Horocyclic Brunn-Minkowski inequality

Rotem Assouline and Bo'az Klartag

Abstract

Given two non-empty subsets A and B of the hyperbolic plane $\mathbb{H}^2$, we define their horocyclic Minkowski sum with parameter $\lambda = 1/2$ as the set $[A : B]_{1/2} \subseteq \mathbb{H}^2$ of all midpoints of horocycle curves connecting a point in A with a point in B. These horocycle curves are parameterized by hyperbolic arclength, and the horocyclic Minkowski sum with parameter $0 < \lambda < 1$ is defined analogously. We prove that when A and B are Borel-measurable,

$$\sqrt{Area([A : B]_\lambda)} \geq (1-\lambda) \cdot \sqrt{Area(A)} + \lambda \cdot \sqrt{Area(B)},$$

where $Area$ stands for hyperbolic area, with equality when A and B are concentric discs in the hyperbolic plane. We also prove horocyclic versions of the Prékopa-Leindler and Borell-Brascamp-Lieb inequalities. These inequalities slightly deviate from the metric measure space paradigm on curvature and Brunn-Minkowski type inequalities, where the structure of a metric space is imposed on the manifold, and the relevant curves are necessarily geodesics parameterized by arclength.

1 Introduction

The Brunn-Minkowski inequality is a geometric inequality that was discovered circa 1887, see Schneider [21, Section 7.1] for its early history. In one of its formulations, it states that for any non-empty, Borel measurable subsets $A, B \subseteq \mathbb{R}^n$ and $0 < \lambda < 1$,

$$Vol_n((1-\lambda)A + \lambda B)^{1/n} \geq (1-\lambda) \cdot Vol_n(A)^{1/n} + \lambda \cdot Vol_n(B)^{1/n}. \tag{1}$$

Here Vol_n is n-dimensional volume in $\mathbb{R}^n$, we write $A + B = \{x + y\, ;\, x \in A, y \in B\}$ for the Minkowski sum, and $\lambda \cdot A = \{\lambda x\, ;\, x \in A\}$. It is perhaps more common to formulate the Brunn-Minkowski inequality as stating that

$$Vol_n(A + B)^{1/n} \geq Vol_n(A)^{1/n} + Vol_n(B)^{1/n}, \tag{2}$$

which is easily seen to be equivalent to (1) by scaling. The advantage of the formulation (1) is that the set $(1-\lambda)A + \lambda B$ is well-defined for any two sets A and B in an *affine* space, while the definition of the set $A + B$ requires the structure of a *linear* space, i.e., it requires a marked point in space referred to as the origin. The Brunn-Minkowski inequality is an indispensible tool in convex geometry, for instance it immediately implies the isoperimetric inequality, see Burago and Zalgaller [4], Gardner [14] or Schneider [21] for more information.

The Brunn-Minkowski inequality has been generalized to other geometries, and in particular to Riemannian manifolds. Given two subsets A and B of a complete, connected, n-dimensional, Riemannian manifold M, we may look at the set of all *midpoints* of minimizing geodesics connecting a point in A with a point in B. Observe that in the case where $M = \mathbb{R}^n$, the set obtained is precisely $(A + B)/2$. Similarly, we can make sense of $(1 - \lambda)A + \lambda B$ for every $0 < \lambda < 1$. Provided that the Ricci curvature of M is non-negative, it is known that the Brunn-Minkowski inequality (1) holds true in M. See Sturm [23] for a discussion of the Riemannian version of (1), and see Cordero-Erausquin, McCann and Schmuckenschläger [8] for a proof of the Prékopa-Leindler inequality, which is a stronger, functional version of the Brunn-Minkowski inequality, for Riemannian manifolds with Ricci curvature bounded below.

In negatively curved manifolds, the Brunn-Minkowski inequality may fail dramatically. Let us demonstrate the issue by looking at the example of the hyperbolic plane $\mathbb{H}^2$, which is the complete, simply-connected, two-dimensional Riemannian manifold of constant Gaussian curvature -1. Consider two discs A and B of area one in $\mathbb{H}^2$ that are far away from each other. It is not too difficult to show that the area of the set of midpoints

$$\left\{ \gamma\left(\frac{1}{2}\right) \, ; \, \gamma(0) \in A, \ \gamma(1) \in B, \text{and } \gamma : [0,1] \to \mathbb{H}^2 \text{ is a minimizing, constant speed geodesic} \right\}$$

can be arbitrarily small. The existence of "tiny bottlenecks" through which all geodesics from A to B have to pass, roughly at their midpoint, is a fundamental property of negatively curved surfaces.

