



HAL
open science

Diffusion by optimal transport in Heisenberg groups

Nicolas Juillet

► **To cite this version:**

Nicolas Juillet. Diffusion by optimal transport in Heisenberg groups. *Calculus of Variations and Partial Differential Equations*, 2014, 50 (3-4), pp.693 - 721. 10.1007/s00526-013-0652-2 . hal-00372216v2

HAL Id: hal-00372216

<https://hal.science/hal-00372216v2>

Submitted on 13 Apr 2018

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

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

DIFFUSION BY OPTIMAL TRANSPORT IN HEISENBERG GROUPS

NICOLAS JUILLET

ABSTRACT. We prove that the hypoelliptic diffusion of the Heisenberg group \mathbb{H}_n describes, in the space of probability measures over \mathbb{H}_n , a curve driven by the gradient flow of the Boltzmann entropy Ent , in the sense of optimal transport. We prove that conversely any gradient flow curve of Ent satisfy the hypoelliptic heat equation. This occurs in the subRiemannian \mathbb{H}_n , which is *not* a space with a lower Ricci curvature bound in the metric sense of Lott–Villani and Sturm. Heisenberg group and optimal transport theory and gradient flow

INTRODUCTION

For some years there has been an alternative representation for the evolution of probability densities. Beside the probabilistic diffusion point of view and the Dirichlet energy approach, one may now also consider this evolution as a curve in terms of optimal transport. The more representative class of examples is certainly given by the heat equation and its transformations. In different papers, covering different frameworks (see for instance [2, 11, 12, 20, 21, 27, 33, 36, 37, 42, 47]) it has been proved that curves of probability measures $(\mu_t)_{t \geq 0}$ with a density satisfying the adequate heat equation are exactly the curves with a speed equal to the opposite of the gradient of the relative Boltzmann entropy Ent , in the sense of optimal transport. Hence, we have the formal paradigm

$$(1) \quad \dot{\mu}_t = -\nabla \text{Ent}(\mu_t) \iff \dot{\rho}_t = \Delta \rho_t$$

where ρ_t on the right-hand side denotes the density of μ_t . Let us give a more precise description. At the origin of this stream is the seminal paper by Jordan, Otto and Kinderlehrer [24] where the Wasserstein space over \mathbb{R}^d , i.e. the space $\mathcal{P}_2(\mathbb{R}^d)$ of probability measures with a covariance matrix, is considered for the first time, formally as an infinite dimensional Riemannian manifold. This approach sometimes called “Otto calculus” as in [47, Chapter 15] made Otto and his coauthors realize that, at least at a formal level, the solutions of the heat equation are densities of measures that describe special curves in $\mathcal{P}_2(\mathbb{R}^d)$. The Boltzmann entropy¹ (with

2010 *Mathematics Subject Classification.* 28A33, 53C17, 60J60.

The author is partially supported by the “Programme ANR ProbaGeo” (ANR-09-BLAN-0364) and the “Programme ANR JCJC GMT” (ANR 2011 JS01 011 01).

¹The correct name would be “H functional of Boltzmann” because the entropy is actually the opposite, $-\int \rho \ln \rho$. A similar remark concerns the sign in front of the gradient in (1): we want to make a potential decrease, as in physics.

respect to a reference measure \mathcal{L} , in \mathbb{R}^d the Lebesgue measure), defined by

$$(2) \quad \text{Ent}(\mu) = \begin{cases} \int \rho \ln \rho \, d\mathcal{L} \in]-\infty, +\infty[& \text{if } \mu = \rho\mathcal{L} \\ +\infty & \text{if } \mu \text{ is not absolutely continuous,} \end{cases}$$

is a function on $\mathcal{P}_2(\mathbb{R}^d)$ because one has $\int (\rho \ln \rho)^- < +\infty$ if $\rho\mathcal{L} \in \mathcal{P}_2(\mathbb{R}^d)$. The entropy is relevant because the diffusion curve evolves with speed and direction determined by the gradient of this functional, formally identified with the vector field “ $\frac{\nabla \rho}{\rho}$ ”. This discovery initiated many studies on gradient flows of different functionals in the Wasserstein space over different spaces. Otto obtained for instance a representation of the porous medium equation [38] thanks to the Rényi entropy in $\mathcal{P}_2(\mathbb{R}^d)$. We have already stressed how successful the optimal transport approach was in representing the heat equation on various metric spaces (X, d, \mathcal{L}) . Those different contributions may adopt different degrees of generality, some of them giving rise to a precise variational analysis of metric spaces. For instance not only smooth curves are possible gradient flow curves, but also curves that are only absolutely continuous. Similarly the slope of Ent at μ is not only the maximal slope along regular curves starting at μ but rather the slope obtained from sequences converging to μ . The most documented book on gradient flows over Wasserstein spaces is certainly the book by Ambrosio, Gigli and Savaré [2] (one can also see [6] that is a kind of simplified version). Another important reference is the book by Villani [47, Chapters 23–25].

There seems to be a central ingredient in the proofs of paradigm (1). Namely, the *displacement convexity* of Ent (roughly speaking the geodesic convexity of Ent with respect to the metric structure of $\mathcal{P}_2(X)$), or its general version as the *K-displacement convexity* of Ent (for some $K \in \mathbb{R}$), offer a particularly good control on the modulus of continuity of the slope of Ent . This condition is so useful that it appears as a hypothesis of a theorem on the stability of gradient flows when the space is varying [19] and it is also the framework of a theory of gradient flows over very general spaces proposed by Ambrosio, Gigli and Savaré [3, 4]. The displacement convexity of Ent has a geometric interpretation: as it has been proved in the papers by Lott and Villani [32, 31] and Sturm [45, 46] (see also [47] for a complete account), one can consider the spaces where Ent is displacement convex as *spaces of non-negative Ricci curvature* (in a weak sense because Ricci curvature is a Riemannian notion) and those where it is *K-displacement convex* as *spaces with Ricci curvature bounded by K from below*. It is not surprising that curvature plays a role in the metric spaces for which paradigm (1) has been established up to now. For instance (1) has been proved for measures on Riemannian manifolds with Ricci curvature bounded from below [11] (– the condition coincides with the weak one by Lott–Villani and Sturm), and on Alexandrov spaces [36, 21, 20]: these spaces are considered as the *metric spaces with sectional curvature bounded from below* (a condition stronger than the displacement convexity of Ent , see [40, 48]).

Unfortunately the Wasserstein space $\mathcal{P}_2(\mathbb{H}_n)$ over the Heisenberg group does not satisfy the convexity of the entropy. This space \mathbb{H}_n may appear as the simplest Lie group after the Euclidean spaces because it involves only a few non-commutativity and its Lie algebra only presents n non-trivial relations (that can be interpreted in quantic physic as the uncertainty principle). The subRiemannian distance, called

Carnot-Carathéodory distance d_c provides an exotic structure that still allows elementary computations. The corresponding metric space (\mathbb{H}_n, d_c) is much studied (see *e.g.*, [17] for geometric measure theory, [28] for conformal geometry, [10] for embedding problems). We will prove that the gradient flow of Ent corresponds to the solutions of the adequate “heat equation”, which is the hypoelliptic diffusion provided by $\Delta_{\mathbb{H}}$, the subRiemannian “sum of square” operator (also called Kohn operator). Notice that the hypoellipticity of this operator is ensured by a famous theorem of Hörmander [23] about “sum of square” operators. In the case of the Heisenberg group, a probabilistic description of the heat distribution is available [18] because it appears in \mathbb{R}^{2n+1} as the joint distribution of n planar Brownian motions with the sum of their Lévy areas.

We want to prove paradigm (1) on \mathbb{H}_n . A formal proof using an adapted Otto calculus could be done easily (and can be found over this paper, see also Remark 5.1 on [27]) but we want to get an equivalence in precise terms, comparable to those of Ambrosio, Gigli and Savaré in [2]. It is not possible to mimic exactly the proof of these authors because of the lack of control on the Ricci curvature. Indeed, it was proved in [26] that the Heisenberg group \mathbb{H}_n is not a space with “Ricci curvature bounded from below” in the sense of Lott, Sturm and Villani. Nevertheless in the present article, we will obtain the complete equivalence between solutions of the “heat equation” in \mathbb{H}_n and the gradient flow of Ent in $\mathcal{P}_2(\mathbb{H}_n)$. It seems that this is the first exotic example of a metric space without a “lower Ricci bound” where one can prove (1). See Remark 5.2 for a more precise discussion on this topic.

We give now some elements of our strategy. The direct inclusion (heat diffusion is a gradient flow) is more or less the result of a direct computation. The more tricky part consists in proving the opposite inclusion. The Heisenberg group has a classical Riemannian approximation, \mathbb{H}_n^ε with a lower bound $\frac{-1}{2\varepsilon^2}$ on the Ricci curvature (see for instance [25] for the computation). In this paper, we will gather all the informations on $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$, a space where results on Ent, the locally absolutely continuous curves and their interactions are already known. This will be useful because \mathbb{H}_n^ε and \mathbb{H}_n are the same topological space and they have the same reference measure. A key result will be Proposition 3.4 (see also Proposition 3.2): it is proved that at points μ with a finite slope for Ent in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$, the slope of Ent at μ in $\mathcal{P}_2(\mathbb{H}_n)$ (that is also finite) equals the maximal slope along regular curves going through μ , which corresponds classically to the square root of the Fisher information of μ . This allows to obtain an important estimate (a strong upper gradient estimate) for regularized curve $(\mu_t^*)_{t \in [0, T]}$ of a gradient flow $(\mu_t)_{t \in [0, T]}$. The measures μ_t^* depend on a parameter $\nu > 0$ and are points of finite slope for Ent in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$. We obtain the result by letting ν go to 0.

The paper is organized as follows: in the first part we introduce the Heisenberg group, the optimal transport theory and state the definition of curves driven by the gradient flow of Ent.

In the second part we review some propositions on the relations between Ent and the Wasserstein distance on the approximating Riemannian manifolds \mathbb{H}_n^ε .

In part three we study the speed of the curves in $\mathcal{P}_2(\mathbb{H}_n)$ (Proposition 3.1) as well as the slope of Ent in this space (Proposition 3.2 and Proposition 3.4). For proving these propositions we use the results of Section 2. In particular, Proposition 3.4 is obtained by approximation using Proposition 2.1 from Section 2.

In the fourth part we prove the theorem (see Section 1 for the definitions, especially Definition 1.7 for the notion of gradient flow of Ent):

Theorem 0.1. *Let T be a positive or infinite time and $(\mu_t)_{t \in [0, T[}$ a locally absolutely continuous curve in $\mathcal{P}_2(\mathbb{H}_n)$. The following two statements are equivalent*

- *For any $t \in]0, T[$, μ_t has a density ρ_t with respect to the Lebesgue measure and $(\rho_t)_{t \in]0, T[}$ is a solution of the hypoelliptic heat equation of \mathbb{H}_n ,*

$$\Delta_{\mathbb{H}} \rho_t = \partial_t \rho_t,$$

where $\Delta_{\mathbb{H}} = \sum_{k=1}^n \mathbf{X}_k^2 + \mathbf{Y}_k^2$.

- *The curve $(\mu_t)_{t \in [0, T[}$ is a gradient flow of the entropy Ent in the sense of optimal transport.*

The strategy of the proof relies on two approximations. First \mathbb{H}_n is compared to the Riemannian manifold \mathbb{H}_n^ε , then we convolve the curve $(\mu_t)_{t \in [0, T[}$ in order to apply what we know on $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ to the measures μ_t^* , which is not possible directly for μ_t .

1. DEFINITIONS

1.1. Heisenberg groups. Let n be a positive integer. The Heisenberg group \mathbb{H}_n is a Lie group that can be represented by $\mathbb{R}^{2n+1} = \mathbb{C}^n \times \mathbb{R}$ with the multiplicative structure

$$(z, u) \cdot (z', u') = \left(z + z', u + u' - \frac{1}{2} \sum_{k=1}^n \Im(z_k \bar{z}'_k) \right)$$

where z writes $(z_1, \dots, z_n) = (x_1 + \mathbf{i}y_1, \dots, x_n + \mathbf{i}y_n)$ and \Im is the imaginary part of a complex number. A basis for the Lie algebra of left-invariant vector fields is given by $(\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n, \mathbf{U})$ where

$$\mathbf{X}_k = \partial_{x_k} - \frac{y_k}{2} \partial_u \quad \text{and} \quad \mathbf{Y}_k = \partial_{y_k} + \frac{x_k}{2} \partial_u \quad \text{for } k \in \{1, \dots, n\}$$

and $\mathbf{U} = \partial_u$. It is possible to define two metric structures on \mathbb{H}_n . For any $\varepsilon > 0$, we note $(\mathbb{H}_n^\varepsilon, d_\varepsilon)$ the Riemannian manifold with $(\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n, \varepsilon \mathbf{U})$ as orthonormal basis in any point. The metric space (\mathbb{H}_n, d_c) will denote the Heisenberg group with the subRiemannian metric with orthonormal frame $(\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n)$. The reference measure of the two spaces is the Lebesgue measure \mathcal{L}^{2n+1} . It is up to a constant the Haar measure of the group (both left- and right-invariant). Up to a constant it is for \mathbb{H}_n^ε the Riemannian volume and the $2n + 2$ -dimensional Hausdorff measure for \mathbb{H}_n . We will simply denote it by \mathcal{L} .

We now make precise what is the Carnot-Carathéodory distance, the subRiemannian distance of \mathbb{H}_n . The functions $\|\cdot\|_\varepsilon$ and $\|\cdot\|_{\mathbb{H}}$ are defined by

$$\begin{aligned} \left\| \left(\sum_{k=1}^n a_k \mathbf{X}_k + b_k \mathbf{Y}_k \right) + c \mathbf{U} \right\|_\varepsilon &= \sqrt{\left(\sum_{k=1}^n a_k^2 + b_k^2 \right) + \frac{c^2}{\varepsilon^2}} \\ \left\| \sum_{k=1}^n a_k \mathbf{X}_k + b_k \mathbf{Y}_k + c \mathbf{U} \right\|_{\mathbb{H}} &= \begin{cases} \sqrt{\sum_{k=1}^n a_k^2 + b_k^2} & \text{if } c = 0, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The distance d_c is obtained by minimizing the subRiemannian length between two points. This length $l_c(\gamma)$ is defined for locally absolutely continuous curves γ in

\mathbb{R}^{2n+1} as the limit of $l_\varepsilon(\gamma)$, the length computed in \mathbb{H}_n^ε when ε tends to 0. Precisely if for almost every t ,

$$\dot{\gamma}(t) = \sum_{k=1}^n (a_k(t)\mathbf{X}_k(\gamma(t)) + b_k(t)\mathbf{Y}_k(\gamma(t)) + c(t)\mathbf{U}(\gamma(t)),$$

then

$$l_c(\gamma) = \int \|\dot{\gamma}(t)\|_{\mathbb{H}} dt = \begin{cases} \int \sqrt{\sum_{k=1}^n (a_k^2 + b_k^2)(t)} dt & \text{if } c = 0 \text{ for almost every } t \text{ (*),} \\ +\infty & \text{otherwise.} \end{cases}$$

The curves satisfying condition (*) are called horizontal. Their speed vectors are contained in $T\mathbb{H}_n := \text{Vect}(\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n)$, that is the horizontal subspace. The distance d_c is finite because of the so-called Hörmander condition [23] (see also [35]): the horizontal subspace and the vector fields that it generates using the Lie bracket span the whole tangent space. Here we have simply $[\mathbf{X}_k, \mathbf{Y}_k] = \mathbf{U}$. Hence any two points of \mathbb{H}_n can be connected by a horizontal curve. Moreover there exists geodesics, *i.e.* curves minimizing the length between two points. As in Riemannian geometry one can even consider an exponential map $\exp^{\mathbb{H}}$ (see *e.g.* [9], or [5, 26, 15]). For $p \in \mathbb{H}_n$, $\vartheta \in \mathbb{R}$ and V a horizontal vector of $T_p\mathbb{H}_n$, the point $\exp_p^{\mathbb{H}}(V, \vartheta)$ is the end of a local geodesic, $\gamma : t \mapsto \exp_p^{\mathbb{H}}(tV, t\vartheta)$, of length $\|V\|_{\mathbb{H}}$ starting in p tangentially to V . The main difference with the exponential map of Riemannian geometry is that there are several such geodesics ; they are parametrized by ϑ . More precisely the projection on \mathbb{C}^n of the local geodesic describes an arc of circle of angle ϑ (up to a multiplicative constant depending on the authors). Note that $\exp^{\mathbb{H}}$ is smooth.

