geq \delta\}) &\leq \eta/2 + \mu(\{x \in K\,;\ d_c(\psi_\varepsilon(x),\tilde{\psi}(x)) \geq \delta\})\\
&\leq \eta/2 + \int_M f(x,\psi_\varepsilon(x))\,d\mu\\
&= \eta/2 + \int_{M\times M} f(x,y)\,d\boldsymbol{\gamma}_\varepsilon.
\end{aligned}$$

It follows

$$\begin{aligned}
\limsup_{\varepsilon\to 0} \mu(\{x \in M\,;\ d_c(\psi_\varepsilon(x),\psi(x)) \geq \delta\}) &\leq \eta/2 + \int_{M\times M} f(x,y)\,d\boldsymbol{\gamma}\\
&= \eta/2 + \int_M \varphi(x)\ g((\psi(x))^{-1}\cdot\tilde{\psi}(x))\,d\mu\\
&\leq \eta/2 + \mu(U\setminus K) \leq \eta.
\end{aligned}$$

We conclude letting $\eta \downarrow 0$. □

In [26], R. Mc Cann gives moreover a characterization of the optimal transport map ψ_ε in terms of some c_ε-concave potential whose gradient -and its norm- specifies the direction -and how far- to move the mass geodesically in M_ε. In our present setting this means that

$$\psi_\varepsilon(x) = \exp_x^\varepsilon(-\nabla^\varepsilon\varphi_\varepsilon(x)) \tag{6.3}$$

for μ-a.e. $x \in M_\varepsilon$, where $\exp_x^\varepsilon$ is the Riemannian exponential of M_ε,

$$\nabla^\varepsilon\varphi_\varepsilon(x) = \sum_{j=1}^n X_j\varphi_\varepsilon(x)X_j + Y_j\varphi_\varepsilon(x)Y_j + Z_\varepsilon\varphi_\varepsilon(x)Z_\varepsilon \in T_xM_\varepsilon$$

is the gradient of φ_ε and φ_ε is a Kantorovich potential, i.e., a c_ε-concave function such that $\boldsymbol{\gamma}$ is optimal if and only if

$$c_\varepsilon(x,y) = \varphi_\varepsilon(x) + \varphi_\varepsilon^c(y) \qquad \boldsymbol{\gamma}\text{-a.e.} \tag{6.4}$$

A byproduct of Mc Cann's argument is also the fact that the curve $s \mapsto \exp_x^\varepsilon(-s\nabla^\varepsilon \varphi_\varepsilon(x))$ is a minimal geodesic in $[0,1]$ for μ-a.e. x. We investigate now the asymptotic behaviour of the Riemannian exponential maps $\exp^\varepsilon$ and of these potentials, our aim being to get an analogue of the representation (6.3) in the sub-Riemannian setting.

Note that, in the approximating Riemannian context, using the left translations in $\mathbb{H}^n$, one can always take without loss of generality 0 as a base point when computing the exponential map $\exp^\varepsilon$. Indeed, for any $x \in M_\varepsilon$ and $w = \sum_{j=1}^n w_j X_j(x) + w_{n+j} Y_j(x) + w_{2n+1} Z_\varepsilon \in T_x M_\varepsilon$, we have

$$\exp_x^\varepsilon w = x \cdot \exp_0^\varepsilon v \tag{6.5}$$

where $v = \sum_{j=1}^n w_j X_j(0) + w_{n+j} Y_j(0) + w_{2n+1} Z_\varepsilon \in T_0 M_\varepsilon$. This follows from the left invariance of the distance d_ε and from the fact that if $l : [0,1] \to M_\varepsilon$ is a C^1 curve and $\tilde{l} : [0,1] \to M_\varepsilon$ is defined by $\tilde{l}(s) = x^{-1} \cdot l(s)$ then

$$\dot{\tilde{l}}(s) = \sum_{j=1}^n \dot{l}_j(s) X_j(\tilde{l}(s)) + \dot{l}_{n+j}(s) Y_j(\tilde{l}(s)) + \dot{l}_\varepsilon(s) Z_\varepsilon .$$

Next we characterize the exponential map $\exp_0^\varepsilon$ and the geodesic equations using locally minimal geodesics in M_ε.