In this paper we propose to remedy the failure of the Brunn-Minkowski inequality in the hyperbolic plane by looking at horocycles in place of geodesics. The idea is that perhaps the geodesic curvature of horocycles can compensate for the negative Gaussian curvature of the underlying hyperbolic space. This is loosely inspired by the success of the Bakry-Émery approach to Riemannian manifolds equipped with a measure, where a negative Ricci tensor can be compensated by the Hessian of the logarithm of the density. See e.g. Bakry, Gentil and Ledoux [1] for information on the Bakry-Émery theory.

A smooth curve $\gamma : [a, b] \to \mathbb{H}^2$ is a *horocycle* if it has constant speed, and if its geodesic curvature equals 1. For instance, consider the Poincaré disc model of hyperbolic geometry, in which the hyperbolic plane is represented in the unit disc $D = \{z \in \mathbb{C} \, ; \, |z| < 1\}$ via the Riemannian metric tensor

$$g = \frac{4|dz|^2}{(1 - |z|^2)^2} \qquad (z \in D).$$

In the Poincaré disc model, each horocycle is a Euclidean circle in the unit disc D tangent to its boundary, i.e., a curve of the form

$$\{z \in D \, ; \, |z - a| = 1 - |a|\}$$

Figure 1: Horocyclic vs. geodesic Minkowski summation in the upper half plane model. The horocyclic $\frac{1}{2}$-Minkowski sum of the two discs appears in blue. The ordinary (geodesic) $\frac{1}{2}$-Minkowski sum, i.e. the set of all midpoints of geodesics joining the two discs, appears in green.

for some $0 \neq a \in D$. We emphasize that our horocycles are constant speed curves with respect to the hyperbolic metric, since this is the only reasonable intrinsic way to parametrize these curves. We refer the reader to Izumiya [15] for a survey of horocyclical geometry, discussing in particular its resemblance to planar Euclidean geometry. An alternative viewpoint on horocycles, is that these are limits of hyperbolic circles passing through a fixed point, whose radii tend to infinity.

Any pair of points in the hyperbolic plane is joined by *two* horocycles, and hence every two points have two different "horocyclic midpoints". It is convenient to restrict our attention to *oriented horocycles*: those horocycles whose velocity and acceleration vectors form an oriented orthogonal basis, after some choice of orientation has been made. In the Poincaré disc model, for example, we can restrict our attention to horocycles parametrized in the counterclockwise direction. Thus, for any two distinct points $x, y \in \mathbb{H}^2$ there is a unique oriented horocycle going from x to y, which is different from the unique oriented horocycle going from y to x.

For $A, B \subseteq \mathbb{H}^2$ and for $0 < \lambda < 1$ we define the *λ-horocycle Minkowski sum of A and B* as

$$[A : B]_\lambda = \{\gamma(\lambda)\,;\, \gamma(0) \in A, \gamma(1) \in B, \text{ and } \gamma \text{ is a constant-speed oriented horocycle}\}\,. \quad (3)$$

By a disc $D(a, r)$ in $\mathbb{H}^2$ we mean the set of all points whose distance from a fixed point $a \in \mathbb{H}^2$ is at most r. When A and B are two concentric discs in $\mathbb{H}^2$, the set $[A : B]_\lambda$ is again a disc. However, when the discs are not concentric, the λ-horocycle Minkowski sum is not necessarily a disc. See Figure 1.

Our main result is a Brunn-Minkowski theorem for the horocyclic Minkowski sum, which resembles the planar Euclidean Brunn-Minkowski inequality.

Theorem 1.1. *Let $A, B \subseteq \mathbb{H}^2$ be non-empty, Borel measurable sets. Then for any $0 < \lambda < 1$,*

$$Area([A : B]_\lambda)^{1/2} \geq (1 - \lambda) \cdot Area(A)^{1/2} + \lambda \cdot Area(B)^{1/2} \tag{4}$$

where $Area$ stands for hyperbolic area in $\mathbb{H}^2$. When A and B are concentric discs, equality holds in (4).

We also prove a functional version of Theorem 1.1, in spirit of the Prékopa-Leindler and the Borell-Brascamp-Lieb inequalities, see Theorem 6.1 below.

Remark 1.2. Theorem 1.1 remains correct if one modifies definition (3) of the λ-horocyclic Minkowski sum, replacing "an oriented horocycle" by "a horocycle". When $A, B \subseteq \mathbb{H}^2$ are concentric discs, equality holds in (4) also with respect to this modified definition.

Theorem 1.1 is analogous to the Brunn-Minkowski inequality in its formulation (1) above. There is also a horocyclic analogue of (2), see Remark 6.5. We still do not know how to fully characterize the equality cases of (4). We also do not consider here the case of *equidistants* in the hyperbolic plane in place of horocycles, i.e., curves whose geodesic curvature equals $\kappa \in (0, 1)$. The higher dimensional case is briefly discussed in Remark 5.2 and Remark 6.8 below.