Another important consequence of the Hörmander condition is the hypoellipticity of the operators $\Delta_{\mathbb{H}} := \sum_{k=1}^n (\mathbf{X}_k^2 + \mathbf{Y}_k^2)$ and $\Delta_{\mathbb{H}} - \partial_t$, which in particular means that distributional solutions of the heat equation on $\mathbb{H}_n \times]0, +\infty[$,

$$\partial_t \rho_t = \Delta_{\mathbb{H}} \rho_t,$$

are smooth. In the precise case of the group \mathbb{H}_n , more is known about the solutions of this equation. For a regular initial condition μ_0 (a probability measure with finite second moment for instance), Gaveau [18] proved that the solutions are given by the convolution with a fundamental solution \mathfrak{h}_t (depending on the dimension n). An explicit expression of \mathfrak{h}_t is

$$\mathfrak{h}_t(z, u) = \frac{2}{(4\pi t)^{n+1}} \int_{\mathbb{R}} \exp\left(\frac{\lambda}{t}(\mathbf{i}u - \frac{|z|^2}{4} \coth \lambda)\right) \left(\frac{\lambda}{\sinh \lambda}\right)^n d\lambda.$$

In fact \mathfrak{h}_t is the density of $X_t^0 = (B_{\sqrt{2t}}^1, \dots, B_{\sqrt{2t}}^n, L_{\sqrt{2t}})$ where the $(B_t^k)_{t \geq 0}$ are n independent planar Brownian motions $B^k = B^{k,1} + \mathbf{i}B^{k,2}$ and L_t is the sum $L_t = L_t^1 + \dots + L_t^n$ of the Lévy areas $L_t^k = \frac{1}{2} \int_0^t (B_s^{k,1} dB_s^{k,2} - B_s^{k,2} dB_s^{k,1})$. Hence \mathfrak{h}_t is positive on $\mathbb{H}_n \times]0, +\infty[$. Note also that $\mathfrak{h}_t(z, u) = \frac{1}{t^{n+1}} \mathfrak{h}_1(z/\sqrt{t}, u/t)$. An expression for ρ_t is

$$(3) \quad \rho_t(p) = (\rho_0 * \mathfrak{h}_t)(p) = \int_{\mathbb{H}_n} \mathfrak{h}_t(q^{-1} \cdot p) d\mu_0(q),$$

which is the distribution of the product $X_t = X_0 \cdot X_t^0$ with independent terms, X_0 having probability law μ_0 . Again ρ_t is positive and $\int \rho_t(p) d\mathcal{L}^{2n+1}(p) = 1$ so that $(\rho_t \mathcal{L})_{t \geq 0}$ is a curve in the space of measures (as noticed after Proposition 1.4, it is even in $\mathcal{P}_2(\mathbb{H}_n)$ as soon as μ_0 is in this space).

1.1.1. *Some additional estimates.* We give here some general estimates and geometric relations on \mathbb{H}_n and \mathfrak{h}_t that we will need in this article.

Lemma 1.1. *Let $p = (z, u) \in \mathbb{H}_n = \mathbb{C}^n \times \mathbb{R}$ be a point of the Heisenberg group and $\varepsilon > 0$. Then*

$$d_\varepsilon \leq d_c \leq d_\varepsilon + 4\pi\varepsilon$$

and

$$\max(|z|, c(|z| + |u|^{1/2})) \leq d_c(p, 0) \leq C(|z| + |u|^{1/2})$$

for some positive constants c and C depending only on n .

Proof. Any horizontal curve has the same length in \mathbb{H}_n and \mathbb{H}_n^ε so that $d_\varepsilon \leq d_c$. For a curve γ from 0 to p with $\dot{\gamma}(t) = \sum_{k=1}^n (a_k(t)\mathbf{X}_k(\gamma(t)) + b_k(t)\mathbf{Y}_k(\gamma(t)))$, the projection on \mathbb{C}^n connects 0 to z and has speed vector $\sum_{k=1}^n a_k(t)\partial_{x_k} + b_k(t)\partial_{y_k}$. These two curves have the same length (the first one in \mathbb{H}_n and the second in \mathbb{C}^n). Thus $|z| \leq d_c(p, 0)$.

The proof of the equivalence between d_c and $N : (z, u) \mapsto (|z| + |u|^{1/2})$ is classical (see [16]) and follows the same lines as the equivalence of norms in a finite vector spaces. One compares those functions on a ‘‘sphere’’ (for instance $\{(z, u) \in \mathbb{H}_n \mid |z| + |u|^{1/2} = 1\}$). With the compactness one can define c and C as the minimum and maximum of d_c . The inequality can be extended to \mathbb{H}_n using dilations, a special feature of the Heisenberg group: for any $\lambda \geq 0$ the distance from 0 to $(\lambda z, \lambda^2 u)$ is $\lambda d_c(0, p)$ and N has the same homogeneity.

The last estimate is a little less usual and its proof requires the study of geodesics in \mathbb{H}_n^ε , for instance through the minimization of the energy of a curve (instead of its length). Details can be found in [5] or [25, subsection 1.6.4]. The analysis of the geodesics shows that there is a point $p' = (z, u')$ with the same z -coordinates as p and $|u - u'| \leq 2\pi\varepsilon^2$ such that $d_\varepsilon(0, p')^2 = d_{\mathbb{H}}(0, p)^2 + d_\varepsilon(p, p')^2$, which is a Pythagorean triple. It follows $|d_{\mathbb{H}}(0, p) - d_\varepsilon(0, p')| \leq d_\varepsilon(p, p')$. But the usual triangle inequality of \mathbb{H}_n^ε is $|d_\varepsilon(0, p) - d_\varepsilon(0, p')| \leq d_\varepsilon(p, p')$ so that $|d_{\mathbb{H}}(0, p) - d_\varepsilon(0, p)| \leq 2d_\varepsilon(p, p')$. But $d_\varepsilon(p, p') \leq 2\pi\varepsilon$ because there is a vertical path in direction \mathbf{U} of length smaller than $2\pi\varepsilon$ connecting $p = (z, u)$ to $p' = (z, u')$. \square

The Heisenberg group is not Abelian but we will need some commutativity relations in the rest of the paper.

Lemma 1.2. *The center of \mathbb{H}_n is $L = \{(z, u) \in \mathbb{H}_n \mid z = 0\}$. For $l \in L$ we have $d_c(p.l, p'.l) = d_c(p, p')$ and $d_\varepsilon(p.l, p'.l) = d_\varepsilon(p, p')$.*

*If $V \in \{\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n, \mathbf{U}\}$ and f and g are two regular functions (for instance of class $C_c^\infty(\mathbb{R}^{2n+1})$), then $V(f * g) = f * Vg$. If now g is a measure concentrated on L , then $V(f * g) = Vf * g$ also holds.*

Proof. the first part of this lemma simply relies on the definition of the multiplicative structure and the fact that d_c and d_ε are left-invariant. The definition of the convolution in \mathbb{H}_n appears in (3) and in this formula a differentiation under the integral sign justifies $V(f * g) = f * Vg$. The last point is a consequence of the two

previous one. Indeed if g is a measure concentrated on L one has $f * g = g * f$ and Vf also commutes with g . \square

The last estimates of this paragraph concerns \mathfrak{h}_t and its derivatives close to infinity. The best results presently available on this topic seem to be the estimates of H.-Q. Li [30]. Less recent results may be also sufficient for Proposition 1.3. For the third assumption see [29, p 376] by the same author.

Proposition 1.3. *Let k be an integer, $t > 0$ and $V_j \in \{\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n\}$ for every $j \in \{1, \dots, k\}$. One has*

- *the function $(V_1 \cdots V_k)\mathfrak{h}_t$ is in L^q for any $q \in [1, +\infty]$ and it is also true for $p \mapsto \max_{t \in [a, b]} (V_1 \cdots V_k)\mathfrak{h}_t(p)$ where $a, b > 0$,*
- *for every polynomial function $P(z, u)$, the product $P(z, u) \cdot (V_1 \cdots V_k)\mathfrak{h}_t(z, u)$ is an integrable function,*
- *the integral $\int \frac{(\mathbf{U}\mathfrak{h}_t)^2}{\mathfrak{h}_t} d\mathcal{L}$ is finite.*

1.2. Wasserstein spaces. Let (X, d) be a Polish metric space. Then the Wasserstein distance of two Borel probability measures μ_0 and μ_1 is defined as

$$(4) \quad W(\mu_0, \mu_1) = \inf_{\pi} \sqrt{\int_{X \times X} d^2(p, q) d\pi(p, q)}$$

where the infimum is taken over all couplings π of μ_0 and μ_1 , *i.e.* over all probability measures π over $X \times X$ such that $\bar{p}_{\#}\pi = \mu_0$ and $\bar{q}_{\#}\pi = \mu_1$. Here \bar{p} (resp. \bar{q}) stands for the first (resp. second) coordinate map and $\#$ is the push-forward symbol. The function W defines a distance on the so-called ‘‘Wasserstein space’’, that is,

$$\mathcal{P}_2(X) = \{\mu \in \mathcal{P}(X) \mid \int d(p, p_0)^2 d\mu(p) < +\infty\}$$

where p_0 is some point of X (the space does not depend on the choice of $p_0 \in X$).

For any Polish space X , there exist minimizing couplings π in (4). In the case of $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ and $\mathcal{P}_2(\mathbb{H}_n)$, it has been proved in [34] (with [13]) and [5] that if μ_0 is absolutely continuous, (4) is achieved by an unique π and this coupling can be expressed by $\pi = (\text{Id} \otimes T)_{\#}\mu_0$ for a μ_0 -almost everywhere uniquely defined map T . Moreover these authors prove that in these two geodesic spaces a geodesic between p and $T(p)$ is μ_0 -almost surely unique. We denote it by $(T_t(p))_{t \in [0, 1]}$ such that in particular $T_0 = \text{Id}$ and $T_1 = T$. Thence, [47, Corollary 7.23] ensure that there exists a unique geodesic $(\mu_t)_{t \in [0, 1]}$ between μ_0 and μ_1 , defined by $\mu_t = (T_t)_{\#}\mu_0$. Still under the assumption that μ_0 is absolutely continuous, the measures μ_t for $t < 1$ are absolutely continuous as well (see [14, 15]).

A related result [47, Corollary 10.44 and Theorem 13.5] states that on a manifold if μ_0 is absolutely continuous, for U_t in a form that we will precise, $\mu_t = (U_t)_{\#}\mu_0$ is also absolutely continuous for t small enough. It is actually sufficient that U_t writes $p \rightarrow \exp_p(tV(p))$ where \exp_p is the usual exponential map of Riemannian geometry and V is a smooth vector field with compact support. This proposition of course applies to \mathbb{H}_n^ε for any ε but there is also an adapted version for \mathbb{H}_n in [25, Section 2.2.3] with a similar proof. We state this result here:

Proposition 1.4. *Let ϑ be in $C_c^\infty(\mathbb{R}^{2n+1})$, V be a smooth horizontal vector field of compact support and μ_0 an absolutely continuous measure on \mathbb{H}_n . Define*

$$U_t(p) = \exp_p^{\mathbb{H}}(tV(p), t\vartheta(p)) \quad \text{and} \quad \mu_t = (U_t)_{\#}\mu_0.$$

There is a $t_0 > 0$ such that for any $t < t_0$, U_t is injective, smooth and the Jacobian determinant $\text{Jac}(U_t)$ is not zero. In particular μ_t is absolutely continuous.

With the notations of the Proposition 1.4, a disintegration of measure argument shows that if $\mu_0 \in \mathcal{P}_2(\mathbb{H}_n)$, the measure μ_t is also in this space, principally because the displacement by U_t is bounded. Similarly a consequence of Proposition 1.3 is that \mathfrak{h}_t is in $\mathcal{P}_2(\mathbb{H}_n)$ so that for any $\mu_0 \in \mathcal{P}_2(\mathbb{H}_n)$, the measure $\mu_t = \mu_0 * \mathfrak{h}_t$ is also in this space. We also notice with Lemma 1.1 that $\mathcal{P}_2(\mathbb{H}_n)$ and $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ are the same topological spaces whatever $\varepsilon > 0$ is.

Consider now Ent , the central functional of the present paper. It is not clear whether in definition (2), the function $\rho \ln(\rho)$ is integrable. On $\mathcal{P}_2(\mathbb{H}_n)$, at least the negative part is.

Lemma 1.5. *If $\mu \in \mathcal{P}_2(\mathbb{H}_n)$ has a density ρ with respect to the Haar (and also Lebesgue) measure \mathcal{L} , then $\int (\rho \ln \rho)^- d\mathcal{L} < +\infty$ so that Ent is well-defined on $\mathcal{P}_2(\mathbb{H}_n)$. It is a functional with values in $]-\infty, +\infty]$.*

Proof. It suffices to notice that this is true for $\rho \mathcal{L} \in \mathcal{P}_2(\mathbb{H}_n^\varepsilon)$. As it is a Lie group with an invariant metric, \mathbb{H}_n^ε is a Riemannian manifold with its Ricci curvature bounded from below (in fact with lower bound $-\frac{1}{2\varepsilon^2}$). In this case it is proved in [11, Lemma 4.1] that $\int [(\rho/a_\varepsilon) \ln(\rho/a_\varepsilon)]^- d(a_\varepsilon \mathcal{L}) < +\infty$ where $a_\varepsilon \mathcal{L}$ is the Riemannian volume of \mathbb{H}_n^ε . Thus

$$\int [\rho \ln(\rho)]^- d\mathcal{L} \leq \int [\rho \ln(\rho/a_\varepsilon)]^- d\mathcal{L} + \int [\rho \ln(a_\varepsilon)]^- d\mathcal{L} < +\infty$$

Notice that the proof could also be directly adapted from [11] because the volume of the balls of radius r of \mathbb{H}_n is well-known. Up to a multiplicative constant it is r^{2n+2} . □

1.3. Absolutely continuous curves. A curve $(\gamma_t)_{t \in I}$ defined on an interval I in a metric space (X, d) is said to be locally absolutely continuous on I if there exists a function m element of $L^1_{\text{loc}}(I)$ such that for any $a < b$ in I ,

$$d(\gamma(a), \gamma(b)) \leq \int_a^b m(t) dt < +\infty.$$

It is proved in [2, Theorem 1.1.2] that if γ is locally absolutely continuous, for almost every $t \in I$ the metric derivative

$$|\dot{\gamma}_t| := \lim_{|h| \rightarrow 0} \frac{d(\gamma_{t+h}, \gamma_t)}{|h|}$$

exists ($|\dot{\cdot}|$ is a notation for the metric speed) and

$$l(\gamma) = \int_a^b |\dot{\gamma}_t| dt$$

equals the classical metric length of the curve γ between a and b . In any metric space $l(\gamma) \geq d(\gamma(a), \gamma(b))$. If $l(\gamma) = d(\gamma(a), \gamma(b))$, γ is a geodesic, up to reparametrizations a constant-speed geodesic. We denote the space of locally absolutely continuous curves by $AC(X)$. Let $AC_2(X) \subset AC(X)$ be the subspace of locally absolutely continuous curves such that $t \mapsto |\dot{\gamma}_t|^2$ is locally integrable. Up to reparametrizations $AC(X)$ -curves are in $AC_2(X)$.