Proposition 6.6. *Let $\varepsilon > 0$ be fixed. Let $v = \sum_{j=1}^n v_j X_j(0) + v_{n+j} Y_j(0) + v_{2n+1} Z_\varepsilon \in T_0 M_\varepsilon$. Then $\exp_0^\varepsilon v = l(1)$, where $l : [0,1] \to M_\varepsilon$ is the unique curve such that $l(0) = 0$ and*

$$\begin{aligned} \dot{l}_j(s) &= \frac{4}{\varepsilon}\, v_{2n+1}\, l_{n+j}(s) + v_j \,, \\ \dot{l}_{n+j}(s) &= -\frac{4}{\varepsilon}\, v_{2n+1}\, l_j(s) + v_{n+j} \,, \\ \dot{l}_\varepsilon(s) &= v_{2n+1} \end{aligned}$$

for all $s \in [0,1]$.

Proof. First recall that since M_ε is connected and $(M_\varepsilon, d_\varepsilon)$ is complete, the map $\exp_0^\varepsilon$ is well defined everywhere on $T_0 M_\varepsilon$. Next it is well known that, being any geodesic in M_ε locally distance minimizing, any geodesic with constant speed locally minimizes the L^2 energy of curves as well: if $v \in T_0 M_\varepsilon$ and $l : [0,1] \to M_\varepsilon$ denotes the geodesic starting from 0 with velocity $\dot{l}(0) = v$, then for all $t \in (0,1)$ one can find $\delta = \delta(t) > 0$ such that

$$\int_{t-\delta}^{t+\delta} |\dot{l}(s)|_\varepsilon^2 \, ds = \min \int_{t-\delta}^{t+\delta} |\dot{\tilde{l}}(s)|_\varepsilon^2 \, ds \tag{6.7}$$

where the minimum is amoung all piecewise C^1 curves $\tilde{l} : [t-\delta, t+\delta] \to M_\varepsilon$ such that $\tilde{l}(t-\delta) = l(t-\delta)$ and $\tilde{l}(t+\delta) = l(t+\delta)$.

Rewriting $|\dot{l}_\varepsilon(s)|_\varepsilon$ in cartesian coordinates and doing standard first variations of the form $\eta \in \mathbb{R} \mapsto l + \eta\, \varphi\, e_j$, where $\varphi \in C_c^1((t-\delta, t+\delta))$ and $j \in \{1, \dots, 2n+1\}$, we get from (6.7) that

$$\begin{aligned} \frac{d}{ds}\left(\dot{l}_j - \frac{2}{\varepsilon}\, \dot{l}_\varepsilon\, l_{n+j}\right) &= \frac{2}{\varepsilon}\, \dot{l}_\varepsilon\, \dot{l}_{n+j}, \\ \frac{d}{ds}\left(\dot{l}_{n+j} + \frac{2}{\varepsilon}\, \dot{l}_\varepsilon\, l_j\right) &= -\frac{2}{\varepsilon}\, \dot{l}_\varepsilon\, \dot{l}_j \end{aligned}$$

for all $j \in \{1, \dots, n\}$ and

$$\frac{d}{ds}(\dot{l}_\varepsilon) = 0$$

on a neighbourhood of all $t \in (0,1)$, hence on $(0,1)$ and finally on $[0,1]$ by continuity. It follows in particular that $\dot{l}_\varepsilon$ is constant and thus $\dot{l}_\varepsilon(s) = \dot{l}_\varepsilon(0) = v_{2n+1}$ for all $s \in [0,1]$. Next we get from the two first families of equations that for all $j \in \{1, \dots, n\}$ one can find two constants α_j and α_{n+j} such that

$$\begin{aligned} \dot{l}_j(s) - \frac{2}{\varepsilon}\, v_{2n+1}\, l_{n+j}(s) &= \frac{2}{\varepsilon}\, v_{2n+1}\, l_{n+j}(s) + \alpha_j\,, \\ \dot{l}_{n+j}(s) + \frac{2}{\varepsilon}\, v_{2n+1}\, l_j(s) &= -\frac{2}{\varepsilon}\, v_{2n+1}\, l_j(s) + \alpha_{n+j} \end{aligned}$$

for all $s \in [0,1]$. Since $l(0) = 0$ and $\dot{l}(0) = v$ we get that $\alpha_j = v_j$ for all $j \in \{1, \dots, 2n\}$ and the Lemma follows. $\square$

The connexion between these exponential coordinates and the map $\exp_{\mathbb{H}}$ is now given by the following lemma.