The approach that we present in this paper is that of a Brunn-Minkowski inequality with respect to a path space, where the paths are not necessarily geodesics. This generalizes the metric-measure space viewpoint on curvature and Brunn-Minkowski type inequalities, where the structure of a metric space is imposed on the manifold, and the relevant curves must be geodesics parameterized by arclength, see Sturm [22, 23] and Lott and Villani [17]. Another interesting development in this direction is the *Lorentzian* approach of McCann [19] and Cavalleti and Mondino [6]. We hope that our approach can be extended to include also other examples of paths of interest, beyond horocycles in the hyperbolic plane. See Theorem 6.7 below for a result in this spirit.

The proof of Theorem 1.1 utilizes L^1-mass transport, or the *needle decomposition* approach for proving geometric inequalities on manifolds that was suggested and developed in the Riemannian setting in [16]. This approach was extended to the setting of a metric-measure space by Cavalleti and Mondino [5] and to the non-reversible Finsler setting by Ohta [20]. A curious aspect of our proof is that we need to work with a certain Finsler metric in order to construct the desired needle decomposition. We learned about this Finsler metric from the work of Crampin and Mestdag [9].

The rest of this paper is organized as follows. Section 2, Section 3 and Section 4 contain a rather extensive introduction to the theory of L^1-mass transport in the two-dimensional case, where the regularity analysis is a simpler. Horocycle geometry is discussed in Section 5 and Section 6, where our Brunn-Minkowski type inequality is proven, as well as its functional versions.

Acknowledgement. Supported by a grant from the Israel Science Foundation (ISF). The second named author would like to thank Itai Benjamini for a fruitful discussion that has led us to conjecture the horocyclic Brunn-Minkowski inequality and for his encouragement along the way.

2 Preliminaries

Let M be an n-dimensional smooth manifold. We write T_xM for the tangent space at the point $x \in M$, and $TM = \cup_{x\in M} T_xM$ is the tangent bundle. A *Finsler structure* on M is a function $\Phi : TM \to [0,\infty)$, smooth away from the zero section, which satisfies the following requirements:

- *positively homogeneous*: $\Phi(\lambda v) = \lambda\Phi(v)$ for all $v \in TM$ and $\lambda > 0$.
- *strongly convex*: Fix $x \in M$. Then the function Φ^2 is convex in the linear space T_xM, and moreover its Hessian at any point $0 \neq v \in T_xM$ is positive definite.

We refer the reader to Bao, Chern and Shen [2] for background on Finsler structures and for proofs of the basic facts mentioned in this section. A Finsler structure induces a metric on M by setting

$$d(x,y) = d_M(x,y) = \inf_{\gamma} \mathrm{Length}(\gamma),$$

where

$$\mathrm{Length}(\gamma) := \int_0^1 \Phi(\dot\gamma(t))dt,$$

and where the infimum is taken over C^1 curves joining x to y. Since the function Φ is only positively homogeneous, $\Phi(-v)$ is not necessarily equal to $\Phi(v)$. As a result, the distance function may not be symmetric. In fact, the metric d is symmetric if and only if $\Phi(v) = \Phi(-v)$ for all $v \in TM$, in which case (M,Φ) is said to be *reversible*.

A *curve* on M is a smooth function from an interval to M. Our notation does not fully distinguish between a parameterized curve $\gamma : [a,b] \to M$ and its image $\gamma([a,b])$ which is just a subset of M, sometimes endowed with an orientation. A curve has constant speed if $\Phi(\dot\gamma(t))$ is constant in t, and it has unit speed if $\Phi(\dot\gamma(t)) = 1$ for all t.

A *minimizing geodesic* or a *forward-minimizing geodesic* is a constant-speed curve $\gamma : [a,b] \to M$ satisfying $\mathrm{Length}(\gamma) = d(\gamma(a),\gamma(b))$. A *geodesic* is a curve $\gamma : [a,b] \to M$ that is locally a minimizing geodesic, i.e., for any $t_0 \in [a,b]$ there exists $\delta > 0$ such that the restriction of γ to the interval $[a,b] \cap (t_0-\delta, t_0+\delta)$ is a minimizing geodesic. Equivalently, a geodesic is a solution to the Euler-Lagrange equation associated with the Lagrangian $\Phi^2/2$. We note that unless Φ is reversible, the curve $\gamma(-t)$ does not have to be a geodesic.

We recall that a minimizing geodesic η cannot intersect a minimizing geodesic $\gamma : (a,b) \to M$ at more than one point unless they overlap.

We say that (M,Φ) is *geodesically convex* if every $x,y \in M$ are joined by a minimizing geodesic. We say that (M,Φ) is *strongly convex* if every $x,y \in M$ are joined by a *unique* minimizing geodesic. Any point in M has a neighborhood that is strongly convex. Any complete Finsler manifold is automatically geodesically convex. We do not require below that (M, Φ) be complete.