1.4. Gradient flows. In this paper, curves will take values in the Wasserstein spaces $\mathcal{P}_2(\mathbb{H}_n)$ and $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$. We will study the slope of the functional entropy Ent introduced in (2). The slope of Ent is defined as

$$\text{Slope}(\text{Ent})(\mu) = \begin{cases} +\infty & \text{if } \text{Ent}(\mu) = +\infty, \\ \max\left(0, \limsup_{\nu \rightarrow \mu} \frac{\text{Ent}(\mu) - \text{Ent}(\nu)}{W(\mu, \nu)}\right) & \text{otherwise.} \end{cases}$$

This quantity is positive and quantify how much the entropy can locally decrease. We keep the notations Slope , W and $|\cdot|$ for the slope, the Wasserstein distance and the metric speed of $\mathcal{P}_2(\mathbb{H}_n)$. In $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$, we use the notations Slope^ε , W^ε and $|\cdot|_\varepsilon$. As $d_c \geq d_\varepsilon$ (recall Lemma 1.1) we have for the Wasserstein distances $W \geq W^\varepsilon$ and $\text{Slope}^\varepsilon(\text{Ent})(\mu) \geq \text{Slope}(\text{Ent})(\mu)$ in every μ . We have also $|\dot{\mu}_t| \geq |\dot{\mu}_t|_\varepsilon$ if these metric speeds make sense.

Remark 1.6 (About infinite slope). Note that the definition $\text{Slope}(\text{Ent})(\mu) = +\infty$ if $\text{Ent}(\mu) = +\infty$ is not only a convenient convention, at least if the metric space is \mathbb{H}_n or \mathbb{H}_n^ε . In fact any measure μ is the limit of measures ν_n obtained by convolution with approximate identity (for instance with $(\mathfrak{h}_t)_{t>0}$). These measures are in the corresponding Wasserstein space so that $\text{Ent}(\nu_n)$ makes sense and is in $] -\infty, +\infty]$. It is not $+\infty$ because the density of ν_n is bounded from above.

We will use in this paper a metric definition for the gradient flow of the entropy. It can be compared to v) in Theorem 5.3 of [6] or to ii) of Proposition 23.1 of [47]. This definition has to be understood as the decomposition of “ $\dot{\mu}_t = -\nabla \text{Ent}(\mu_t)$ ” in (1) into the norm equality and the indication of the direction.

Definition 1.7. Let I be an interval. A curve in the Wasserstein space $(\mu_t)_{t \in I}$ is said to be a gradient flow of Ent if

- it is a locally absolutely continuous curve in $\mathcal{P}_2(X)$ and for almost every $t \in I$,

$$|\dot{\mu}_t| = \text{Slope}(\text{Ent})(\mu_t),$$

- the function $E(t) = \text{Ent}(\mu_t)$ is non-increasing and for almost every $t \in I$,

$$\dot{E}(t) = -\text{Slope}(\text{Ent})(\mu_t) \cdot |\dot{\mu}_t|.$$

Remark 1.8. With the notations of Definition 1.7, the function E is differentiable at almost every $t \in I$ because non-increasing functions are of bounded variation. Moreover $\dot{E}(t) = -|\dot{\mu}_t|^2$ for almost every t , so that $t \mapsto |\dot{\mu}_t|^2$ has a finite integral on $[t_0, t_1]$ provided that both $E(t_0)$ and $E(t_1)$ are finite. This condition holds if $X = \mathbb{H}_n$ and $t_0, t_1 > \inf(I)$ as is proved in Lemma 4.2 (or see directly Remark 1.9 and Lemma 1.5). Moreover it will be proved in Proposition 4.5 that E is locally absolutely continuous on $]0, T[$.

Remark 1.9 (About infinite entropy and infinite slope in gradient flows). The entropy can not be $+\infty$ on an interval greater than a single time. Indeed if $\text{Ent}(\mu_t) = +\infty$, for every $t \in [a, b]$ with $a < b$, we have also $|\dot{\mu}_t| = +\infty$. This contradicts the fact that μ_t is an AC-curve in $\mathcal{P}_2(X)$. Therefore if $(\mu_t)_{t \in [0, T[}$ is a gradient flow of the entropy, $\text{Ent}(\mu_t)$ is finite for $t > 0$. In particular μ_t admits a density function ρ_t for any $t > 0$.

1.5. Functional spaces, tangent spaces. The vector spaces we will introduce now are defined using weak formulations that require test functions. In the sequel, the space of test functions is $\mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$, the space of smooth functions of \mathbb{R}^{2n+1} with a compact support. For any $k \in \{1, \dots, n\}$ and f a function of $L_{\text{loc}}^1(\mathbb{H}_n)$ (a \mathcal{L} -locally integrable function on \mathbb{H}_n), if the distributional derivatives of f through $\mathbf{X}_k, \mathbf{Y}_k$ and \mathbf{U} are functions of $L_{\text{loc}}^1(\mathbb{H}_n)$ we will note them $\mathbf{X}_k f, \mathbf{Y}_k f$ and $\mathbf{U}f$. We state for instance the defining relation of $\mathbf{U}f$:

$$\forall \psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1}), \int (\mathbf{U}f)\psi d\mathcal{L} + \int f(\mathbf{U}\psi) d\mathcal{L} = 0$$

Note that these notations are coherent with the usual ones when f is smooth. Actually in this case, $f\psi$ has a compact support. Then (this time for \mathbf{X}_k because it is less obvious than for \mathbf{U})

$$\begin{aligned} \int \mathbf{X}_k(f\psi) d\mathcal{L} &= \int (\partial_{x_k} - \frac{1}{2}y_k\partial_u)(f\psi) d\mathcal{L} \\ &= \int \partial_{x_k}(f\psi) d\mathcal{L} - \frac{1}{2} \int \left(\int (y_k\partial_u)(f\psi) du \right) dx dy = 0 \end{aligned}$$

But the first integral is also $\int (\mathbf{X}_k f)\psi d\mathcal{L} + \int f(\mathbf{X}_k\psi) d\mathcal{L}$ as we want.

We say that f has a gradient $\nabla_{\mathbb{H}}f$ if it has a weak derivative Vf for any $V \in \{\mathbf{X}_1, \mathbf{Y}_1, \dots, \mathbf{X}_n, \mathbf{Y}_n\}$. Then we can define the \mathbb{H} -gradient $\nabla_{\mathbb{H}}f$ by

$$\nabla_{\mathbb{H}}f := \sum_{k=1}^n (\mathbf{X}_k f \mathbf{X}_k + \mathbf{Y}_k f \mathbf{Y}_k).$$

If moreover f has a weak \mathbf{U} -derivative, we can define the ε -gradient $\nabla_\varepsilon f$,

$$\nabla_\varepsilon f := \nabla_{\mathbb{H}}f + (\varepsilon\mathbf{U})f(\varepsilon\mathbf{U}) = \nabla_{\mathbb{H}}f + (\varepsilon^2\mathbf{U}f)\mathbf{U}.$$

We note $W_{\text{loc}}^{1,1}(\mathbb{H}_n)$ the space of locally integrable functions f with a locally integrable \mathbb{H} -gradient $\nabla_{\mathbb{H}}f$. Let $W_{\text{loc}}^{1,1}(\mathbb{H}_n^\varepsilon) \subset W_{\text{loc}}^{1,1}(\mathbb{H}_n)$ be the subspace of the functions f with a weak derivative $\mathbf{U}f$.

We note div the simple divergence operator of \mathbb{R}^{2n+1} . One can check that for $V = \sum_{k=1}^n a_k \mathbf{X}_k + \sum_{k=1}^n b_k \mathbf{Y}_k + c\mathbf{U}$, $\text{div} V$ can be written $\sum \mathbf{X}_k a_k + \sum \mathbf{Y}_k b_k + \mathbf{U}c$. Hence we have the following integration by part where $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$ and V is smooth,

$$\int \langle V | \nabla_\varepsilon \psi \rangle_\varepsilon d\mathcal{L} = - \int \psi \text{div}(V) d\mathcal{L},$$

for any $\varepsilon > 0$, and if V is horizontal

$$\int \langle V | \nabla_{\mathbb{H}} \psi \rangle_{\mathbb{H}} d\mathcal{L} = - \int \psi \text{div}(V) d\mathcal{L}.$$

Note that if $V = \nabla_{\mathbb{H}}\varphi$, the function $\text{div} V$ is $\Delta_{\mathbb{H}}\varphi$. Later div will also be used in the distributional sense, especially in the continuity equations (5) and (8).

For a given absolutely continuous measure μ , we define $\text{Tan}_{\mathbb{H}}(\mu)$, the tangent space at $\mu \in \mathcal{P}_2(\mathbb{H}_n)$ as the smallest Hilbert space that contains the vector fields $\nabla_{\mathbb{H}}\psi$ for a test function ψ . Precisely

$$\text{Tan}_{\mathbb{H}}(\mu) = \overline{\{\nabla_{\mathbb{H}}\psi \mid \psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})\}}^{L^2(\mu)}$$

with $L_{\mathbb{H}}^2(\mu) = \{\xi \in \Gamma(T\mathbb{H}_n) \mid \int_{\mathbb{H}_n} \|\xi\|_{\mathbb{H}}^2 d\mu < +\infty\}$. Similarly at $\mu \in \mathcal{P}_2(\mathbb{H}_n^\varepsilon)$, the tangent space $\text{Tan}_\varepsilon(\mu)$ is defined by

$$\text{Tan}_\varepsilon(\mu) = \overline{\{\nabla_\varepsilon \psi \mid \psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})\}}^{L_\varepsilon^2(\mu)}$$

where $L_\varepsilon^2(\mu) = \{\xi \in \Gamma(T\mathbb{H}_n^\varepsilon) \mid \int_{\mathbb{H}_n^\varepsilon} \|\xi\|_\varepsilon^2 d\mu < +\infty\}$. See [2, Chapter 12] or [47, Chapter 13 and 15] and the references therein for the identification of these ‘‘tangent spaces’’. The good reasons for this appellation became first apparent with the geometric formalism introduced by Otto [38].

2. SOME RESULTS ON THE ‘‘RIEMANNIAN’’ WASSERSTEIN SPACES $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$

We state here, for the approximating manifolds \mathbb{H}_n^ε a proposition that Erbar [11, Proposition 4.3.] proved for Riemannian manifolds. It is the translation of the same statement for Euclidean spaces by Ambrosio and Savaré [6, Theorem 4.16]. The proof of Erbar also strongly relies on Theorem 23.14 of [47] whose proof is long and difficult. Note that these results can be applied to \mathbb{H}_n^ε because it has Ricci curvature bounded from below by $-\frac{1}{2\varepsilon^2}$ (see for instance [9]).

Proposition 2.1 (On the slope of Ent). *Let $\varepsilon > 0$ and $\mu \in \mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ that has a density ρ . Then the following statements are equivalent:*

- (i) $\text{Slope}^\varepsilon(\text{Ent})(\mu) < +\infty$,
- (ii) $\rho \in W_{\text{loc}}^{1,1}(\mathbb{H}_n^\varepsilon)$, and $\nabla_\varepsilon \rho = \rho w^\varepsilon$ \mathcal{L} -almost everywhere for some $w^\varepsilon \in L_\varepsilon^2(\mu)$.

In this case $w^\varepsilon \in \text{Tan}_\varepsilon(\mu)$ and $\text{Slope}^\varepsilon(\text{Ent})(\mu) = \|w^\varepsilon\|_{L_\varepsilon^2(\mu)}$.

Remark 2.2. The vector field w^ε can be simply written ‘‘ $\nabla_\varepsilon \rho / \rho$ ’’. Actually as $\mu = \rho \mathcal{L}$, the function ρ is μ -almost surely non-zero such that $\nabla_\varepsilon \rho / \rho$ makes sense in $L_\varepsilon^2(\mu)$. The interpretation of this vector field in the Otto calculus is ‘‘ ∇Ent ’’. Note moreover that if condition (ii) is satisfied $\{\rho = 0\} \cap \{\nabla_\varepsilon \rho \neq 0\}$ has Lebesgue measure zero. We will use both the notations w^ε and $\nabla_\varepsilon \rho / \rho$ for the same vector field. Nevertheless when ρ is positive as in paragraph 4.1 about the heat curve, the second notation will be used more frequently.

Remark 2.3. The quantity $\|w^\varepsilon\|_{L_\varepsilon^2(\mu)}^2 = \int \frac{\|\nabla_\varepsilon \rho\|_\varepsilon^2}{\rho} d\mathcal{L}$ is traditionally called the Fisher information of $\mu = \rho \mathcal{L}$. In the Riemannian manifold \mathbb{H}_n^ε , one will represent it by $I_\varepsilon(\mu)$. Thus according to Proposition 2.1, if it exists $\sqrt{I_\varepsilon(\mu)} = \text{Slope}^\varepsilon(\text{Ent})(\mu)$.

Remark 2.4. If statement (ii) is true for some ε , it also holds for other $\varepsilon' > 0$. It follows that if the slope is finite in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$, it is also finite in the other Wasserstein spaces $\mathcal{P}_2(\mathbb{H}_n^{\varepsilon'})$ even for $\varepsilon' > \varepsilon$.

We state now a mix of propositions by Villani [47, Theorem 23.14 and Particular Case 23.15] and by Erbar [11, Proposition 2.5].

Proposition 2.5. *Let $\varepsilon > 0$ and I be an open interval. Let $(\mu_t)_{t \in I}$ be an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$. Assume that for almost every $t \in I$, the slope $\text{Slope}^\varepsilon(\text{Ent})(\mu_t)$ is finite. Let $w_t^\varepsilon \in L_\varepsilon^2(\mu_t)$ be the corresponding vector field given by Proposition 2.1. Then there is a subset $I' \subset I$ of full-measure such that for any $t \in I'$ the speed $|\dot{\mu}_t|_\varepsilon$ is finite and there is a vector field $v_t^\varepsilon \in \text{Tan}_\varepsilon(\mu_t)$ with $\|v_t^\varepsilon\|_\varepsilon = |\dot{\mu}_t|_\varepsilon$ that satisfies the two statements (5) and (6) below. The vector field v_t^ε can be chosen to be Borel measurable as a function $(t, p) \mapsto v_t^\varepsilon(p)$.*

- One has $\frac{d}{dt}\mu_t + \operatorname{div}(v_t^\varepsilon \mu_t) = 0$, i.e. for every test function $\psi \in \mathcal{C}_c^\infty$,

$$(5) \quad \frac{d}{dt} \int \psi d\mu_t = \int \langle v_t^\varepsilon \mid \nabla_\varepsilon \psi \rangle_\varepsilon d\mu_t,$$

- along the curve the entropy satisfies

$$(6) \quad \operatorname{Ent}(\mu_s) \geq \operatorname{Ent}(\mu_t) + \int \langle w_t^\varepsilon \mid (s-t)v_t^\varepsilon \rangle_\varepsilon d\mu_s + o(|s-t|),$$

when s goes to t .