Lemma 6.8. *Let $v^\varepsilon = \sum_{j=1}^n v_j^\varepsilon X_j(0) + v_{n+j}^\varepsilon Y_j(0) + v_{2n+1}^\varepsilon Z_\varepsilon \in T_0 M_\varepsilon$ be such that the geodesic $l^\varepsilon(s) := \exp_0^\varepsilon(sv^\varepsilon)$ is minimal in $[0,1]$ and assume that $\lim_{\varepsilon\to 0} \exp_0^\varepsilon v^\varepsilon = x$ for some $x \in \mathbb{H}^n \setminus \{0\}$. Let $(A+iB, w) \in \mathbb{C}^n \setminus \{0\} \times [-\pi/2, \pi/2]$ be such that $x = \exp_{\mathbb{H}}(A+iB, w)$. Then we have*

$$\lim_{\varepsilon\to 0} \frac{v_{2n+1}^\varepsilon}{\varepsilon} = w.$$

If moreover $x \in \mathbb{H}^n \setminus L$, we also have

$$\lim_{\varepsilon\to 0} v_j^\varepsilon = A_j\,, \quad \lim_{\varepsilon\to 0} v_{n+j}^\varepsilon = B_j\,.$$

Proof. First we notice that $|v^\varepsilon|_\varepsilon = d_\varepsilon(0, \exp_0^\varepsilon v^\varepsilon)$ is bounded hence $(v_j^\varepsilon)_{\varepsilon>0}$ is bounded for all $j \in \{1, \cdots, 2n+1\}$. Set $l^\varepsilon(s) := \exp_0^\varepsilon(sv^\varepsilon)$, $s \in [0,1]$. We show first that $v_{2n+1}^\varepsilon = O(\varepsilon)$. Indeed, by integrating the ODE given by Proposition 6.6 we get

$$\begin{aligned} l_j^\varepsilon(s) &= \frac{\varepsilon\, v_{n+j}^\varepsilon}{4\, v_{2n+1}^\varepsilon}\left(1 - \cos\frac{4\, v_{2n+1}^\varepsilon\, s}{\varepsilon}\right) + \frac{\varepsilon\, v_j^\varepsilon}{4\, v_{2n+1}^\varepsilon}\sin\frac{4\, v_{2n+1}^\varepsilon\, s}{\varepsilon}\,, \\ l_{n+j}^\varepsilon(s) &= -\frac{\varepsilon\, v_j^\varepsilon}{4\, v_{2n+1}^\varepsilon}\left(1 - \cos\frac{4\, v_{2n+1}^\varepsilon\, s}{\varepsilon}\right) + \frac{\varepsilon\, v_{n+j}^\varepsilon}{4\, v_{2n+1}^\varepsilon}\sin\frac{4\, v_{2n+1}^\varepsilon\, s}{\varepsilon} \\ l_{2n+1}^\varepsilon(s) &= \varepsilon\, v_{2n+1}\, s + 2\left(\frac{\varepsilon}{4\, v_{2n+1}^\varepsilon}\right)^2 \left(\frac{4\, v_{2n+1}^\varepsilon\, s}{\varepsilon} - \sin\frac{4\, v_{2n+1}^\varepsilon\, s}{\varepsilon}\right) \sum_{j=1}^{2n} (v_j^\varepsilon)^2 \end{aligned}$$

for any $s \in [0,1]$ and $j \in \{1, \dots, n\}$. In particular, if by contradiction $|v_{2n+1}^{\varepsilon_i}|/\varepsilon_i \to \infty$ for some $\varepsilon_i \to 0$, we obtain that $l_j^{\varepsilon_i}(s) \to 0$ for all $j \in \{1, \dots, 2n+1\}$, hence $x = \lim_i l^{\varepsilon_i}(1) = 0$.