A Borel measure on a smooth manifold M is said to be *absolutely continuous* if it is absolutely continuous in local coordinates. By the change of variables formula, the property of a Borel measure being *absolutely continuous with smooth density* does not depend on the choice of local coordinates. Similarly, we say that $A \subseteq M$ is a set of measure zero if in any local chart it has measure zero. When we say that almost any point in M satisfies a certain condition, we mean that the set of points that do not satisfy this condition is of measure zero.

Given $x \in M$ and a covector $\alpha \in T_x^* M$, the fiberwise strong convexity of $\Phi^2/2$ implies that the supremum

$$\sup_{v \in T_x M} \left[\alpha(v) - \frac{\Phi^2(v)}{2} \right] \tag{5}$$

is uniquely attained at a tangent vector $\mathcal{L}(\alpha) \in T_x M$, depending smoothly on α. The resulting diffeomorphism $\mathcal{L} : T^* M \to TM$ is called the *Legendre transform* associated with the function $\Phi^2/2$. (In convexity theory it is common to refer to the expression in (5) as the Legendre transform, and not to the unique maximizer, but here we try to stick to standard terminology in Finsler geometry).

Let $f : M \to \mathbb{R}$ be a smooth function. The Legendre transform $\mathcal{L}(df)$ of the differential of f is referred to as the *gradient* of f and denoted by ∇f. Note that the gradient $\nabla f(x)$ points at the direction of steepest infinitesimal ascent of the function f.

Suppose that M and N are smooth manifolds, $A \subseteq M$ an arbitrary set and $f : A \to N$. We say that f is *locally Lipschitz* if for every $p \in A$ there exist open neighbourhoods $p \in U \subseteq M$ and $f(p) \in V \subseteq N$, each contained in a coordinate chart, such that f is Lipschitz in these local coordinates. We write $| \cdot |$ for the Euclidean norm in $\mathbb{R}^n$.

3 Partition with a guiding function

Let (M, Φ) be a geodesically convex Finsler manifold. A function $u : M \to \mathbb{R}$ is said to be L-Lipschitz if for all $x, y \in M$,

$$-L \cdot d(y, x) \leq u(y) - u(x) \leq L \cdot d(x, y).$$

(note that the left-hand side inequality follows from the right-hand one). The minimal $L \geq 0$ for which u is L-Lipschitz is denoted by $\|u\|_{Lip}$. Let μ_1, μ_2 be two absolutely-continuous, finite Borel measures on M satisfying the mass balance condition

$$\mu_1(M) = \mu_2(M).$$

Let us furthermore make the mild regularity assumption that there exists $x_0 \in M$ such that

$$\int_M [d(x_0, x) + d(x, x_0)] d\mu_i(x) < \infty \qquad (i = 1, 2).$$

Consider the Monge-Kantorovich optimization problem

$$W_1(\mu_1, \mu_2) = \sup_{\|u\|_{Lip} \le 1} \left[\int_M u d\mu_1 - \int_M u d\mu_2 \right]. \tag{6}$$

It follows from the Arzela-Ascoli theorem that the supremum in (6) is finite and is actually a maximum. Let us fix a 1-Lipschitz function u that attains the supremum in (6), and refer to this function as the *guiding function* or the *Kantorovich potential*. As in Ohta [20, Section 4.1] we define

$$\Gamma_u = \{(x, y) \in M \times M \, ; \, u(y) - u(x) = d(x, y)\} .$$

The collection of *strain points* of u is

$$Strain[u] = \{x \in M \, ; \, \exists w, y \in M \setminus \{x\}, \;\; (w, x), (x, y) \in \Gamma_u\} ,$$

and the collection of *loose points* is

$$Loose[u] = \{x \in M \, ; \, \forall y \in M \setminus \{x\}, \;\; (x, y), (y, x) \not\in \Gamma_u\} .$$

Clearly $Strain[u]$ and $Loose[u]$ are disjoint sets. A *transport ray* of u is a minimizing, unit-speed geodesic $\gamma : I \to M$, with $I \subseteq \mathbb{R}$ connected, open and non-empty, such that for all $s, t \in I$,

$$u(\gamma(t)) - u(\gamma(s)) = t - s, \tag{7}$$

and such that γ is maximal: there is no minimizing, unit-speed geodesic $\tilde{\gamma} : J \to M$, with $J \subseteq \mathbb{R}$ connected, open and strictly containing I, such that $\tilde{\gamma}|_I = \gamma$ and (7) holds true for all $s, t \in J$.

Proposition 3.1 (Properties of the guiding function)**.**

(i) Almost any point in M belongs either to $Strain[u]$ or to $Loose[u]$. Moreover, almost all points in the support of $\mu_1 - \mu_2$ belong to $Strain[u]$.