Moreover as a complement to (6), for any μ and ν in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ such that $\operatorname{Slope}^\varepsilon(\operatorname{Ent})(\mu) < +\infty$,

$$(7) \quad \operatorname{Ent}(\nu) \geq \operatorname{Ent}(\mu) - \|w^\varepsilon\|_{L_\varepsilon^2(\mu)} W^\varepsilon(\mu, \nu) - \frac{1}{2 \cdot 2\varepsilon^2} W^\varepsilon(\mu, \nu)^2.$$

Proof. The space $(\mathbb{H}_n^\varepsilon, d_\varepsilon)$ is a complete Riemannian manifold with Ricci curvature uniformly bounded from below by $K = -1/2\varepsilon^2$. We can apply Proposition 2.5 of [11] and we obtain a measurable vector field v_t^ε of $I \times \mathbb{H}_n^\varepsilon$ satisfying $v_t^\varepsilon \in \operatorname{Tan}_\varepsilon(\mu_t)$ for almost every time and the space-time version of (5) for test functions ψ of compact support in $I \times \mathbb{H}_n^\varepsilon$. In order to localize the continuity equation for times in a set of full measure I' we first consider “rectangular” test functions of type $(t, p) \mapsto g(t)\psi(p)$ where $g(t)$ approximates the indicator of a segment. We deduce that $t \mapsto \int \psi d\mu(t)$ has a derivatives on a set of full measure. In order to cancel the dependance of this set with respect to ψ we take a countable family of test functions, following the same method as at the end of [47, proof of Theorem 13.8] (see also the end of Proposition 3.1). Note that we obtain (5) with a slightly different normalization but it does not matter (see Remark 2.6). Another way to obtain directly (5) would be to follow the proof of Theorem 13.8 replacing the Lipschitz curve by an absolutely continuous curve as suggested in [47, Remark 13.9]. However it would be not clear how to conclude that $(t, p) \mapsto v_t^\varepsilon(p)$ is Borel measurable up to a null set.

The equation (6) is obtained as an easy consequence of Particular Case 23.15 in [47] by using Lemma 2.7 of [11]. This lemma permits to approximate the optimal transport vector field $v^\varepsilon(\mu_t, \mu_s)$ that relates optimally μ_t to μ_s . Precisely for almost every $t \in I$ we have the Taylor expansion

$$v^\varepsilon(\mu_t, \mu_s) = (s-t)v_t^\varepsilon + o(|s-t|),$$

in $L_\varepsilon^2(\mu_t)$.

In the Riemannian manifold \mathbb{H}_n^ε we can apply Particular Case 23.15 and Remark 23.16 of [47]. For almost every t , according to the hypothesis and Proposition 2.1, we have $\mu_t \in W_{\operatorname{loc}}^{1,1}$ as we need for this particular case. Thus we obtain (7) in a slightly different normalization (see Remark 2.6). \square

Remark 2.6. The different quantities ρ_t , Ent , v_t^ε , w_t^ε in Proposition 2.5 may depend on the choice of the reference measure, \mathcal{L} or $\operatorname{vol}_\varepsilon$. We have $\rho_t = c_\varepsilon^{-1} \rho_t^\varepsilon$ where c_ε is the constant defined by $\mathcal{L} = c_\varepsilon \operatorname{vol}_\varepsilon$. The vector field $w_t^\varepsilon = \nabla \rho_t^\varepsilon / \rho_t^\varepsilon$ is simply w_t and the entropy with respect to $\operatorname{vol}_\varepsilon$ is $\operatorname{Ent}(\mu_t) + \ln(c_\varepsilon)$. Like w_t^ε , the vector field v_t^ε is not modified. Finally the inequalities of the proposition are equivalent with \mathcal{L} or $\operatorname{vol}_\varepsilon$ as reference measure. The form with $\operatorname{vol}_\varepsilon$ in the Riemannian manifold $(\mathbb{H}_n^\varepsilon, d_\varepsilon, \operatorname{vol}_\varepsilon)$ are actually the original statements of (5), (6) and (7) that can be found in [11] and [47].

Remark 2.7. The type of inequality like (7) is called a HWI inequality. It was established the first time by Otto and Villani [39]. On a Riemannian manifold it can be written $H(\nu) \geq H(\mu) - W^\varepsilon(\mu, \nu) \sqrt{I_\varepsilon(\mu)} + \frac{1}{2}K \cdot W^\varepsilon(\mu, \nu)^2$, where H stands for Ent (it is the ‘‘H-functional of Boltzmann’’) and I_ε is the Fisher information (see Remark 2.3). The real K is an uniform lower bound on the Ricci curvature. In \mathbb{H}_n^ε the greatest possible bound is $\frac{-1}{2\varepsilon^2}$. Note that it tends to $-\infty$ as ε goes to 0.

We state a result from [11, Proposition 3.6] (see also [2, Definition 1.2.1. and Corollary 2.4.10.]) for the special Riemannian manifolds \mathbb{H}_n^ε . Contrarily to the estimates of Proposition 2.5 it is an integrated and not a pointwise estimate. For this statement one usually says that $\text{Slope}^\varepsilon(\text{Ent})$ is an upper gradient of Ent in $\{\mu \in \mathcal{P}_2(\mathbb{H}_n^\varepsilon), \text{Ent}(\mu) < +\infty\}$.

Proposition 2.8 (Upper gradient estimate). *Let I be an open interval and $(\mu_t)_{t \in I}$ an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$. Assume moreover that $\text{Ent}(\mu_t) < +\infty$ for every t and for any $t_0, t_1 \in I$ the integral $\int_{t_0}^{t_1} [\text{Slope}^\varepsilon(\text{Ent})(\mu_t)] \cdot |\dot{\mu}_t|_\varepsilon dt$ is finite. Then $t \mapsto \text{Ent}(\mu_t)$ is locally absolutely continuous with*

$$|\text{Ent}(\mu_{t_1}) - \text{Ent}(\mu_{t_0})| \leq \int_{t_0}^{t_1} [\text{Slope}^\varepsilon(\text{Ent})(\mu_t)] \cdot |\dot{\mu}_t|_\varepsilon dt$$

for any $t_0, t_1 \in I$.

In Proposition 3.1 we will prove a similar statement to (5) for the ‘‘true’’ Heisenberg group \mathbb{H}_n . In subsection 3.2 we will let ε go to zero in (7) and get results on the slope of the entropy in $\mathcal{P}_2(\mathbb{H}_n)$, the Wasserstein space of the ‘‘true’’ Heisenberg group. Inequality (6) will also be interpreted in the context of $\mathcal{P}_2(\mathbb{H}_n)$ at the end of paragraph 4.2. With Proposition 2.8, for ε going to 0 we will obtain successively Proposition 4.5 and Proposition 4.6, two major results of the proof-paragraph 4.2.

3. SPEED OF CURVES AND SLOPE OF Ent IN $\mathcal{P}_2(\mathbb{H}_n)$

3.1. Speed and velocity. Equality (5) in Proposition 2.5 shows that for the AC_2 -curves $(\mu_t)_{t \in I}$ in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ (where a metric speed $|\dot{\mu}_t|_\varepsilon$ is available) it is possible to define a velocity (speed and direction) thanks to a vector field v_t^ε . We show that there is a similar velocity for $\mathcal{P}_2(\mathbb{H}_n)$. The next proposition will appear in the paragraph 4.2 on the proof of Theorem 0.1 where we will apply it to the curves driven by the gradient flow of Ent.

Proposition 3.1. *Let I be an open interval and $(\mu_t)_{t \in I}$ an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n)$. Then there is a subset $I' \subset I$ of full measure such that in any $t \in I'$ there is a vector field $v_t \in L^2_{\mathbb{H}}(\mu_t)$ so that*

$$(8) \quad \frac{d}{dt} \mu_t + \text{div}(v_t \mu_t) = 0$$

in the weak sense. It means that for every function $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$,

$$\frac{d}{dt} \int \psi d\mu_t = \int \langle \nabla_{\mathbb{H}} \psi | v_t \rangle_{\mathbb{H}} d\mu_t.$$

The vector field v_t can be chosen to be Borel measurable as a function $(t, p) \mapsto v_t(p)$. Moreover, $\|v_t\|_{L^2_{\mathbb{H}}(\mu_t)} \leq |\dot{\mu}_t|$ for any $t \in I'$.

Conversely if $(v_t)_{t \in I}$ is a Borel vector field satisfying the continuity equation and $\int_{t_0}^{t_1} \|v_t\|_{L^2_{\mathbb{H}}(\mu_t)}^2 dt < +\infty$ for every $t_0 < t_1$, one has $|\dot{\mu}_t| \leq \|v_t\|_{L^2_{\mathbb{H}}(\mu_t)}$ for almost every

$t \in I$. Hence for the vector field introduced in the first part, I' can be restricted so that $\|v_t\|_{L^2_{\mathbb{H}}(\mu_t)} = |\dot{\mu}_t|$ for any $t \in I'$.

Proof. As $(\mu_t)_{t \in I}$ is an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n)$, it is also in $AC_2(\mathcal{P}_2(\mathbb{H}_n^\varepsilon))$ for any $\varepsilon > 0$. Note that for $\psi \in C_c^\infty(\mathbb{R}^{2n+1})$, the map $\mu \in \mathcal{P}_2(\mathbb{H}_n^\varepsilon) \mapsto \int \psi d\mu$ is Lipschitz continuous with Lipschitz constant equal to the Lipschitz constant of ψ in \mathbb{H}_n^ε . Thus $t \mapsto \int \psi d\mu_t$ is locally absolutely continuous and according to Proposition 2.5 there are Borel vector fields $v_t^\varepsilon \in L^2_{\mathbb{H}}(\mu_t)$ such that for any $t_0 < t_1$,

$$(9) \quad \int \psi d\mu_{t_1} - \int \psi d\mu_{t_0} = \int_{t_0}^{t_1} \int \langle \nabla_\varepsilon \psi \mid v_t^\varepsilon \rangle_\varepsilon d\mu_t dt$$

and

$$\int \|v_t^\varepsilon\|_\varepsilon^2 d\mu_t \leq |\dot{\mu}_t|_\varepsilon^2 \leq |\dot{\mu}_t|^2$$

for almost every $t \in [t_0, t_1]$ with moreover $|\dot{\mu}_t| \in L^2_{\text{loc}}(I)$. This last estimates writes

$$(10) \quad (\|v_t^{\varepsilon, \mathbb{H}}\|_{\mathbb{H}})^2 + \frac{1}{\varepsilon^2} \int (v_t^{\varepsilon, \mathbf{U}})^2 d\mu_t \leq |\dot{\mu}_t|^2$$

where $v_t^{\varepsilon, \mathbb{H}} + v_t^{\varepsilon, \mathbf{U}} \mathbf{U} = v_t^{\varepsilon, \mathbb{H}} + \frac{1}{\varepsilon} v_t^{\varepsilon, \mathbf{U}} \varepsilon \mathbf{U}$ is the decomposition of v_t^ε into the horizontal frame and the \mathbf{U} vector. By weak compactness of balls in Hilbert spaces and a Cantor diagonalization argument, we can find a sequence v^{ε_i} with $\varepsilon_i \downarrow 0$ and a Borel vector field v , such that for any compact interval $K = [t_0, t_1] \subset I$, the vector field v^{ε_i} converges to v in $L^2_1(\bar{\mu}_K)$. Here $\bar{\mu}_K$ is the measure on $\mathbb{H}_n \times K$ with time marginals the measures μ_t . The norm of a vector w in the Hilbert space $L^2_1(\bar{\mu}_K)$ is

$$\|w\| = \sqrt{\int_{t_0}^{t_1} \int \|w_t\|_1^2 d\mu_t dt}$$

where 1 stands for $\varepsilon = 1$. From (10) we learn that $v_t^{\varepsilon_i, \mathbf{U}}$ tends to 0 as i goes to infinity. It follows that v is horizontal (for a.e. t , v_t is horizontal) because the norm of its \mathbf{U} coordinate has to be smaller than the lower limit of the corresponding norms in the sequence. But for any $i \in \mathbb{N}$ and almost every $t \in I$,

$$\begin{aligned} \int \langle \nabla_{\varepsilon_i} \psi \mid v_t^{\varepsilon_i} \rangle_{\varepsilon_i} d\mu_t &= \int \langle \nabla_{\mathbb{H}} \psi \mid v_t^{\varepsilon_i, \mathbb{H}} \rangle_{\mathbb{H}} d\mu_t + \frac{1}{\varepsilon_i^2} \int v_t^{\varepsilon_i, \mathbf{U}} \varepsilon_i^2 \mathbf{U} \psi d\mu_t \\ &= \int \langle \nabla_{\mathbb{H}} \psi \mid v_t^{\varepsilon_i, \mathbb{H}} \rangle_1 d\mu_t + \int v_t^{\varepsilon_i, \mathbf{U}} \mathbf{U} \psi d\mu_t. \end{aligned}$$

Because the horizontal part of v^{ε_i} weakly converges to the horizontal part of v (that is v itself) and $v_t^{\varepsilon_i, \mathbf{U}}$ to 0, it follows

$$\lim_{i \rightarrow +\infty} \int \langle \nabla_{\varepsilon_i} \psi \mid v_t^{\varepsilon_i} \rangle_{\varepsilon_i} d\mu_t = \int \langle \nabla_{\mathbb{H}} \psi \mid v_t \rangle_1 d\mu_t = \int \langle \nabla_{\mathbb{H}} \psi \mid v_t \rangle_{\mathbb{H}} d\mu_t$$

This occurs for almost every $t \in [t_0, t_1]$. In (9), we put the limit under the integral sign and can use the Lebesgue dominated convergence because $\int \langle \nabla_\varepsilon \psi \mid v_t^\varepsilon \rangle_\varepsilon d\mu_t \leq \|\nabla_\varepsilon \psi\|_\varepsilon \cdot |\dot{\mu}_t|$. But $(\mu_t)_{t \in I}$ is an AC -curve in $\mathcal{P}_2(\mathbb{H}_n)$ and $\|\nabla_\varepsilon \psi\|_\varepsilon^2 = \|\nabla_{\mathbb{H}} \psi\|_{\mathbb{H}}^2 + \varepsilon^2 (\mathbf{U} \psi)^2$ is smaller than $\|\nabla_1 \psi\|_1^2$ as soon as $\varepsilon \leq 1$.

It follows that $\zeta_\psi : t \mapsto \int \psi d\mu_t$ is almost everywhere differentiable and this derivative is $\int \langle \nabla_{\mathbb{H}} \psi \mid v_t \rangle_{\mathbb{H}} d\mu_t$. The estimate $\|v_t\|_{L^2_{\mathbb{H}}(\mu_t)} \leq |\dot{\mu}_t|$ has been proved during the proof as a limit of (10). Observe that the proof is not totally complete. Actually we do not have a correct set I' as in the statement because $C_c^\infty(\mathbb{R}^{2n+1})$

is uncountable. A way to get over this problem is to consider a family $(\psi_k)_{k \in \mathbb{N}}$ with $(\nabla_{\mathbb{H}} \psi_k)_{k \in \mathbb{N}}$ dense in the set $\{\nabla_{\mathbb{H}} \psi \mid \psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})\}$ for the uniform norm. Hence it is dense in $\text{Tan}_{\mathbb{H}}(\mu_t)$ for any t . Let I' be the set of points t such that all functions $\zeta_k : t \mapsto \int \psi_k d\mu_t$ are differentiable with derivative $\int \langle \nabla_{\mathbb{H}} \psi_k | v_t \rangle_{\mathbb{H}} d\mu_t$ at t . But $\zeta_k(t)$ is $\zeta_k(t) = \zeta_k(t_0) + \int_{t_0}^t (\int \langle \nabla_{\mathbb{H}} \psi_k | v_s \rangle_{\mathbb{H}} d\mu_s) ds$. Let ψ be in $\mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$ and assume without loss of generality that $(\nabla_{\mathbb{H}} \psi_k)_{k \in \mathbb{N}}$ converges to $\nabla_{\mathbb{H}} \psi$. Hence $\int \langle \nabla_{\mathbb{H}} \psi_k | v_s \rangle_{\mathbb{H}} d\mu_s$ converges pointwise to $\int \langle \nabla_{\mathbb{H}} \psi | v_s \rangle_{\mathbb{H}} d\mu_s$ and moreover

$$\left| \int \langle \nabla_{\mathbb{H}} \psi | v_s \rangle_{\mathbb{H}} d\mu_s - \int \langle \nabla_{\mathbb{H}} \psi_k | v_s \rangle_{\mathbb{H}} d\mu_s \right| \leq \|\nabla_{\mathbb{H}}(\psi - \psi_k)\|_\infty \cdot |\dot{\mu}_s|.$$

This difference is locally integrable because $(\mu_t)_{t \in I}$ is an AC -curve in $\mathcal{P}_2(\mathbb{H}_n)$. Hence the derivative of ζ_ψ exists at any $t \in I'$ and it is the limit of $\zeta'_k(t)$, *i.e.* $\int \langle \nabla_{\mathbb{H}} \psi | v_t \rangle_{\mathbb{H}} d\mu_t$.