Let $v = (v_1, \dots, v_{2n}, 0)$ be any limit point of v^ε, along some sequence ε_i, and assume that $v_{2n+1}^{\varepsilon_i}/\varepsilon_i \to v_{2n+1}$. The monotone convergence of d_ε to d_c ensures that the limit curve

$l : [0,1] \to \mathbb{H}^n$ satisfies $l(0) = 0$, $l(1) = x$,

$$l_j(s) = \frac{v_{n+j}\,(1-\cos(4\,v_{2n+1}\,s)) + v_j\,\sin(4\,v_{2n+1}\,s)}{4\,v_{2n+1}},$$

$$l_{n+j}(s) = \frac{-v_j\,(1-\cos(4\,v_{2n+1}\,s)) + v_{n+j}\,\sin(4\,v_{2n+1}\,s)}{4\,v_{2n+1}},$$

$$\dot{l}_{2n+1}(s) = 2\sum_{j=1}^{n} l_{n+j}(s)\,\dot{l}_j(s) - l_j(s)\,\dot{l}_{n+j}(s)$$

and is a minimal geodesic with respect to the d_c metric. A simple differentiation in time also shows that l has a constant speed, which must be equal to $r = d_c(0,x)$. In other words, setting $\gamma(s) = l(s/r)$, then $\gamma : [0,r] \to \mathbb{H}^n$ is a sub-unit minimal geodesic in $(\mathbb{H}^n, d_c)$ joining 0 to x. Alternatively, with the terminology of §3, one can also recognize that γ is a curve with parameter $((v_j + iv_{n+j})/r, 4\,v_{2n+1}, r)$. Then it follows from the description of sub-unit minimal geodesics given in Theorem 3.4 and the definition of $\exp_{\mathbb{H}}$ that one must have $v_{2n+1} = w$. Note that when $x \in L^*$ then $A + iB$ is not uniquely determined and one cannot say more about v_j for $j \in \{1, \dots, 2n\}$. On the other hand when $x \in \mathbb{H}^n \setminus L$ it follows from the uniqueness and explicit description of sub-unit minimal geodesic from 0 to x that one must have $v_j = A_j$ and $v_{n+j} = B_j$ for all $j \in \{1, \dots, n\}$. □

We now turn our attention to the c_ε-concave potentials φ_ε, showing that any limit potential φ is c-concave.

Lemma 6.9. *Let $Y \subset \mathbb{H}^n$ be a bounded set with respect to d_c. For $\varepsilon > 0$, let $\phi_\varepsilon : Y \to \mathbb{R} \cup \{+\infty\}$, $\phi_\varepsilon \not\equiv +\infty$. Let us define*

$$\varphi_\varepsilon(x) = \inf_{y \in Y} c_\varepsilon(x,y) + \phi_\varepsilon(y)$$

and assume that $\varphi_\varepsilon(0) = 0$. Then the family $(\varphi_\varepsilon)_{\varepsilon > 0}$ is locally uniformly bounded in the space of Lipschitz functions on $(\mathbb{H}^n, d_c)$ and hence is relatively compact as $\varepsilon \to 0$ with respect to the local uniform convergence. Moreover any limit point φ is c-concave.

Proof. Let $K \subset \mathbb{H}^n$ be any bounded set. It follows from (6.1) that Y_ε is bounded with respect to d_ε and then we know by Lemma 4.5 that each φ_ε is Lipschitz on K with respect to the distance d_ε with Lipschitz constant less than $\sup_{y \in Y,\, z \in K} d_\varepsilon(y,z)$. Using (6.1) one gets that φ_ε is Lipschitz on K with respect to the distance d_c with Lipschitz constant less than $\sup_{y \in Y,\, z \in K} d_c(y,z)$. Then the first claim follows by Ascoli–Arzelá theorem, remembering the normalization $\varphi_\varepsilon(0) = 0$.

Let now φ be a limit point of φ_ε, along some sequence $\varepsilon_i \to 0$. Although a direct and elementary proof could be given, we prove that φ is c-concave using some general tools from the theory of Γ-convergence. Without loss of generality we may assume that Y is compact and ϕ_{ε_i} are lower semicontinuous (if not, we replace Y by its closure and ϕ_{ε_i} by its lower semicontinuous envelope, defined on the whole closure). By the compactness properties of Γ-convergence (see for instance [11]) we can also assume that ϕ_{ε_i} Γ-converge as $i \to \infty$ to some lower semicontinuous function ϕ, and since $c_{\varepsilon_i}(x,\cdot) \to c(x,\cdot)$ uniformly on Y the functions $c_{\varepsilon_i}(x,\cdot) + \phi_{\varepsilon_i}$ Γ-converge to $c(x,\cdot) + \phi$. Therefore, the minimum values of these functions, i.e., φ_{ε_i}, converge to the minimum value of $c(x,\cdot) + \phi$. It follows that