(ii) The function u is differentiable at any point $x \in Strain[u]$, and its Finsler gradient $\nabla u(x) \in T_x M$ is a unit vector.

(iii) The set $Strain[u]$ is the disjoint union of all transport rays of u. Furthermore, u is differentiable in $Strain[u]$, and each transport ray $\gamma : I \to M$ is an integral curve of ∇u, i.e.

$$\nabla u(\gamma(t)) = \dot{\gamma}(t) \qquad \textit{for all } t \in I. \tag{8}$$

(iv) If $(x, y) \in \Gamma_u$ for $x \neq y$, then the relative interior of any forward-minimizing geodesic from x to y is contained in a transport ray.

(v) A union of transport rays, which is also a Borel subset of M, is called a transport set. *For any transport set $A \subseteq M$ we have the mass balance condition*

$$\mu_1(A) = \mu_2(A).$$

Proposition 3.1 is well-known, and the main ideas behind it go back to Evans and Gangbo [11], with further contributions by Feldman and McCann [13], the second named author [16], Cavalleti and Mondino [5] and in particular Ohta [20] who considered the non-reversible, Finslerian case. The proofs of (ii), (iii) and (iv) are short and simple, and are presented with all details in [20, Section 4.1]. There is no need to repeat them here. The proofs of (i) and (v) require some preparation. The two-dimensional case, which is rather central to us, is somewhat simpler, as remarked already by Evans and Gangbo [11]. Thus from now on we assume in this section that

$$n = 2.$$

We hope that our relatively short presentation in two dimensions is of value for some readers. In the remainder of this section we prove (i) and (v), as well as some regularity properties of u and of the foliation by transport rays.

By Proposition 3.1(iii), every point $x \in Strain[u]$ is contained in a unique transport ray

$$\gamma : (-\alpha(x), \beta(x)) \to M$$

satisfying

$$\dot{\gamma}(0) = \nabla u(x).$$

By the definition of $Strain[u]$, both α and β are well-defined, positive functions on $Strain[u]$, possibly attaining the value $+\infty$. These two functions are Borel mesurable, essentially by the same arument as in [16, Lemma 2.1.12]. Let

$$Strain_\varepsilon[u] := \{x \in Strain[u]\, ;\, \alpha(x), \beta(x) \geq \varepsilon\}.$$

Thus

$$Strain[u] = \bigcup_{\varepsilon>0} Strain_\varepsilon[u].$$

Lemma 3.2. *For any $\varepsilon > 0$, the gradient ∇u is locally-Lipschitz in $Strain_\varepsilon[u]$.*

The proof of Lemma 3.2 requires the following two-dimensional lemma, whose proof is deferred to the Appendix below.

Lemma 3.3 ("Disjoint geodesics that are close to each other at one point have similar tangents")**.** *Let $p \in M$. Then there exist $C, c > 0, \sigma_0 \in (0,1)$ and a smooth coordinate chart $\psi : V \to U$, where V is an open subset of $\mathbb{R}^2$ and U is a neighbourhood of p, such that the following holds.*

Suppose that $0 < \sigma \leq \sigma_0$ and $\gamma_1, \gamma_2 : (-\sigma, \sigma) \to U$ are two disjoint, unit-speed geodesics. Set

$$\eta_i = \psi^{-1} \circ \gamma_i,$$

and assume that $|\dot{\eta}_1(0) - \dot{\eta}_2(0)| < c$. Then,

$$|\dot{\eta}_1(0) - \dot{\eta}_2(0)| \leq \frac{C}{\sigma} \cdot |\eta_1(0) - \eta_2(0)|.$$

Proof of Lemma 3.2. First, we argue that ∇u is continuous on $Strain_\varepsilon[u]$. Let $p_n \in Strain_\varepsilon[u]$ converge to $p \in Strain_\varepsilon[u]$. Then there exist unit-speed, minimizing geodesics $\gamma_n, \gamma : (-\varepsilon, \varepsilon) \to M$ with

$$\gamma_n(0) = p_n, \quad \gamma(0) = p, \quad \dot\gamma_n(0) = \nabla u(p_n) \quad \text{and} \quad \dot\gamma(0) = \nabla u(p),$$

all satisfying (7) for all $t, s \in (-\varepsilon, \varepsilon)$. Assume by contradiction that $\dot\gamma_n(0)$ has a subsequence $\dot\gamma_{n_k}(0)$ converging to some $v \neq \nabla u(p)$. By the continuity of u and of the Finslerian distance function d, the pointwise limit of γ_{n_k} is a unit-speed, minimizing geodesic γ_0 satisfying $\gamma_0(0) = p$, $\dot\gamma_0(0) = v$ and

$$u(\gamma_0(t)) - u(\gamma_0(s)) = t - s \qquad \text{for all } t, s \in (-\varepsilon, \varepsilon).$$