For the converse part, let $(v_t)_{t \in I}$ be a Borel vector field satisfying the continuity equation and $\int_{t_0}^{t_1} \|v_t\|_{L_{\mathbb{H}}^2(\mu_t)}^2 dt < +\infty$ for every $t_0 < t_1$. This also means $\int_{t_0}^{t_1} \|v_t\|_{L_{\mathbb{H}}^1(\mu_t)}^2 dt < +\infty$ because v_t is horizontal (here $\varepsilon = 1$). Hence one can apply Theorem 5.8 in [7] in the setting of the continuity equation on Riemannian manifolds. We obtain a probability measure Π on $\mathcal{C}(I, \mathbb{H}_n^1)$, the space of curves over \mathbb{H}_n^1 with the Borel sigma-field, satisfying the conditions:

- Π is concentrated on $AC_2(\mathbb{H}_n^1)$.
- the curve γ is $\Pi(\gamma)$ -almost certainly an integral curve of the vector field v_t , so that it is a horizontal curve.
- for every $t \in I$, the law of the point $\gamma(t)$ with respect to Π is μ_t .

The following computation will conclude the proof of the converse part.

$$\begin{aligned} W(\mu_s, \mu_t)^2 &\leq \int_{\mathcal{C}(I, \mathbb{H}_n)} d_c(\gamma_s, \gamma_t)^2 d\Pi(\gamma) \leq \int_{\mathcal{C}(I, \mathbb{H}_n)} \left(\int_s^t \|\dot{\gamma}(\tau)\|_{\mathbb{H}} d\tau \right)^2 d\Pi(\gamma) \\ &\leq (t-s) \int \int_s^t \|v_t(\gamma(\tau))\|_{\mathbb{H}}^2 d\tau d\Pi(\gamma) \leq (t-s) \int_s^t \int \|v_t\|_{\mathbb{H}}^2 d\mu_\tau d\tau. \end{aligned}$$

Thus $\|v_t\|_{L_{\mathbb{H}}^2(\mu_t)} = |\dot{\mu}_t|$ holds for almost every $t \in I$. \square

3.2. Slope. After the previous part we can represent the velocity of $(\mu_t)_{t \in I}$ by a vector field. Propositions 3.2 and 3.4 makes the picture more precise and allow to identify “the gradient of the entropy” as a vector field, at least at some points $\mu \in \mathcal{P}_2(\mathbb{H}_n)$.

Proposition 3.2 (On the slope of Ent, 2). *Let $\varepsilon > 0$ and $\mu \in \mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ that has a density ρ . Assume moreover $\text{Slope}(\text{Ent})(\mu) < +\infty$. Then there is a horizontal vector field $w^{\mathbb{H}} \in L_{\mathbb{H}}^2(\mu)$ such that \mathcal{L} -almost everywhere $\nabla_{\mathbb{H}} \rho = \rho w^{\mathbb{H}}$.*

Moreover,

$$\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu)} = \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho\|_{\mathbb{H}}^2}{\rho} d\mathcal{L}} \leq \text{Slope}(\text{Ent})(\mu).$$

Proof. Let V be a smooth horizontal vector field with compact support and $\vartheta \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$. We consider then $U_t(p) = \exp_p^{\mathbb{H}}(tV(p), t\vartheta(p))$. Because of Proposition 1.4 for t small enough the map U_t is smooth, one-to-one and $\text{Jac}(U_t)$ does not vanish. Note that $V(p)$ is the speed vector of $t \rightarrow U_t(p)$ at time $t = 0$ and that

this curve is a geodesic of (constant) speed $\|V(p)\|_{\mathbb{H}}$. Then the metric speed of $((U_t)_\# \mu)_{t>0}$ is smaller than $\|V\|_{L^2_{\mathbb{H}}(\mu)}$. The derivative in time of $\text{Ent}(\mu_t)$ can be classically computed using the fact that $(p, t) \rightarrow U_t(p)$ is the flow of a smooth vector field corresponding to V in $t = 0$. Hence we have

$$\frac{d\text{Ent}(\mu_t)}{dt} \Big|_{t=0} = - \int \rho \operatorname{div}(V) d\mathcal{L}.$$

Note that if the entropy grows, which we want to avoid, we can replace V by $-V$. Therefore the following inequality with slope and speed on the right hand-side holds:

$$\left| \int \rho \operatorname{div}(V) d\mathcal{L} \right| \leq \operatorname{Slope}(\text{Ent})(\mu) \cdot \|V\|_{L^2_{\mathbb{H}}(\mu)}.$$

But $\operatorname{Slope}(\text{Ent})(\mu)$ is finite too. Hence the Riesz representation theorem provides a horizontal vector field $w^{\mathbb{H}} \in L^2_{\mathbb{H}}(\mu)$ with $\|w^{\mathbb{H}}\|_{L^2_{\mathbb{H}}(\mu)} \leq \operatorname{Slope}(\text{Ent})(\mu)$ such that for any $V \in \Gamma(T\mathbb{H}_n)$ with compact support,

$$- \int \rho \operatorname{div}(V) d\mathcal{L} = \int \langle w^{\mathbb{H}} | V \rangle_{\mathbb{H}} d\mu.$$

Thence one observes that $\rho \in W_{\text{loc}}^{1,1}(\mathbb{H})$ with $\nabla_{\mathbb{H}} \rho = \rho w^{\mathbb{H}}$. \square

Remark 3.3. As in Remark 2.2, $w^{\mathbb{H}}$ can be written “ $\nabla_{\mathbb{H}} \rho / \rho$ ” and as in Remark 2.3 one will say that $I_{\mathbb{H}}(\mu) = \int \frac{\|\nabla_{\mathbb{H}} \rho\|_{\mathbb{H}}^2}{\rho} d\mathcal{L}$ is the Fisher information of μ in \mathbb{H}_n . An important issue of this paper is to find out if $\operatorname{Slope}(\text{Ent})(\mu)$ equals $\sqrt{I_{\mathbb{H}}(\mu)}$, which is a sort of “maximal slope of Ent along regular curves starting from μ ”.

Contrarily to Proposition 2.1, we were not able to state that the finiteness of $\operatorname{Slope}(\text{Ent})$ is equivalent to the existence of a $w^{\mathbb{H}} \in L^2_{\mathbb{H}}(\mu)$ and that these conditions imply $w^{\mathbb{H}} \in \operatorname{Tan}_{\mathbb{H}}(\mu)$. Nevertheless this can be established under a further condition.

Proposition 3.4 (On the slope of Ent, 3). *Let $\mu \in \mathcal{P}_2(\mathbb{H}_n)$ with a density ρ and such that $\operatorname{Slope}^\varepsilon(\text{Ent})(\mu) < +\infty$ for some (or any) $\varepsilon > 0$. Then $\operatorname{Slope}(\text{Ent})$ is also finite in μ and*

$$\|w^{\mathbb{H}}\|_{L^2(\mathbb{H})} = \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L}} = \operatorname{Slope}(\text{Ent})(\mu).$$

Moreover $w_{\mathbb{H}} \in \operatorname{Tan}_{\mathbb{H}}(\mu)$.

Proof. Let $\mu \in \mathcal{P}_2(\mathbb{H}_n)$ with finite entropy and such that $\operatorname{Slope}^\varepsilon(\text{Ent})(\mu) < +\infty$ for some (or any, see Remark 2.4) $\varepsilon > 0$. As $W^\varepsilon \leq W$, we have also $\operatorname{Slope}(\text{Ent})(\mu) \leq \operatorname{Slope}^\varepsilon(\text{Ent})(\mu) < +\infty$. From Proposition 3.2 we know that ρ has a weak derivative $\nabla_{\mathbb{H}}$ and that there is $w^{\mathbb{H}} \in L^2_{\mathbb{H}}(\mu)$ such that $\rho w^{\mathbb{H}} = \nabla_{\mathbb{H}} \rho$ and $\|w^{\mathbb{H}}\|_{L^2_{\mathbb{H}}(\mu)} \leq \operatorname{Slope}(\text{Ent})(\mu)$. The fact that the slope of Ent is also finite in $\mathcal{P}_2(\mathbb{H}^\varepsilon)$ means that $\nabla_{\varepsilon} \rho$ is also well defined and that there exists $w^\varepsilon \in L^2_\varepsilon(\mu)$ such that $\rho w^\varepsilon = \nabla_{\varepsilon} \rho$. Thus the weak derivative $\mathbf{U}\rho$ is well-defined and satisfies $\varepsilon(\mathbf{U}\rho)\varepsilon\mathbf{U} = \nabla_{\varepsilon} \rho - \nabla_{\mathbb{H}} \rho$. Moreover the function $w^{\mathbf{U}}$ defined by $\varepsilon w^{\mathbf{U}}(\varepsilon\mathbf{U}) = w^\varepsilon - w^{\mathbb{H}}$ is in $L^2(\mu)$. In fact $\rho w^{\mathbf{U}} = \mathbf{U}\rho$ \mathcal{L} -almost everywhere.

We can show $w^{\mathbb{H}} \in \operatorname{Tan}_{\mathbb{H}}(\mu)$ from the fact that $w^\varepsilon \in \operatorname{Tan}_\varepsilon(\mu)$. Indeed, it is possible to approach w^ε in $L^2(\mathbb{H}^\varepsilon)$ by a sequence $(\nabla_\varepsilon \psi_k)_{k \in \mathbb{N}}$ where every ψ_k is a

function of $\mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$. It follows that $(\nabla_{\mathbb{H}}\psi_k)_{k \in \mathbb{N}}$ tends to the horizontal part of w^ε . Hence $w^{\mathbb{H}}$ is in the tangent space $\text{Tan}_{\mathbb{H}}(\mu)$.

We already know the inequality $\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu)} \leq \text{Slope}(\text{Ent})(\mu)$, we will prove the opposite inequality thanks to inequality (7) in Proposition 2.5. In this inequality, we first replace every $W^\varepsilon(\mu, \nu)$ by $W(\mu, \nu)$. It is allowed because the second is greater. Then we take ε as a function of ν with $\varepsilon = W(\mu, \nu)^{1/4}$. But

$$\begin{aligned} \|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu)} &\leq \|w^\varepsilon\|_{L_{\mathbb{H}}^2(\mu)} \leq \sqrt{\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu)}^2 + (\varepsilon\|w^{\mathbb{U}}\|_{L^2(\mu)})^2} \\ &\leq \|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2} + \frac{1}{2} \frac{(\varepsilon\|w^{\mathbb{U}}\|_{L^2(\mu)})^2}{\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu)}} \end{aligned}$$

because the graph of $\sqrt{\cdot}$ is under the tangent line in $\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2}^2$. It follows

$$\begin{aligned} \text{Ent}(\nu) &\geq \text{Ent}(\mu) - \left[\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2} + \frac{W(\mu, \nu)^{1/2}}{2\|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2}} \|w^{\mathbb{U}}\|_{L^2(\mu)}^2 \right] \cdot W(\mu, \nu) \\ &\quad - (1/4)W(\mu, \nu)^{3/2} \\ &\geq \text{Ent}(\mu) - \|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2} \cdot W(\mu, \nu) - O(W(\mu, \nu)^{3/2}). \end{aligned}$$

when ν tends to μ . Thus $\text{Slope}(\text{Ent})(\mu) \leq \|w^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu)}$ and the equality follows. \square

Remark 3.5. In the previous proof, we have established a HWI inequality (see Remark 2.7) for the subRiemannian \mathbb{H}_n . Nevertheless it only holds for the points μ such that $I_\varepsilon(\mu)$ is finite. Note that this condition is different from $I_{\mathbb{H}}(\mu) < +\infty$.

4. PROOF OF THE THEOREM

In this section, we prove Theorem 0.1 in two parts, corresponding to the two inclusions. In paragraph 4.1 we prove the direct inclusion and in paragraph 4.2 the indirect inclusion.

4.1. Heat diffusion is a gradient flow of Ent. Let $(\mu_t)_{t \in [0, T[}$ be a curve in $\mathcal{P}_2(\mathbb{H}_n)$ such that $\mu_t = \rho_t \mathcal{L} = \mu_0 * \mathfrak{h}_t$ for any $t \in]0, T[$. We will show that both the curve (μ_t) and $E : t \in]0, T[\mapsto \text{Ent}(\mu_t)$ are locally absolutely continuous and compute \dot{E} . We will also compute the slope of Ent at μ_t and estimate the metric speed of μ_t . We will obtain

$$(11) \quad \begin{cases} \dot{E}(t) = - \left\| \frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \right\|_{L_{\mathbb{H}}^2(\mu_t)}^2 \\ \text{Slope}(\text{Ent})(\mu_t) = \left\| \frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \right\|_{L_{\mathbb{H}}^2(\mu_t)} \\ |\dot{\mu}_t| \leq \left\| \frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \right\|_{L_{\mathbb{H}}^2(\mu_t)} \end{cases}$$

at almost every $t \in]0, T[$. But because of the differentiability properties of $E(t)$, we have the chain rule estimate

$$(12) \quad |\dot{E}(t)| \leq \text{Slope}(\text{Ent})(\mu_t) \cdot |\dot{\mu}_t|$$

at almost every t . Comparing with system (11), we see that inequality (12) is an equality for almost every $t \in [0, T[$.

Assuming (11) we have proved that $t \mapsto \mu_t$ is a gradient flow of Ent on $]0, T[$. For the result on $[0, T[$, according to Definition 1.7 it is sufficient to prove that the curve is locally absolutely continuous on $[0, T[$. In fact if μ_0 is a Dirac mass we have $|\dot{\mu}_t| = \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \mathfrak{h}_t\|_{\mathbb{H}}^2}{\mathfrak{h}_t}} = \sqrt{\frac{1}{t}} \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \mathfrak{h}_1\|_{\mathbb{H}}^2}{\mathfrak{h}_1}}$ so that $\int_0^{t_1} |\dot{\mu}_t| dt < +\infty$ for every $t_1 > 0$. For general μ_0 the estimate $|\dot{\mu}_t| \leq \sqrt{\frac{1}{t}} \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \mathfrak{h}_1\|_{\mathbb{H}}^2}{\mathfrak{h}_1}}$ holds as well. Actually for $t_0, t_1 > 0$ a coupling π_{t_0, t_1} between μ_{t_0} and μ_{t_1} can be defined as $P_{\#}(\mu_0 \otimes \pi^0)$ with the map $P : (p, (q_0, q_1)) \in \mathbb{H}_n \times (\mathbb{H}_n)^2 \mapsto (p \cdot q_0, p \cdot q_1)$ and an optimal coupling π^0 between $\mathfrak{h}_{t_0} \mathcal{L}$ and $\mathfrak{h}_{t_1} \mathcal{L}$. It follows

$$W(\mu_{t_0}, \mu_{t_1}) \leq W(\mathfrak{h}_{t_0} \mathcal{L}, \mathfrak{h}_{t_1} \mathcal{L}) \leq \int_{t_0}^{t_1} \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \mathfrak{h}_t\|_{\mathbb{H}}^2}{\mathfrak{h}_t}} d\mathcal{L} dt.$$

4.1.1. *Computation of \dot{E} .* We start with the formal computation of \dot{E} . It shows the different steps of the rigorous computation. Setting $H(\rho) = \rho \ln(\rho)$,

$$\begin{aligned} \left\langle \frac{d}{dt} \text{Ent}(\mu_t) \right\rangle &= \int \frac{dH(\rho_t)}{dt} d\mathcal{L} \\ &= \int (1 + \ln(\rho_t)) \Delta_{\mathbb{H}} \rho_t \\ &= \int \langle -\nabla_{\mathbb{H}}(1 + \ln(\rho_t)) \mid \nabla_{\mathbb{H}} \rho_t \rangle_{\mathbb{H}} \\ &= - \int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L}. \end{aligned}$$

There are several problems. Among them the vector field $(1 + \ln(\rho_t)) \nabla_{\mathbb{H}} \rho_t$ does not have a compact support so that it is not sure that the full integral of its derivatives is zero, which is what we need for an integration by parts. Another difficulty is about the regularity of $E : t \mapsto \text{Ent}(\mu_t)$ and the possibility to differentiate under the integral sign. We will avoid those technical problems by defining an approximation of Ent by some functionals $\int H_k(\rho)$ where the functions H_k are defined thanks to cut-off functions.