$$\varphi(x) = \min_{y \in Y} c(x,y) + \phi(y) \tag{6.10}$$

is c-concave. $\square$

To conclude, the next Theorem gives the behaviour as ε goes to zero of the c_ε-concave potentials and justifies once more our description in terms of $\mathbb{H}$-exponential coordinates of the optimal transport map in $\mathbb{H}^n$. Notice that the gradients of φ_ε are uniquely determined μ-a.e., and therefore the functions φ_ε are uniquely determined up to an additive constant only if $\mathcal{L}^{2n+1} \ll \mu$. This is the reason why we only talk about convergence of gradients in the next theorem.

Theorem 6.11 (Convergence of gradients of dual potentials). *Let μ and ν be two Borel probability measures in $\mathbb{H}^n$ with μ absolutely continuous with respect to $\mathcal{L}^{2n+1}$, $\operatorname{spt}\nu$ compact and $\int_{\mathbb{H}^n} d_c(x)^2\, d\mu(x) < +\infty$. For $\varepsilon > 0$, let $\varphi_\varepsilon : M_\varepsilon \to \mathbb{R}$ be c_ε-concave dual potentials as in (6.4), normalized in such a way that $\varphi_\varepsilon(0) = 0$. Then there exists a locally Lipschitz and c-concave map $\varphi : \mathbb{H}^n \to \mathbb{R}$ such that*

$$X_j\varphi_\varepsilon \overset{L^p_{loc}(d\mu)}{\longrightarrow} X_j\varphi, \quad Y_j\varphi_\varepsilon \overset{L^p_{loc}(d\mu)}{\longrightarrow} Y_j\varphi \qquad \forall j \in \{1,\dots,n\}$$

for all $1 \le p < +\infty$ and $Z\varphi_\varepsilon$ converges in μ-measure to $Z\varphi$ on the set $F := \{x \in \mathbb{H}^n\,;\, \psi(x) \neq x\}$ where ψ is the optimal transport from μ to ν relative to the cost c. Any limit point φ of such c_ε-concave dual potentials φ_ε as $\varepsilon \downarrow 0$ has this property.

Proof. First recall that it follows from [26] that one can find $Y \subset \operatorname{spt}\nu$ and for any $\varepsilon > 0$ some $\phi_\varepsilon : Y \to \mathbb{R}\cup\{+\infty\}$ such that $\varphi_\varepsilon(x) = \inf_{y\in Y} c_\varepsilon(x,y) + \phi_\varepsilon(y)$. It follows from Lemma 6.9 that the family $(\varphi_\varepsilon)_{\varepsilon>0}$ is relatively compact with respect to the local uniform convergence. Then let φ be a limit point of $(\varphi_\varepsilon)_{\varepsilon>0}$ along some sequence $\varepsilon_k \to 0$ so that $\varphi_k := \varphi_{\varepsilon_k}$ converges to φ locally uniformly on $\mathbb{H}^n$. Note that φ is a locally Lipschitz function on $\mathbb{H}^n$ equipped with the distance d_c, that φ is c-concave by Lemma 6.9 and more precisely satisfies (6.10) for some suitable ϕ. Note also that, according to Lemma 6.9 and Theorem 2.5, $X_j\varphi_k$, $Y_j\varphi_k$, $X_j\varphi$ and $Y_j\varphi$ are well-defined $\mathcal{L}^{2n+1}$-a.e. on $\mathbb{H}^n$ and belong to $L^\infty_{loc}(\mathbb{H}^n, dx)$ with

$$\sup_k(\|X_j\varphi_k\|_{L^\infty(K,dx)}, \|Y_j\varphi_k\|_{L^\infty(K,dx)}) < +\infty \tag{6.12}$$

for any compact set K and $j \in \{1,\dots,n\}$. Using Theorem 6.2 and passing to a subsequence if necessary one can also assume that the optimal transport maps

$$\psi_k(x) := x \cdot \exp_0^\varepsilon(-\nabla^{\varepsilon_k}\varphi_k(x))$$

relative to c_{ε_k} converge μ-a.e. to the optimal transport map ψ relative to c. On the other hand, as stressed just before, φ satisfies (5.4) and since $\operatorname{spt}\nu$ is compact one also have (5.5), hence one gets from the characterization of minimality in Theorem 5.1 that