Since u is 1-Lipschitz, it follows that $\nabla u(p) = v$, in contradiction. This completes the proof that ∇u is continuous. Next, let $p \in Strain_\varepsilon[u]$. Let the coordinate chart $\psi : V \to U$ and the constants $C, c, \sigma_0 > 0$ be as in Lemma 3.3. For ease of reading and with abuse of notation, in this proof we identify $U \subseteq M$ with $V \subseteq \mathbb{R}^2$, and we identify TU with $TV \cong V \times \mathbb{R}^2$. Since ∇u is continuous on $Strain_\varepsilon[u]$, we may shrink the coordinate chart and assume that

$$|\nabla u(q_1) - \nabla u(q_2)| \leq c \qquad \text{for all } q_1, q_2 \in U \cap Strain_\varepsilon[u]. \tag{9}$$

Choose a neighbourhood $U' \subset U$ of p and a number $\sigma_1 > 0$ such that any geodesic $\gamma : (-\sigma_1, \sigma_1) \to M$ with $\gamma(0) \in U'$ is contained in U. Set

$$\sigma = \min\{\sigma_0, \sigma_1, \varepsilon\}.$$

For any $q \in U' \cap Strain_\varepsilon[u]$, there exists a unit-speed geodesic $\gamma_q : (-\sigma, \sigma) \to U'$, contained in the transport ray through q, with $\gamma_q(0) = q$ and

$$\dot\gamma_q(0) = \nabla u(q).$$

Let $q_1, q_2 \in U' \cap Strain_\varepsilon[u]$. If q_1, q_2 belong to different transport rays, then γ_{q_i} are disjoint by Proposition 3.1(iii), hence by (9) and Lemma 3.3,

$$|\nabla u(q_1) - \nabla u(q_2)| \leq \frac{C}{\sigma}|q_1 - q_2|.$$

If q_1, q_2 lie on the same transport ray then $\nabla u(q_1)$ and $\nabla u(q_2)$ are unit vectors tangent to the same geodesic, and by the smoothness of geodesics, they differ by at most $C'|q_1 - q_2|$, where C' is independent of q_1, q_2. It follows thus that ∇u is locally Lipschitz on $U' \cap Strain_\varepsilon[u]$. □

The set $Strain[u]$, which comprises of disjoint transport rays, can be divided into countably many "ray clusters", each consisting of a union of transport rays, and admitting a convenient parametrization.

Lemma 3.4. *We may decompose $Strain[u]$ into a countable disjoint union of transport sets $A_1, A_2, \ldots$ such that for any $k \geq 1$,* *(10) There exists a Borel subset $Y_k \subseteq \mathbb{R}$ and a locally Lipschitz one-to-one map $\varphi_k : Y_k \to A_k$ such that for each transport ray $\gamma \subseteq A_k$, the set $\varphi_k(Y_k) \cap \gamma$ is a singleton. Moreover, ∇u is locally-Lipschitz in $\varphi_k(Y_k)$.*

Proof. Let $p \in Strain[u]$. Then $p \in Strain_\varepsilon[u]$ for some $\varepsilon > 0$. Let $v \in T_pM$, $v \neq \nabla u(p)$ and let $\eta : (-s_0, s_0) \to M$ be a minimizing geodesic with $\dot{\eta}(0) = v$, for some $s_0 > 0$. Thus η is transverse to the transport ray through p. Since a transport ray is a minimizing geodesic and since η is a minimizing geodesic defined on an open interval, no transport ray can intersect η more than once.

Let $A_{p,\varepsilon}$ denote the union of all transport rays intersecting $\eta \cap Strain_\varepsilon[u]$. Then $A_{p,\varepsilon}$ is Borel (even sigma-compact, as in [16, Lemma 2.1.12]) hence is a transport set. We claim that there exists an open disc D containing p such that $D \cap Strain_\varepsilon[u] \subseteq A_{p,\varepsilon}$. Indeed, otherwise there exists a sequence $p_k \in Strain_\varepsilon[u]$ converging to p such that the transport rays through p_k do not intersect η, but the transport ray through p is transverse to η, which is a contradiction to the continuity of ∇u on $Strain_\varepsilon[u]$ (since we are in dimension two).

Let $Y_{p,\varepsilon} = \eta^{-1}(Strain_\varepsilon[u])$, and $\varphi_{p,\varepsilon} = \eta|_{Y_{p,\varepsilon}}$. Then $Y_{p,\varepsilon}$ is Borel and $\varphi_{p,\varepsilon}$ is locally Lipschitz and one-to-one, and ∇u is locally Lipschitz on $\varphi_{p,\varepsilon}(Y_{p,\varepsilon})$ by Lemma 3.2. By construction, $\varphi_{p,\varepsilon}(Y_{p,\varepsilon}) \cap \gamma$ is a singleton for every transport ray $\gamma \subseteq A_{p,\varepsilon}$.