For any $k \geq 3$, let $\psi_k : \mathbb{R}^+ \rightarrow [0, 1]$ be a cut-off function defined by

$$\psi_k(x) = \begin{cases} 1 & \text{if } x \in [1/k, k] \\ 0 & \text{if } x \notin [1/(k+1), k+1]. \end{cases}$$

We consider $H'_k(x) = \int_{1/k}^x \psi_k(u) \frac{1}{u} du + 1 - \ln(k)$ and $H_k(x) = \int_0^x H'_k(u) du$. We will see that H_k pointwise converges to $H : x \mapsto x \ln(x)$. We start with a few remarks: note that $H_k(0) = H(0) = 0$, that H_k is convex but less convex than H because $H''(x) = \frac{1}{x} \geq \psi_k(x) \frac{1}{x} = H''_k(x)$. Note also that $H'(x) = H'_k(x)$ on $[1/k, k]$. Let $x > 0$ be a real number and k great enough to have $x \in [1/k, k]$ so that H and H_k are “parallel” on this interval. Then $H_k(x) - H(x) = H_k(1/k) - H(1/k)$. But

$$\frac{1}{k} \ln(1/k) = H(1/k) \leq H_k(1/k) \leq 0,$$

because $H'(u) \leq H'_k(u) \leq 0$ for any $u \leq 1/k$. All these terms tend to 0 so that we have the pointwise convergence.

Let us now prove a domination property. The non-negative function

$$K : x \mapsto \begin{cases} |x \ln(x)| & \text{if } x \leq 1/3 \\ 1 + |x \ln(x)| & \text{if } x > 1/3. \end{cases}$$

uniformly dominates the functions $|H_k|$ because on $[0, 1/3]$, $H_k(x)$ is negative and $H_k(x) - H(x) = H_k(x) - x \ln(x) \geq 0$ and on $[1/3, +\infty[$ we have $H_k(x) - x \ln(x) \leq \frac{\ln(k)}{k} \leq 1$ and $H_k \geq -e^{-1} > -1$. Moreover if ρ is the density of some $\mu \in \mathcal{P}_2(\mathbb{H})$ satisfying $\text{Ent}(\mu) < +\infty$, we have $\int K(\rho) d\mathcal{L} < +\infty$ because $\int |\rho \ln(\rho)| d\mathcal{L} < \infty$ and $\int_{\{\rho > 1/3\}} 1 d\mathcal{L} < 3$. Thus using the dominated convergence theorem one has

$$\lim_{k \rightarrow +\infty} \int H_k(\rho) d\mathcal{L} = \int \rho \ln(\rho) d\mathcal{L} = \text{Ent}(\mu).$$

The cut-off functions ψ_k give us the possibility to make the former formal computation correct.

$$(13) \quad \frac{d}{dt} \int H_k(\rho_t) = \int \frac{dH_k(\rho_t)}{dt} d\mathcal{L} \\ = \int H'_k(\rho_t) \Delta_{\mathbb{H}} \rho_t$$

$$(14) \quad = - \int \psi_k(\rho_t) \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L}.$$

Let us see the justifications. For (13) we do not have to much difficulties because $H_k(\rho_t)$ has derivative $H'_k(\rho_t) \Delta_{\mathbb{H}} \rho_t$ and absolute value $g_t = |H'_k(\rho_t)(\rho_0 * \Delta_{\mathbb{H}} \mathfrak{h}_t)|$, the functions H'_k and $\Delta_{\mathbb{H}} \mathfrak{h}_t$ are bounded and ρ_0 is integrable. That together proves that g_t is integrable. Moreover $x \mapsto \sup_{t \in I} g_t(x)$ is also integrable for compact intervals I included in $]0, +\infty[$ (the $\sup_{t \in I} |\Delta_{\mathbb{H}} \mathfrak{h}_t|$ is integrable, see Proposition 1.3). It follows that $t \mapsto \int H_k(\rho_t)$ is in $\mathcal{C}^1(]0, +\infty[)$. Let us now justify the integration by part (14). In fact we have already shown that $H'_k(\rho_t) \Delta_{\mathbb{H}} \rho_t$ is integrable. But we will see that $\text{div}(H'_k(\rho_t) \nabla_{\mathbb{H}} \rho_t)$ is also integrable and thanks to Lemma 4.1 that this integral is 0, which will justify the integration by part (14). But $\psi_k(\rho_t) \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} \leq (k+1) \sup(\|\nabla_{\mathbb{H}} \mathfrak{h}_t\|^2) \chi_{\rho_t \geq \frac{1}{k+1}}$. This is enough for the integrability of $\text{div}(H'_k(\rho_t) \nabla_{\mathbb{H}} \rho_t)$. Observe that $(z, u) \in \mathbb{H}_n \mapsto (1 + \|z\|_{\mathbb{C}^n})(H'_k(\rho_t) \nabla_{\mathbb{H}} \rho_t)(z, u)$ is also integrable for similar reasons (H'_k is bounded, $\mu \in \mathcal{P}_2(\mathbb{H}_n)$, $\nabla_{\mathbb{H}} \rho_t = \rho_0 * \nabla_{\mathbb{H}} \mathfrak{h}_t$ and Proposition 1.3). Hence we just need to apply Lemma 4.1 to $V = H'_k(\rho_t) \nabla_{\mathbb{H}} \rho_t$ because for any smooth ρ , $\|\nabla_{\mathbb{H}} \rho\|_{\mathbb{R}^{2n+1}}(z, u) \leq 4n^2(1 + |z|) \cdot \|\nabla_{\mathbb{H}} \rho\|_{\mathbb{H}}$.

Lemma 4.1. *Let V a smooth horizontal vector such that $\|V\|_{\mathbb{R}^{2n+1}} \in L^1(\mathbb{H}_n)$. Suppose also that $\text{div } V$ is integrable. Then $\int \text{div } V = 0$.*

Proof. Let $\mathcal{B}(r)$ and $\mathcal{S}(r)$ be the Euclidian balls and spheres of radius r and center 0 in \mathbb{R}^{2n+1} . Then $|\int_{\mathcal{B}(r)} \text{div } V| \leq \int_{\mathcal{S}(r)} \|V\|$ where the norm is the one of \mathbb{R}^{2n+1} . But $\iint_{\mathcal{S}(r)} \|V\| d\sigma dr = \int \|V\| < +\infty$. Hence there is a sequence (r_n) with $r_n \rightarrow +\infty$ and $\int_{\mathcal{S}(r_n)} \|V\| \rightarrow 0$. It follows $\int \text{div } V = 0$. \square

Hence

$$\int H_n(\rho_{t_0}) - \int H_n(\rho_{t_1}) = \int_{t_0}^{t_1} \left(\int \psi_n \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L} \right) dt.$$

The last step is to let n go to ∞ and use the monotone convergence theorem.

$$(15) \quad \text{Ent}(\mu_{t_0}) - \text{Ent}(\mu_{t_1}) = \int_{t_0}^{t_1} \left(\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L} \right) dt.$$

Note that in (15) the entropies are finite. Actually, from Lemma 1.5, we have $\text{Ent} > -\infty$ and ρ_t is bounded from above by $\sup \mathfrak{h}_t$, so that $\rho_t \ln \rho_t \leq C \rho_t$ for some positive constant C . This implies $\text{Ent}(\mu_t) \leq C < +\infty$. Hence $E : t \mapsto \text{Ent}(\mu_t)$ is locally absolutely continuous and for almost every t ,

$$-\dot{E}(t) = \int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L} < +\infty.$$

4.1.2. *Computation of the slope of Ent.* We will apply Proposition 3.4 to μ_t for almost every t . We have already proved in (15) that $\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L}$ is finite. Consider the Cauchy-Schwarz inequality

$$\begin{aligned} (\rho_0 * |\mathbf{U}\mathfrak{h}_t|)^2 &= \left(\rho_0 * \left(\frac{|\mathbf{U}\mathfrak{h}_t|}{\sqrt{\mathfrak{h}_t}} \sqrt{\mathfrak{h}_t} \right) \right)^2 \\ &\leq \left(\rho_0 * \frac{|\mathbf{U}\mathfrak{h}_t|^2}{\mathfrak{h}_t} \right) (\rho_0 * \mathfrak{h}_t). \end{aligned}$$

Finally

$$\frac{|\mathbf{U}\rho_t|^2}{\rho_t} \leq \left(\frac{(\rho_0 * |\mathbf{U}\mathfrak{h}_t|)^2}{\rho_0 * \mathfrak{h}_t} \right) \leq \rho_0 * \left(\frac{|\mathbf{U}\mathfrak{h}_t|^2}{\mathfrak{h}_t} \right).$$

After integrating one obtains

$$\int \frac{|\mathbf{U}\rho_t|^2}{\rho_t} \leq \int \frac{|\mathbf{U}\mathfrak{h}_t|^2}{\mathfrak{h}_t},$$

because \mathcal{L} is translation-invariant. Proposition 1.3 and Proposition 3.4 gives us $\text{Slope}(\text{Ent})(\mu_t) = \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L}}$.

4.1.3. *Computation of the speed of (μ_t) .* Let us finally consider the speed of $(\mu_t)_{t \in]0, T[}$. Let $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$. We recall that ρ_t is positive, smooth in space and time for $t > 0$. Moreover, $\psi \nabla_{\mathbb{H}} \rho_t$ has a compact support. Then

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{H}} \psi \rho_t d\mathcal{L} &= \int_{\mathbb{H}} \psi \partial_t \rho_t d\mathcal{L} = \int_{\mathbb{H}} \psi \Delta_{\mathbb{H}} \rho_t d\mathcal{L} \\ &= \int_{\mathbb{H}} \langle \nabla_{\mathbb{H}} \psi \mid -\nabla_{\mathbb{H}} \rho_t \rangle_{\mathbb{H}} d\mathcal{L} = \int_{\mathbb{H}} \langle \nabla_{\mathbb{H}} \psi \mid -\frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \rangle_{\mathbb{H}} d\mu_t. \end{aligned}$$

Hence $\partial_t(\mu_t) + \text{div} \left(-\frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \mu_t \right) = 0$ holds for every t . Recall from (15) that $-\frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t}$ is in $L_{\mathbb{H}}^2(\mu_t)$ and $\left\| \frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \right\|_{L_{\mathbb{H}}^2(\mu_t)}^2$ is locally integrable on $]0, T[$. Using now the converse part of Proposition 3.1 we can conclude that $|\dot{\mu}_t| \leq \left\| \frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \right\|_{L_{\mathbb{H}}^2(\mu_t)}$ holds at almost every time.

Thus we have computed the speed of μ_t , the slope of Ent at μ_t and \dot{E} . See the beginning of the subsection for the end of the proof.

4.2. Gradient flows of Ent are heat diffusions. We want to prove that gradient flows $(\mu_t)_{t \in]0, T[}$ of Ent in $\mathcal{P}_2(\mathbb{H}_n)$ are solutions of the hypoelliptic heat equation. We will state two important propositions, Proposition 4.5 and Proposition 4.6, that are based on the following assumption:

$$(16) \quad \int_{t_0}^{t_1} \|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} \cdot |\dot{\mu}_t| dt \quad \text{is finite.}$$

with the conventions $\|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} = +\infty$ each time $\text{Slope}(\text{Ent})(\mu_t)$ is infinite and $|\dot{\mu}_t| = +\infty$ if the metric derivative of $(\mu_t)_t$ in $\mathcal{P}_2(\mathbb{H}_n)$ does not exist at t . Before we prove that gradient flows of Ent satisfy (16), let us stress that we do not need these conventions in the case of a gradient flow, because $\text{Ent}(\mu_t)$ is finite (except maybe at one time) and $|\dot{\mu}_t| = \text{Slope}(\text{Ent})(\mu_t)$ is finite for almost every t . Furthermore $\|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} < +\infty$ by Proposition 3.2.

Lemma 4.2. *Let T be a finite or infinite time. Gradient flows $(\mu_t)_{t \in]0, T[}$ of Ent in $\mathcal{P}_2(\mathbb{H}_n)$ are AC_2 -curves on the open set $]0, T[$ and they satisfy condition (16) for any positive bounds t_0, t_1 .*

Proof. As $(\mu_t)_{t \in]0, T[}$ is a gradient flow of Ent in $\mathcal{P}_2(\mathbb{H}_n)$ we have

$$0 \leq \text{Ent}(\mu_{t_0}) - \text{Ent}(\mu_{t_1}) < +\infty$$

for any $0 < t_0 < t_1 < T$. Indeed because of Lemma 1.5, $\text{Ent}(\mu_t) > -\infty$ and it can also not be $+\infty$ as explained in Remark 1.9. Therefore, as $E : t \mapsto \text{Ent}(\mu_t)$, that is a non-increasing function, has bounded variations we obtain

$$\int_{t_0}^{t_1} [\text{Slope}(\text{Ent})(\mu_t)] \cdot |\dot{\mu}_t| = - \int_{t_0}^{t_1} \dot{E} \leq E(t_0) - E(t_1) < +\infty.$$

But $\|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} \leq \text{Slope}(\text{Ent})(\mu_t)$ so that $\int \|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} \cdot |\dot{\mu}_t| < +\infty$.

The curve $t \mapsto \mu_t$ is an AC_2 -curve on $]0, T[$ in $\mathcal{P}_2(\mathbb{H}_n)$ because it satisfies $|\dot{\mu}|^2 = [\text{Slope}(\text{Ent})(\mu_t)] \cdot |\dot{\mu}_t|$, which is integrable on any $[t_0, t_1] \subset]0, T[$. \square

We will prove a result similar to Proposition 2.8 for the Heisenberg group in the two propositions of this paragraph. For that, we will use two kinds of approximations: the first one is the comparison of \mathbb{H}_n with \mathbb{H}_n^ε and the second one is the approximation of measures μ_t by some others μ_t^* in the following way: the measure μ_t^* is obtained by convolution with a measure ξ_v concentrated on the center $L = \{(z, u) \in \mathbb{H}_n, |z| = 0\}$ and with density $u \mapsto \frac{\xi(u/v)}{v}$ where ξ is a function of support $[-1, 1] \subset L$ that differs from

$$u \mapsto e^{\frac{-1}{1-u^2}}$$

only by a normalization constant. We will simply write μ_t^* even if it depends on $v > 0$. While in the proofs $\varepsilon > 0$ is a fixed parameter, μ^* depends in fact on $v > 0$ and μ^* tends to μ while this parameter goes to 0. Recall Lemma 1.2 on the possibility to differentiate one or the other term of the convolution in μ_t^* . The convolution of a measure on \mathbb{H}_n with an horizontal vector field has to be understood in the sense of left- or right- translation of vectors ; for ξ_v because of Lemma 1.2 left or right means the same and the convolution with functions or vector fields is the same as the convolution in \mathbb{R}^{2n+1} .

We will apply many times the following lemma with one of the measures ξ_ν or μ_t in the role of ν .