$$\psi(x) = x \cdot \exp_{\mathbb{H}}(-X\varphi(x) - iY\varphi(x), -Z\varphi(x))$$

for μ-a.e. x. Then, using Lemma 6.8 and the fact that $x^{-1}\cdot\psi(x) \notin L$ for μ-a.e. $x \in F$ (remember Lemma 4.7) we get that

$$\lim_{k\to+\infty} X_j\varphi_k(x) = X\varphi(x)\,, \quad \lim_{k\to+\infty} Y_j\varphi_k(x) = Y\varphi(x)\,, \quad \lim_{k\to+\infty} Z\varphi_k(x) = Z\varphi(x)\,, \tag{6.13}$$

for μ-a.e. $x \in F$. In particular it follows that $Z\varphi_\varepsilon$ converges in μ-measure to $Z\varphi$ on F. On $\mathbb{H}^n \setminus F$ we notice that $X\varphi + iY\varphi = 0$ and since

$$|X\varphi_k(x) + iY\varphi_k(x)| \le d_{\varepsilon_k}(0, x^{-1}\cdot\psi_k(x))$$

the μ-a.e. convergence of ψ_k to ψ yields the μ-a.e. convergence of $X\varphi_k$ to $X\varphi$ and of $Y\varphi_k$ to $Y\varphi$ also on $\mathbb{H}^n \setminus F$. By (6.12) we obtain the local $L^p(d\mu)$-convergence of the horizontal derivatives remembering that μ is absolutely continuous with respect to $\mathcal{L}^{2n+1}$. □

We have not been able to prove convergence of $Z\varphi_\varepsilon$ on $\{\psi(x) = x\}$, due to the fact that the exponential coordinates of the identity are not uniquely defined (since $\exp_{\mathbb{H}}(0, w) = 0$ for any $w \in [-\pi/2, \pi/2]$). The convergence of $Z\varphi_\varepsilon$ on this set presumably requires an analysis similar to the one made in Lemma 4.13 for the distance d_g^2.

7. Extensions and open problems

In this section we discuss some open problems and possible extensions of our results.

(a). More general spaces and cost functions. Our results still hold, with minor modifications, if $c = d^2/2$ is replaced by $c = d^p/p$ for some $p \geq 2$ (here $d = d_c$ or $d = d_g$). In the case $c = d_c^p/p$ the formula for the optimal transport map becomes

$$\psi(x) = x \cdot \exp_{\mathbb{H}} \left(-|\nabla_{\mathbb{H}} \varphi|^{\frac{2-p}{p-1}} (X\varphi + iY\varphi), -|\nabla_{\mathbb{H}} \varphi|^{\frac{4-2p}{p-1}} Z\varphi \right) \qquad \mu\text{-a.e.}$$

Concerning the extension of our results to general groups, we already pointed out that some of our results depend on explicit computation of geodesics. In particular it would be interesting to know whether the property that $Zd_c^2 \in L^\infty$ has a more general validity, at least in step 2 Carnot groups. More generally, one can look for general semiconcavity estimates for d_c^2 along the horizontal directions in general Carnot groups.

(b). Regularity of the optimal transport map. Assume that $\mu = f\mathcal{L}^{2n+1}$ and $\nu = g\mathcal{L}^{2n+1}$. In the Euclidean case the optimal transport map ψ is the gradient of a convex real-valued function, the function $\phi := \frac{1}{2}|x|^2 - \varphi$ (here φ is the Kantorovich potential). Since gradients of convex functions are differentiable a.e. (see for instance [1], where these properties are studied in a more general framework), one can use the area formula (see for instance [25]) to write the condition $\psi_{\#}\mu = \nu$ in a pointwise form:

$$\det(\nabla^2 \phi) = \det(\nabla \psi) = \frac{g(\nabla \phi)}{f}$$

Moreover, Caffarelli's regularity theory [9, 10] for the Monge–Ampére equation yields the smoothness of ψ, provided f, g are smooth and $\operatorname{spt} g$ is convex.

All these facts are still to be understood in $\mathbb{H}^n$. In particular we don't know whether the map ψ is sufficiently regular to apply the area formula in Carnot groups, recently established in [27], [37]. This problem might be related to the convexity properties recently studied in [12] and [24] and with the formulation of the Monge-Ampere equation in $\mathbb{H}^n$.