Since each $A_{p,\varepsilon}$ contains a neighbourhood of p in $Strain_\varepsilon[u]$, it is possible to take a countable subcollection $\{(A_k, Y_k, \varphi_k)\}$ of the collection $\{(A_{p,\varepsilon}, Y_{p,\varepsilon}, \varphi_{p,\varepsilon})\}$ such that the union of the transport sets A_k covers $Strain[u]$. We may also take A_k to be disjoint, by replacing A_k by $A_k \setminus \bigcap_{k'<k} A_{k'}$. This does not change the property that $\varphi(Y_k) \cap \gamma$ is a singleton for every transport ray γ in A_k. ∎

For future reference we consider the notion of a *parallel line-cluster* for general n, even though now we are only interested in the case $n = 2$. A *parallel line-cluster* is a subset $B \subseteq \mathbb{R}^{n-1} \times \mathbb{R}$ of the following form: there exists a Borel set $Y \subseteq \mathbb{R}^{n-1}$ and Borel functions $a : Y \to [-\infty, 0)$ and $b : Y \to (0, +\infty]$ such that

$$B = \left\{(y, t) \in \mathbb{R}^{n-1} \times \mathbb{R}\,;\, y \in Y,\ a_y < t < b_y\right\}. \tag{11}$$

Corollary 3.5 ("Local straightening of the partition")**.** *For each of the transport sets A_k from Lemma 3.4 there is a parallel line-cluster B_k and a locally-Lipschitz bijection $F_k : B_k \to A_k$ such that the following hold:*

(a) The relation (11) *holds true with $Y = Y_k$, where Y_k and $\varphi = \varphi_k$ satisfy (10). Additionally, $F_k(y, 0) = \varphi_k(y)$ for all $y \in Y_k$.*

(b) For any $y \in Y_k$, the curve $t \mapsto F_k(y, t)$ defined for $a_y < t < b_y$ is a transport ray.

(c) Write $E_k \subseteq M$ for the set of all limit points of the form $\lim_{t \to a_y^+} F_k(y, t)$ or $\lim_{t \to b_y^-} F_k(y, t)$, whenever the limit exists, where y ranges over the set Y_k. Then E_k has measure zero.

Proof. Fix $k \geq 1$ and let $A = A_k, Y = Y_k, \varphi = \varphi_k$ be as in Lemma 3.4. For each $y \in Y$ let γ_y be the transport ray satisfying $\dot{\gamma}_y(0) = \nabla u(\varphi(y))$. Let

$$a_y = -\alpha(\varphi(y)), \qquad b_y = \beta(\varphi(y))$$

and

$$F(y,t) = \gamma_y(t), \qquad y \in Y, t \in (a_y, b_y).$$

The transport ray γ_y is a unit-speed geodesic, and in the case where $\lim_{t \to a_y^+} \gamma_y(t)$ exists we define $F(y, a_y)$ by continuity. We similarly define $F(y, b_y)$ by continuity whenever $\lim_{t \to b_y^-} \gamma_y(t)$ exists.

By Lemma 3.4, the function $\nabla u \circ \varphi$ is locally Lipschitz on Y, whence the function F is locally Lipschitz on the set B defined in (11). The functions a and b are Borel. The locally-Lipschitz function F is a bijection on B by Lemma 3.4, and properties (a) and (b) hold with $F_k = F$ and $B = B_k$ by construction. The set E_k is the image under the map F of the set

$$\left\{ (y,t) \in B \,\middle|\, t \in \{a_y, b_y\},\ F_y(t) \text{ is defined} \right\} \subseteq \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R} \times \mathbb{R},$$

which is contained in the union of two graphs of measurable functions (of one variable, as $n = 2$). Since F is locally-Lipschitz, we conclude that E_k has measure zero (see [16, Lemma 3.1.8]). □

We may now discuss assertions (i) and (v) of Proposition 3.1. We claim that all points which are neither in $Strain[u]$ nor in $Loose[u]$ are limit points of the form described in Corollary 3.5(c). Indeed, if $y \notin Strain[u] \cup Loose[u]$ then we know that $(x, y) \in \Gamma_u$ or $(y, x) \in \Gamma_u$ for some $x \neq y$. In the first case, the relative interior of any minimizing geodesic from x to y is contained in a transport ray, and in the second case, the relative interior of any minimizing geodesic from y to x is contained in a transport ray. This follows from Proposition 3.1(iv). However, the point y does not belong to any transport ray, since it is not in $Strain[u]$. Therefore y must be a limit point as in (c). It now follows from Corollary 3.5(c) that $Strain[u] \cup Loose[u]$ is a set of full measure in M.

Property (v) follows exactly as in [20, Section 5], and the "In fact" part of property (i) is proven as in [16, Theorem 1.5(B)]. These arguments require only the most straightforward modifications, if at all, in order to fit our setting.