Lemma 4.3. *Let μ be a probability measure on \mathbb{R}^N with density ρ and f a μ -measurable function. Let also ν be a probability measure. Define $F = \rho f$. Then*

$$\|F * \nu / \mu * \nu\|_{L^2(\mu * \nu)} \leq \|F / \mu\|_{L^2(\mu)} = \|f\|_{L^2(\mu)}.$$

Moreover if ν_n converges in law to the Dirac mass in 0 and $\|f\|_{L^2(\mu)} < +\infty$, the sequence $\|F * \nu_n / \mu * \nu_n\|_{L^2(\mu * \nu_n)}$ converges and has limit $\|f\|_{L^2(\mu)}$.

Proof. The result is trivial if f is not in $L^2(\mu)$. So we assume that it belongs to this space. Hence it is also in $L^1(\mu)$ so that one can identify F with $f\mu$, a finite Radon measure. The wanted estimate relies on the Cauchy-Schwarz inequality.

$$\begin{aligned} (|F| * \nu)^2 &= \left(\left(\frac{|F|}{\sqrt{\rho}} \sqrt{\rho} \right) * \nu \right)^2 \\ &\leq \left(\frac{|F|^2}{\rho} * \nu \right) (\rho * \nu). \end{aligned}$$

But $|F * \nu| \leq |F| * \nu$, then

$$\int \frac{|F^*|^2}{\rho^*} d\mathcal{L} \leq \int \frac{|F|^2}{\rho} d\mathcal{L}.$$

For the second part we consider the fact that $F * \nu_n$ and $\mu * \nu_n$ weakly* converges to F and μ respectively. Moreover $f \in \mathbb{R} \mapsto f^2 \in \mathbb{R}$ is superlinear. Hence we have gathered the assumptions for Theorem 2.34 of [1] and can apply it as in Example 2.36 of the same book. More precisely the theorem is applied to the finite Radon measures F^* and F . It follows that

$$\liminf_{\nu \downarrow 0} \int \frac{|F^*|^2}{\rho^*} d\mathcal{L} \geq \int \frac{|F|^2}{\rho} d\mathcal{L}.$$

□

Remark 4.4. The first part of Lemma 4.3 comes from the Cauchy-Schwarz inequality. Alternatively it follows from the Blachman–Stam inequality [8, 43] or the Jensen inequality as in [6, Lemma 2.13 and Lemma 2.14].

We can now start with the two propositions.

Proposition 4.5. *Let T be a positive or infinite time and $(\mu_t)_{t \in]0, T[}$ be an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n)$. Let us assume that $\text{Ent}(\mu_t)$ is finite for any t and that for any $t_0, t_1 > 0$, condition (16) is satisfied:*

$$\int_{t_0}^{t_1} \|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} \cdot |\dot{\mu}_t| dt < +\infty.$$

(About this condition recall also the convention $\|w_t^{\mathbb{H}}\| = +\infty$ if the slope of Ent in μ_t is not finite.) Then $t \mapsto \text{Ent}(\mu_t)$ is locally absolutely continuous. Moreover

$$|\text{Ent}(\mu_{t_1}) - \text{Ent}(\mu_{t_0})| \leq \int_{t_0}^{t_1} \|w_t^{\mathbb{H}}\|_{L_{\mathbb{H}}^2(\mu_t)} \cdot |\dot{\mu}_t| dt$$

for any $t_0, t_1 > 0$.

Proof. We will apply Proposition 2.8 to the curves $(\mu_t^*)_{t \in]0, T[}$ built as explained above from $(\mu_t)_{t \in]0, T[}$. First we are indeed mostly interested in the fact that $E^* : t \mapsto \text{Ent}(\mu_t^*)$ is locally absolutely continuous. We will prove this by using the theory on $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$ where $\varepsilon > 0$ is chosen arbitrary. Note that $\nabla_{\mathbb{H}}(\rho * \xi_v) = (\nabla_{\mathbb{H}}\rho) * \xi_v$ and $\nabla_\varepsilon(\rho * \xi_v) = (\nabla_\varepsilon\rho) * \xi_v$ because of Lemma 1.2.

One thing we want to check in order to apply Proposition 2.8 is that $t \mapsto \mu_t^*$ is an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$. This is quite obvious because as explained in Lemma 1.2, $d_\varepsilon(p, q) = d_\varepsilon(p.l, q.l)$ for any $l \in L$. It follows $W^\varepsilon(\mu_s^*, \mu_{s'}) \leq W^\varepsilon(\mu_s, \mu_{s'})$ because each transport plan π of μ_s and $\mu_{s'}$ gives rise to a transport plan π^* defined by $\int \phi(p, q) d\pi^*(p, q) = \int \phi(p.l, q.l) d\pi(p, q) \otimes d\xi_v(l)$. Hence we have

$$W^\varepsilon(\mu_s^*, \mu_{s'}^*) \leq W^\varepsilon(\mu_s, \mu_{s'}) \leq W(\mu_s, \mu_{s'}) \leq \int_s^{s'} |\dot{\mu}_t| dt.$$

Therefore $t \mapsto \mu_t^*$ is locally absolutely continuous. Moreover it holds $|\dot{\mu}_t^*|_\varepsilon \leq |\dot{\mu}_t|$, t -almost everywhere. Thus $t \mapsto \mu_t^*$ is an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n^\varepsilon)$.

Another fact that we have to check before applying Proposition 2.8 concerns the slope of Ent at point μ_t^* in $\mathcal{P}_2(\mathbb{H}_\varepsilon)$. In fact we will state

$$(17) \quad \int \frac{\|\nabla_\varepsilon \rho_t^*\|_\varepsilon^2}{\rho_t^*} = \int \frac{\|\nabla_{\mathbb{H}} \rho_t^*\|_{\mathbb{H}}^2}{\rho_t^*} + \varepsilon^2 \int \frac{|\mathbf{U}\rho_t^*|^2}{\rho_t^*} \leq \int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} + \varepsilon^2 \int \frac{|\mathbf{U}\xi_v|^2}{\xi_v},$$

with the convention that an integral is $+\infty$ when the integrand is not well-defined. Actually as $\text{Ent}(\mu_t) < +\infty$ the density ρ_t exists but the vectors of type $\nabla\rho/\rho$ may not be well-defined as elements of some L^2 -space (recall Remark 2.2 and paragraph 3.2).

The equality in (17) is only the definition of ∇_ε and the inequality is the consequence of two estimates. For the first estimate we apply Lemma 4.3 to $\frac{\mathbf{U}\xi_v}{\xi_v}$ and μ_t , for the second estimate we apply the same lemma to the coordinates of $\frac{\nabla_{\mathbb{H}}\rho_t}{\rho_t}$ and the measure ξ_v . It follows

$$\sqrt{\int \frac{\|\nabla_\varepsilon \rho_t^*\|_\varepsilon^2}{\rho_t^*}} \cdot |\dot{\mu}_t^*|_\varepsilon \leq \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t}} \cdot |\dot{\mu}_t| + \varepsilon \sqrt{\int \frac{|\mathbf{U}\xi_v|^2}{\xi_v}} \cdot |\dot{\mu}_t^*|_\varepsilon.$$

Thence we can apply Proposition 2.8 because the last estimate shows that $t \mapsto \text{Slope}^\varepsilon(\text{Ent})(\mu_t^*) \cdot |\dot{\mu}_t^*|_\varepsilon$ is integrable. Therefore E^* is locally absolutely continuous. We can now turn to the second part of the proposition because for almost every t , the derivative of $E^* : t \mapsto \text{Ent}(\mu_t^*)$ is smaller than $\text{Slope}^{\mathbb{H}}(\text{Ent})(\mu_t^*) \cdot |\dot{\mu}_t^*|$. But because of estimate (17) and Proposition 3.4, $\text{Slope}^{\mathbb{H}}(\text{Ent})(\mu_t^*) = \sqrt{\int \frac{|\nabla_{\mathbb{H}} \rho_t^*|^2}{\rho_t^*}}$. It follows

$$\begin{aligned} |\text{Ent}(\mu_{t_1}^*) - \text{Ent}(\mu_{t_0}^*)| &\leq \int_{t_0}^{t_1} \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho_t^*\|_{\mathbb{H}}^2}{\rho_t^*}} d\mathcal{L} \cdot |\dot{\mu}_t^*| dt \\ &\leq \int_{t_0}^{t_1} \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t}} d\mathcal{L} \cdot |\dot{\mu}_t| dt. \end{aligned}$$

Letting v go to 0 in the left-hand we obtain both results we wish: the absolute continuity of E and the estimate. In fact if $\mu^{\mathbb{C}}$ and $(\mu_t|_z)_{z \in \mathbb{C}^n}$ (resp. $(\mu_t^*|_z)_{z \in \mathbb{C}^n}$) are the marginal and the conditional measures of μ_t (resp. μ_t^*) in the projection

$(z, u) \mapsto z$, this disintegration provides

$$(18) \quad \text{Ent}(\mu_t^*) - \text{Ent}(\mu_t) = \int_{\mathbb{C}^n} \left(\text{Ent}(\mu_{t|z}^*) - \text{Ent}(\mu_{t|z}) \right) d\mu_t^{\mathbb{C}}(z),$$

both for $t = t_1$ and $t = t_0$. By the dominated convergence theorem, this goes to 0 as v tends to 0. \square

Let us immediately improve the upper bound of Proposition 4.5.

Proposition 4.6. *Let T be a positive or infinite time and $(\mu_t)_{t \in]0, T[}$ be an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}^n)$ and $(v_t)_{t \in]0, T[}$ be μ_t -measurable vector fields given by Proposition 3.1. Under the hypothesis that $\text{Ent}(\mu_t)$ is finite for any t and (16) holds for any $t_0, t_1 > 0$, we have*

$$(19) \quad \text{Ent}(\mu_{t_0}) - \text{Ent}(\mu_{t_1}) \leq \int_{t_0}^{t_1} \int \left\langle \frac{-\nabla_{\mathbb{H}} \rho_t}{\rho_t} \mid v_t \right\rangle_{\mathbb{H}} d\mu_t dt$$

for any $t_0, t_1 > 0$.

Proof. We divide the proof in three steps.

Continuity equation (8) for $(\mu_t^*)_{t > 0}$: Let (μ_t) be a locally absolutely continuous curve such that $\int_{t_0}^{t_1} \sqrt{\int \frac{\|\nabla_{\mathbb{H}} \rho_t\|_{\mathbb{H}}^2}{\rho_t} d\mathcal{L}} \cdot |\dot{\mu}_t| dt$ is finite for any $t_0, t_1 > 0$. Observe that both μ_t and μ_t^* satisfy the assumptions of Proposition 4.5 as we have seen in the proof of this proposition itself. Hence we know that E and E^* are locally absolutely continuous and $w_t^{\mathbf{U},*} = \mathbf{U}\rho_t^*/\rho_t^*$ is in $L^2(\mu_t^*)$. We will start to prove (19) for μ_t^* together with the μ_t^* -measurable vector fields $(v_t^*)_{t \in I}$ defined by $v_t^* \rho_t^* := (\rho_t v_t) * \xi_v$ for any $t \in]0, T[$. We first check that v_t^* is a speed vector of μ_t^* in the sense of (8). We mean that it satisfies the continuity equation (8). Let $\psi \in C_c^\infty(\mathbb{R}^{2n+1})$ be a test function. Then $\int \psi d\mu_t^* = \int (\psi * \xi_v) d\mu_t$. As $\psi * \xi_v \in C_c^\infty(\mathbb{R}^{2n+1})$, the derivative of this quantity is

$$\int \langle \nabla_{\mathbb{H}}(\psi * \xi_v) \mid v_t \rangle_{\mathbb{H}} d\mu_t = \int \langle \nabla_{\mathbb{H}} \psi \mid (\rho_t v_t) * \xi_v \rangle_{\mathbb{H}} d\mathcal{L} = \int \langle \nabla_{\mathbb{H}} \psi \mid v_t^* \rangle_{\mathbb{H}} d\mu_t^*.$$

Because of Lemma 4.3 applied to the coordinates of $\rho_t v_t$ and ξ_v , we have $\int \|v_t^*\|_{\mathbb{H}}^2 d\mu_t^* \leq \int \|v_t\|_{\mathbb{H}}^2 d\mu_t = |\dot{\mu}_t|^2$ (but we do not know whether $\int \|v_t^*\|_{\mathbb{H}}^2 d\mu_t^* \leq |\dot{\mu}_t^*|^2$).

The curve $(\mu_t^*)_{t > 0}$ satisfies (19): First notice in (16) that if $v_t = 0$ for almost every t , both \dot{E} and \dot{E}^* are zero such that $-\dot{E}^* \leq \int \left\langle \frac{-\nabla_{\mathbb{H}} \rho_t^*}{\rho_t^*} \mid v_t^* \right\rangle_{\mathbb{H}} d\mu_t^*$. Generally under

(16), $\text{Slope}(\text{Ent})(\mu_t)$ and $\|w_t^{\mathbb{H}}\|_{L^2_{\mathbb{H}}(\mu_t)}$ are finite on $\{t \in]0, T[, |\mu_t| \neq 0\}$ for almost every t . Thus at those times t we can apply Proposition 2.1 and Proposition 2.5 to μ_t^* because $w_t^{\mathbb{H}} = \nabla_{\mathbb{H}} \rho_t / \rho_t$ is square-integrable. Indeed we have proved in Proposition 4.5 (especially estimate (17)) that this condition implies that $\nabla_{\varepsilon} \rho_t^* / \rho_t^*$ is in $L^2_{\varepsilon}(\mu_t^*)$.

From there with Proposition 2.5, one has $-\dot{E}^*(t) \leq \int \left\langle \frac{-\nabla_{\varepsilon} \rho_t^*}{\rho_t^*} \mid v_t^{\varepsilon,*} \right\rangle_{\varepsilon} d\mathcal{L}$ for a vector field $v_t^{\varepsilon,*}$ as in (6). As $w_t^{\varepsilon,*} = w_t^{\mathbb{H},*} + \varepsilon^2 w_t^{\mathbf{U},*} = \nabla_{\varepsilon} \rho_t^* / \rho_t^*$ is in $L^2_{\varepsilon}(\mu_t^*)$, according to Proposition 2.1 it is also a vector field of $\text{Tan}_{\varepsilon}(\mu_t^*)$. Thence let $(\psi_k)_{k \in \mathbb{N}}$ be a sequence converging to $w_t^{\varepsilon,*}$ in $L^2_{\varepsilon}(\mu_t^*)$ where the $(\psi_k)_{k \in \mathbb{N}}$ are test functions. Therefore the sequence $(\nabla_{\mathbb{H}} \psi_k)_{k \in \mathbb{N}}$ tends to $w_t^{\mathbb{H},*}$ and the derivative of $t \mapsto \int \psi_k d\mu_t^*$ has two limits, the first one due to the theory in $\mathcal{P}_2(\mathbb{H}_n^{\varepsilon})$, the other to the one in $\mathcal{P}_2(\mathbb{H}_n)$. One obtains

$$\lim_{k \rightarrow \infty} \frac{d}{dt} \int \psi_k d\mu_t^* = \lim_{k \rightarrow \infty} \int \langle \nabla_{\mathbb{H}} \psi_k \mid v_t^* \rangle_{\mathbb{H}} d\mu_t^* = \int \left\langle \frac{\nabla_{\mathbb{H}} \rho_t^*}{\rho_t^*} \mid v_t^* \right\rangle_{\mathbb{H}} d\mu_t^*.$$

and

$$\lim_{k \rightarrow \infty} \frac{d}{dt} \int \psi_k d\mu_t^* = \lim_{k \rightarrow \infty} \int \langle \nabla_\varepsilon \psi_k \mid v_t^{\varepsilon,*} \rangle_\varepsilon d\mu_t^* = \int \left\langle \frac{\nabla_\varepsilon \rho_t^*}{\rho_t^*} \mid v_t^{\varepsilon,*} \right\rangle_\varepsilon d\mu_t^*.$$