(c). Non-linear interpolation in $\mathbb{H}^n$. Let $\mathcal{P}_2(\mathbb{H}^n)$ the space of probability measures in $\mathbb{H}^n$ with finite second order moments, endowed with the Wasserstein metric

$$W_2(\mu, \nu) := \left(\min_\gamma \int_{\mathbb{H}^n \times \mathbb{H}^n} \frac{1}{2} d_c^2(x, y) \, d\gamma \right)^{1/2}$$

(here γ varies in the family $\Pi(\mu, \nu)$ of admissible transport plannings). If $\mu \ll \mathcal{L}^{2n+1}$ and $\psi = \exp_{\mathbb{H}}(-X\varphi + iY\varphi, -Z\varphi)$ is the optimal transport map between μ and ν, one can easily show that the family of measures

$$\mu_t := \psi_{t\#}\mu \quad \text{with} \quad \psi_t(x) := x \cdot \exp_{\mathbb{H}}(-tX\varphi - itY\varphi, -t^2 Z\varphi)$$

is a geodesic in $\mathcal{P}_2(\mathbb{H}^n)$ connecting μ and ν, i.e. $W_2(\mu,\mu_t)+W_2(\mu_t,\nu)=W_2(\mu,\nu)$ for any $t\in[0,1]$. It is not presently clear whether, as in the Euclidean case, all measures μ_t for $t\in[0,1)$ are absolutely continuous.

(d). Relation between the two distances $\mathbf{d_c}$ and $\mathbf{d_g}$. By definition d_c is a geodesic metric, i.e. for any pair of points x, $y\in\mathbb{H}^n$ there exists a Lipschiz curve connecting x and y whose length, computed with respect to d_c, is $d_c(x,y)$. The relation between d_c and d_g is discussed in [22], where it is shown that the length of any Lipschitz curve computed with respect to d_g coincides with the length computed with respect to d_c (the heuristic reason for that is that infinitesimally both metrics have the same behaviour in the horizontal directions). In particular this implies that the geodesic metric $\tilde{d}_g$ associated to d_g, defined by

$$\tilde{d}_g(x,y):=\inf\left\{\mathrm{length}_{d_g}(\gamma):\ \gamma:[0,1]\to\mathbb{H}^n\ \text{Lipschitz},\ \gamma(0)=x,\ \gamma(1)=y\right\}$$

coincides with d_c. While the choice of d_c is more natural, especially in connection with the Riemannian approximation, explicit computations are usually much simpler with d_g. It is for these reasons that both metrics have been considered in this paper.

Acknowledgement. The first author thanks F.Otto for a useful conversation on the subject of this paper. The second author wishes to thank the hospitality of the Scuola Normale Superiore where part of the work was done.

References

[1] G. Alberti and L. Ambrosio, *A geometric approach to monotone functions in* $\mathbf{R}^n$, Matematische Zeitschrift, **230** (1999), 259–316.

[2] L. Ambrosio and V. Magnani, *Some fine properties of BV functions on stratified groups*, to appear on Math. Z.

[3] L. Ambrosio and A. Pratelli, *Existence and stability results in the* L^1 *theory of optimal transportation*, to appear on the proceedings of the CIME course in Martina Franca (September 2001), Lecture Notes in Mathematics.

[4] Z.M. Balogh and J. Tyson, *Polar coordinates in Carnot groups*, preprint 2001.

[5] A. Bellaïche, *The tangent space in sub-Riemannian geometry*, in Sub-Riemannian geometry, Progr. Math., vol. 144, Birkhäuser, Basel, 1996, 1–78.

[6] Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs,* C.R. Acad. Sci. Paris, Sér I Math., **305** (1987), 805–808.

[7] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions,* Comm. Pure Appl. Math., **44** (1991), 375–417.

[8] L. Caffarelli, *Boundary regularity of maps with a convex potential,* Commun. Pure Appl. Math., **45** (1992), 1141–1151.

[9] L. Caffarelli, *A localization property of viscosity solutions to the Monge–Ampére equation,* Ann. of Math., **131** (1990), 129–134.

[10] L. Caffarelli, *Interior* $W^{2,p}$ *estimates for the solutions of the Monge–Ampére equation,* Ann. of Math., **131** (1990), 135–150.