Hence for the locally absolutely continuous E^* we have

$$(20) \quad \text{Ent}(\mu_{t_0}^*) - \text{Ent}(\mu_{t_1}^*) = \int_{t_0}^{t_1} (-\dot{E}^*) \leq \int_{t_0}^{t_1} \int \left\langle \frac{-\nabla_{\mathbb{H}} \rho_t^*}{\rho_t^*} \mid v_t^* \right\rangle_{\mathbb{H}} d\mu_t^* dt.$$

The curve $(\mu_t)_{t>0}$ satisfies (19): We would like now to let the $*$ -parameter v go to 0. This will be done if $\|w_t^*\|_{L^2_{\mathbb{H}}(\mu_t^*)}$, $\|v_t^*\|_{L^2_{\mathbb{H}}(\mu_t^*)}$ and $\|w_t^* + v_t^*\|_{L^2_{\mathbb{H}}(\mu_t^*)}$ all three converge to $\|w_t\|_{L^2_{\mathbb{H}}(\mu_t)}$, $\|v_t\|_{L^2_{\mathbb{H}}(\mu_t)}$ and $\|w_t + v_t\|_{L^2_{\mathbb{H}}(\mu_t)}$ respectively, because it would mean that $\int \left\langle \frac{\nabla_{\mathbb{H}} \rho_t^*}{\rho_t^*} \mid v_t^* \right\rangle_{\mathbb{H}} d\mu_t^*$ tends to $\int \left\langle \frac{\nabla_{\mathbb{H}} \rho_t}{\rho_t} \mid v_t \right\rangle_{\mathbb{H}} d\mu_t$. This is attested by the equality part of Lemma 4.3 applied to the coordinates of v_t , w_t and $v_t + w_t$. We conclude as in Proposition 4.5 for the left-hand side of (20). For the right-hand side we dominate $\int \left\langle \frac{\nabla_{\mathbb{H}} \rho_t^*}{\rho_t^*} \mid v_t^* \right\rangle_{\mathbb{H}} d\mu_t^*$ by $\|w_t^{\mathbb{H}}\|_{L^2(\mathbb{H})} \cdot |\dot{\mu}_t|_{\mathbb{H}}$ and let v go to 0. (Indeed $\|w_t^{\mathbb{H}}\|_{L^2(\mathbb{H})} \geq \|w_t^{\mathbb{H}*}\|_{L^2(\mathbb{H})}$ because of Lemma 4.3) \square

For the end of the proof, recall that Lemma 4.2 tells us that a gradient flow $(\mu_t)_{t \in]0, T[}$ of Ent satisfies (16) and is an AC_2 -curve in $\mathcal{P}_2(\mathbb{H}_n)$ on $]0, T[$. Therefore one can apply Proposition 4.5 and Proposition 4.6 to $(\mu_t)_{t>0}$. Hence we put the estimate of \dot{E} obtained in (20) in equation (8). Considering that $-\dot{E}(t) = \text{Slope}(\text{Ent})(\mu_t) \cdot |\dot{\mu}_t| \geq \|w_t^{\mathbb{H}}\|_{L^2(\mathbb{H})} \cdot \|v_t\|_{L^2(\mathbb{H})}$ must be smaller than $\left\langle \frac{-\nabla_{\mathbb{H}} \rho_t}{\rho_t} \mid v_t \right\rangle_{\mathbb{H}}$, one can deduce $v_t = -w_t^{\mathbb{H}} = -\nabla_{\mathbb{H}} \rho_t / \rho_t$. Replacing v_t , one can therefore rewrite the continuity equation of Proposition 3.1:

$$\frac{d}{dt} \mu_t + \text{div}(-w_t^{\mathbb{H}} \mu_t).$$

Because of relation $\nabla_{\mathbb{H}} \rho_t = \rho_t w_t^{\mathbb{H}}$ it is also

$$(21) \quad \frac{d}{dt} \mu_t + \text{div}(-\nabla_{\mathbb{H}} \rho_t d\mathcal{L})$$

where $\nabla_{\mathbb{H}} \rho_t$ is the weak gradient of ρ_t . Remind that (21) means that for any $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$,

$$(22) \quad \frac{d}{dt} \int \psi d\mu_t = \int \langle \nabla_{\mathbb{H}} \psi \mid -\nabla_{\mathbb{H}} \rho_t \rangle_{\mathbb{H}} d\mathcal{L}.$$

We know from the proof of Proposition 3.1 that $\zeta_\psi = \int \psi d\mu_t$ is locally absolutely continuous on $]0, T[$. It follows that we can integrate (22) on an interval $[t_0, t_1]$ and obtain

$$\int \psi \rho_{t_0} d\mathcal{L} - \int \psi \rho_{t_1} d\mathcal{L} = - \int_{t_0}^{t_1} \int \langle \nabla_{\mathbb{H}} \psi \mid \nabla_{\mathbb{H}} \rho_t \rangle_{\mathbb{H}} d\mathcal{L} dt$$

for any $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2n+1})$. We recognize a weak formulation of the ‘‘hypoelliptic heat equation’’. By using classical references about hypoelliptic operator as [41] or [44] and the references therein, this concludes the proof for $(\mu_t)_{t \in]0, T[}$. The continuity of the curve in 0 states that $(\mu_t)_t$ is the heat diffusion on the whole $[0, T[$.

Remark 4.7. We have proved that μ_t satisfies the hypoelliptic heat equation. But $\Delta_{\mathbb{H}}$ commutes with \mathbf{U} so that μ_t^* is actually also driven by heat diffusion in \mathbb{H}_n .

5. FINAL REMARKS

In this final section, we compare the main theorem and our techniques to similar results and other techniques.

Remark 5.1. In the last paragraph of their paper concerning a nonholonomic Moser theorem [27, Subsection 6.3], Khesin and Lee prove a result similar to one of the two inclusions of Theorem 0.1: “heat diffusion is a gradient flow of $\Delta_{\mathbb{H}}$ ”. The paper takes place in the wide class of bracket-generating distribution τ on a connected and compact manifold M (it includes the Albanese torus, a compact quotient of \mathbb{H}_n). The authors also approximate the metric space by Riemannian manifolds completing the horizontal tangent space by the other directions. However, their result is different from the comparable inclusion in Theorem 0.1 because it does not apply to the non-compact \mathbb{H}_n . The proof of Khesin and Lee is more algebraic than the proofs of the present paper and the Wasserstein space they are considering is a “smooth” Wasserstein space, restricted to smooth probability measures. Actually the authors begin to give the smooth tangential structure of $\mathcal{P}_2(M_\tau)$, and define the length of curves and the Wasserstein distance from there. Hence the definition of gradient flows is different: the solution of the hypoelliptic equation goes in the “smooth” direction with the greatest slope but a rough $\text{Slope}(\text{Ent})$ is not defined.

In [27], the other inclusion “gradient flows of Ent are heat diffusions of $\Delta_{\mathbb{H}}$ ” is not addressed.

We say a few words on the most recent results by Ambrosio, Gigli and Savaré.

Remark 5.2. In [3], Theorem 8.5 states in a very general metric setting that if $\text{Slope}(\text{Ent})$ is both — i) a strong upper gradient, ii) lower semicontinuous on the sublevels of Ent , the gradient flow of Ent exists and corresponds to the L^2 -gradient flow of the suitable Dirichlet energy (the hypoelliptic semigroup for \mathbb{H}_n). According to Theorem 9.3 in [3], conditions i) and ii) are satisfied in metric spaces “with Ricci curvature bounded from below”. In the paper, we have proved i) for the Heisenberg group (Proposition 4.4) but I do not know whether ii) holds.

According to recent discussions with the authors, even the assumptions of Theorem 8.5 in [3] could be impressively removed. Only the basic estimate (4.3) of this paper — on the growth of balls — seems to be necessary. This growth estimate is satisfied for Riemannian manifolds with a lower bound on the Ricci curvature and for the Heisenberg groups as well.

The last remark of this paper concerns possible generalizations of our proof.

Remark 5.3. It is a natural question to ask whether the techniques of this paper apply to more general subRiemannian structures. We would like to review some of the key-points that seem to be necessary to make our proof of paradigm (1) work and try to define a possible framework. A first natural assumption is to replace \mathbb{H}_n by a Carnot group \mathbb{G} . These groups are isomorphic to vector spaces of finite dimension with a polynomial multiplicative structure. The Lebesgue measure is left- and right- invariant so that one has $\int fV^2g \, d\mathcal{L} + \int Vf.Vg \, d\mathcal{L} = 0$ for any left-invariant vector field V and compactly supported and smooth f and g (see [22] on Carnot groups). As I’m not aware of a Carnot group where estimates concerning the diffusion are as well known as in the Heisenberg group, it is reasonable to set up in a compact space. Actually in a non-compact setting we would need estimates such as those in Lemma 1.3 for the direct part of the proof corresponding to paragraph

4.1. Hence we choose to work in a nilmanifold, that is the quotient manifold of \mathbb{G} by a cocompact subgroup.

In the indirect part of the proof, it will be necessary to know a lower bound of the Ricci curvature for the approximating manifolds. This is necessary for establishing an HWI inequality as in Proposition 3.4 at the measures μ that have a density ρ with a square-integrable Riemannian gradient. The second important item is the possibility to replace a gradient flow curve $(\mu_t)_{t \in I}$ by a curve made of more regular measures μ_t^* obtained doing a convolution in the direction of the center of \mathbb{G} (the commutativity would unblock the continuity of both the entropy and the subRiemannian Fisher information during the 1-parameter convolution). This operation aims to give the density ρ_t^* a derivative in every direction, in other words a Riemannian gradient. A consequence is that the operator $\Delta_{\mathbb{G}}$ should be defined as $\sum_i V_i^2$ where the left invariant vector fields V_i are linearly independent and span the whole Lie algebra except the center (that can be more than 1-dimensional). This is a sub-Laplacian if \mathbb{G} is a 2-step Carnot group and another hypoelliptic operator otherwise.

Acknowledgements. I wish to thank Hervé Pajot and Karl-Theodor Sturm that were my advisers when I started this research. I thank Luigi Ambrosio, Nicola Gigli and Cédric Villani for helpful discussions as well as my ex-colleagues Matthias Erbar and Miguel Rodrigues. The referee of the first version gave me good advice (for instance on alternative proofs in Proposition 3.1 and paragraph 4.1) and the motivation for avoiding a non-necessary extra hypothesis in paragraph 4.2. I would like to thank her or him very much. I thank also the other referees for their careful reading

REFERENCES

- [1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.
- [2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
- [3] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *preprint*, 2011.
- [4] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *preprint*, 2011.
- [5] L. Ambrosio and S. Rigot. Optimal mass transportation in the Heisenberg group. *J. Funct. Anal.*, 208(2):261–301, 2004.
- [6] L. Ambrosio and G. Savaré. Gradient flows of probability measures. In *Handbook of differential equations: evolutionary equations. Vol. III*, Handb. Differ. Equ., pages 1–136. Elsevier/North-Holland, Amsterdam, 2007.
- [7] P. Bernard. Young measures, superposition and transport. *Indiana Univ. Math. J.*, 57(1):247–275, 2008.
- [8] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Information Theory*, IT-11:267–271, 1965.
- [9] L. Capogna, D. Danielli, S. D. Pauls, and J. T. Tyson. *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, volume 259 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 2007.
- [10] J. Cheeger and B. Kleiner. Differentiating maps into L^1 , and the geometry of BV functions. *Ann. of Math. (2)*, 171(2):1347–1385, 2010.
- [11] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):1–23, 2010.

- [12] S. Fang, J. Shao, and K.-T. Sturm. Wasserstein space over the Wiener space. *Probab. Theory Related Fields*, 146:535–565, 2010.
- [13] A. Fathi and A. Figalli. Optimal transportation on non-compact manifolds. *Israel J. Math.*, 175:1–59, 2010.
- [14] A. Figalli and N. Juillet. Absolute continuity of Wasserstein geodesics in the Heisenberg group. *J. Funct. Anal.*, 255(1):133–141, 2008.
- [15] A. Figalli and L. Rifford. Mass transportation on sub-Riemannian manifolds. *Geom. Funct. Anal.*, 20(1):124–159, 2010.
- [16] G. B. Folland and E. M. Stein. *Hardy spaces on homogeneous groups*, volume 28 of *Mathematical Notes*. Princeton University Press, Princeton, N.J., 1982.
- [17] B. Franchi, R. Serapioni, and F. Serra Cassano. Rectifiability and perimeter in the Heisenberg group. *Math. Ann.*, 321(3):479–531, 2001.
- [18] B. Gaveau. Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents. *Acta Math.*, 139(1-2):95–153, 1977.
- [19] N. Gigli. On the heat flow on metric measure spaces: existence, uniqueness and stability. *Calc. Var. Partial Differential Equations*, 39(1-2):101–120, 2010.
- [20] N. Gigli, K. Kuwada, and S.-i. Ohta. Heat Flow on Alexandrov spaces. *Accepted at Comm. Pure Appl. Math.*, 2012.
- [21] N. Gigli and S.-I. Ohta. First Variation Formula in Wasserstein Spaces over Compact Alexandrov Spaces. *Canad. Math. Bull.*, 55(4):723–735, 2012.
- [22] J. Heinonen. Calculus on Carnot groups. In *Fall School in Analysis (Jyväskylä, 1994)*, volume 68 of *Report*, pages 1–31. Univ. Jyväskylä, Jyväskylä, 1995.
- [23] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.
- [24] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
- [25] N. Juillet. *Optimal transport and geometric analysis in Heisenberg groups*. PhD thesis, Université Grenoble 1, Bonn Universität, <http://tel.archives-ouvertes.fr/tel-00345301>, 2008.
- [26] N. Juillet. Geometric inequalities and generalized Ricci bounds in the Heisenberg group. *International Mathematical Research Notices*, 2009, 2009.
- [27] B. Khesin and P. Lee. A nonholonomic Moser theorem and optimal transport. *J. Symplectic Geom.*, 7(4):381–414, 2009.
- [28] A. Korányi and H. M. Reimann. Quasiconformal mappings on the Heisenberg group. *Invent. Math.*, 80(2):309–338, 1985.
- [29] H.-Q. Li. Estimation optimale du gradient du semi-groupe de la chaleur sur le groupe de Heisenberg. *J. Funct. Anal.*, 236(2):369–394, 2006.
- [30] H.-Q. Li. Estimations asymptotiques du noyau de la chaleur sur les groupes de Heisenberg. *C. R. Math. Acad. Sci. Paris*, 344(8):497–502, 2007.
- [31] J. Lott and C. Villani. Weak curvature conditions and functional inequalities. *J. Funct. Anal.*, 245(1):311–333, 2007.
- [32] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.
- [33] J. Maas. Gradient flows of the entropy for finite markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
- [34] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001.
- [35] R. Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.
- [36] S.-i. Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *Amer. J. Math.*, 131(2):475–516, 2009.
- [37] S.-i. Ohta and K.-T. Sturm. Heat flow on Finsler manifolds. *Comm. Pure Appl. Math.*, 62(10):1386–1433, 2009.
- [38] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [39] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [40] A. Petrunin. Alexandrov meets Lott-Villani-Sturm. *Münster J. Math.*, 4:53–64, 2011.

- [41] L. P. Rothschild and E. M. Stein. Hypoelliptic differential operators and nilpotent groups. *Acta Math.*, 137(3-4):247–320, 1976.
- [42] G. Savaré. Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds. *C. R. Math. Acad. Sci. Paris*, 345(3):151–154, 2007.
- [43] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.
- [44] E. M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.
- [45] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.
- [46] K.-T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.
- [47] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, 2009.
- [48] H.-C. Zhang and X.-P. Zhu. Ricci curvature on Alexandrov spaces and rigidity theorems. *Comm. Anal. Geom.*, 18(3):503–553, 2010.

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE, UMR 7501, UNIVERSITÉ DE STRASBOURG
ET CNRS, 7 RUE RENÉ DESCARTES, 67 000 STRASBOURG, FRANCE
E-mail address: nicolas.juillet@math.unistra.fr