[11] G. Dal Maso, *An introduction to* Γ*-convergence*, Birkhäuser, 1993.

[12] D. Danielli, D. Nhieu and N. Garofalo, *Notions of convexity in Carnot groups,* Preprint, 2002.

[13] H. Dietrich, *Zur c-konvexität und c-Subdifferenzierbarkeit von Funktionalen,* Optimization, **19** (1988), 355–371.

[14] L.C. Evans, *Partial Differential Equations and Monge–Kantorovich Mass Transfer,* Current Developments in Mathematics, 1997, 65–126.

[15] G.B. Folland, *A fundamental solution for a subelliptic operator,* Bull. Amer. Math. Soc., **79** (1973), 373–376.

[16] W. Gangbo, *An elementary proof of the polar factorization theorem for functions,* Arch. Rat. Mech. Anal., **128** (1994), 381–399.

[17] W. Gangbo and R.J. McCann, *The geometry of optimal transportation,* Acta Math., **177** (1996), 113–161.
[18] N. Garofalo and D.M. Nhieu, *Lipschitz continuity, global smooth approximations and extension theorems for Sobolev functions in Carnot-Carathéodory spaces*, J. Anal. Math. 74 (1998), 67–97.
[19] 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.
[20] M. Gromov, *Structures métriques pour les variétés riemaniennes*, CEDIC, Paris, 1981
[21] D. Jerison and A. Sanchez Calle, *Subelliptic, second order differential operators*, in Complex analysis, III (College Park, Md., 1985–86), 46–77, Lecture Notes in Math., 1277, Springer, Berlin, 1987.
[22] A. Korányi, *Geometric properties of Heisenberg-type groups,* Adv. in Math., **56** (1985), 28–38.
[23] A. Korànyi and H.M. Reimann, *Quasiconformal mappings on the Heisenberg group*, Invent. Math. 80 (1985), no. 2, 309–338.
[24] G. Lu, J. Manfredi and B. Stroffolini, *Convex functions in the Heisenberg group,* Preprint, 2002.
[25] R. McCann, *A convexity principle for interacting gases,* Adv. Math., **128** (1997), 153–179.
[26] R. McCann, *Polar factorization of maps on Riemannian manifolds,* Geom. Funct. Anal., **11** (2001), 589–608.
[27] V. Magnani, *Differentiability and area formula on stratified Lie groups,* Houston Math. J., **27** (2001), 297–323.
[28] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, 91. American Mathematical Society, Providence, RI, 2002. xx+259 pp.
[29] R. Monti, *Some properties of Carnot-Carathéodory balls in the Heisenberg group*, Rend. Mat. Acc. Lincei 11 (2000), 155–167.
[30] R. Monti and F. Serra Cassano, *Surfaces measures in some CC spaces,* Calc. Var. Partial Differential Equations **13** (2001), 339–376.
[31] P. Pansu, *Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un*, Ann. of Math. (2) 129 (1989), no. 1, 1–60.
[32] S.T. Rachev and L. Rüschendorf, *A characterization of random variables with minimum L^2 distance,* J. Multivariate Anal., **32** (1990), 48–54.
[33] S.T. Rachev and L. Rüschendorf, *Mass transportation problems,* Vol I: Theory, Vol. II: Applications. Probability and its applications, Springer, 1998.
[34] L. Rüschendorf, *On c-optimal random variables,* Statistics & Probability Letters, **27** (1996), 267–270.
[35] C. Smith and M. Knott, *On the optimal transportation of distributions,* J. Optim. Theory Appl., **52** (1987), 323–329.
[36] C. Smith and M. Knott, *On Hoeffding–Fréchet bounds and cyclic monotone relations,* J. Multivariate Anal., **40** (1992), 328–334.
[37] A.D. Ukhlov and S.Vodop'yanov, *Approximately differentiable transformations and change of variables on nilpotent groups,* Siberian Math. J., **37** (1996), 62–78.
[38] C. Villani, *Topics in mass transportation,* forthcoming book by AMS.

(L. AMBROSIO) Scuola Normale Superiore, Piazza Cavalieri 7, 56100 Pisa, Italy
E-mail address: `luigi@ambrosio.sns.it`

(S. RIGOT) Université de Paris-Sud, Mathématiques, Bâtiment 425, 91405 Orsay cedex, France
E-mail address: `Severine.Rigot@math.u-psud.fr`