4 Needle decomposition with respect to a path space

In the previous section we worked with a Finsler manifold M. In this section the setting is a bit different, we do not require the Finsler structure, and henceforth we assume that M is a smooth manifold, and that μ_1 and μ_2 are two finite, absolutely-continuous measures μ_1, μ_2 satisfying the mass balance condition

$$\mu_1(M) = \mu_2(M).$$

We furthermore assume that we are given a family Γ of parametrized curves in M, and we would like to obtain a result analogous to Proposition 3.1, with the curves of Γ playing the role of the geodesics. That is, we would like to obtain a partition covering almost all of the support of $\mu_1 - \mu_2$ into transport rays with the mass balance condition, yet this time we would like the transport rays to belong to Γ.

The previous section covers the case where Γ consists of all the unit-speed geodesics of some Finsler metric. The actual parametrization of the curves of Γ is of little importance to us in this section, it is only their orientation that really matters.

Definition 4.1. An *n-dimensional path space* is a pair (M, Γ), where M is a smooth n-dimensional manifold, and Γ is a family of smooth parametrized curves in M, such that the following hold:

(i) For each $x \in M$ and $0 \neq v \in T_x M$ there is a unique $\gamma_v \in \Gamma$ which emanates from x in the direction of v, i.e. $\dot{\gamma}_v(0) = \lambda v$ for some $\lambda > 0$. The map $(v, t) \mapsto \gamma_v(t)$ is smooth in both its variables when $t \neq 0$.

(ii) The family Γ is projectively Finsler-metrizable. This means that, possibly after applying smooth, orientation-preserving reparametrizations to each of its members, the family Γ coincides with the family of unit-speed geodesics of some geodesically-convex Finsler metric on M.

Condition (i) follows from Condition (ii), yet we view Condition (ii) as a mere technical requirement. In two dimensions, requirement (ii) always holds locally, see Matsumoto [18]. See Bacataru and Mzsnay [3] for a general discussion on Finsler metrizability of path spaces.

Remark 4.2. Implicit in (ii) is the assumption that if γ is a curve in Γ, then so is every restriction of γ to a smaller subinterval, as well as every precomposition of γ with a translation in $\mathbb{R}$. In particular, we stipulate that every singleton is a member of Γ. The uniqueness in assumption (i) is, of course, only up to restriction or extension of the domain.

Definition 4.3. Let (M, Γ) be an n-dimensional path space. A *regular partition* is a locally-Lipschitz, one-to-one map $F : B \to M$, where B is a parallel line-cluster, with Y, a_y, b_y as in (11), such that:

(i) The image $A = F(B)$ is a Borel subset of M.

(ii) For any $y \in Y$, the curve $t \mapsto F(y, t)$, defined for $a_y < t < b_y$ is a curve in Γ. We call it a *transport ray* associated with the regular partition.

(iii) The map

$$(y, t) \mapsto \left. \frac{\partial F(y, t)}{\partial t} \right|_{t=0} \tag{12}$$

is locally-Lipschitz.

Theorem 4.4. *Let (M, Γ) be an n-dimensional path space. Let μ_1 and μ_2 be compactly-supported, absolutely-continuous, Borel measures on M with $\mu_1(M) = \mu_2(M)$. Then there is a disjoint union $\sqcup_k A_k \subseteq M$, covering the support of $\mu_1 - \mu_2$ up to a set of measure zero, with the following properties:*

(i) For each k, the set A_k is a Borel set admitting a regular partition $F_k : B_k \to M$ with $A_k = F_k(B_k)$, in the sense of Definition 4.1.

(ii) Consider the collection of all transport rays associated with all of the regular partitions of A_k $(k = 1, 2, \ldots)$. A union of such transport rays, which is also a Borel subset of M, is called a transport set. Then for any transport set $S \subseteq M$ we have

$$\mu_1(S) = \mu_2(S). \tag{13}$$

Thanks to Definition 4.1(ii), we already have all of the ingredients required for the proof of Theorem 4.4 in the two-dimensional case. Indeed, note that if Γ is replaced by the collection of geodesics of a geodesically-convex Finsler metric on M, then Theorem 4.4 follows from Proposition 3.1, Lemma 3.4 and Corollary 3.5. However, by Definition 4.1(ii), the collection Γ differs from the collection of geodesics of some geodesically-convex Finsler metric Φ on M by an orientation-preserving reparametrization of its curves. Suppose that F_k are the partitions corresponding to the Finsler structure Φ. Then each F_k can be precomposed with a homeomorphism of the parallel line cluster B_k, smoothly reparametrizing each curve $F_k(y, \cdot)$, which is a geodesic of Φ, into a member of Γ. The new set of regular partitions $\tilde{F}_k$ will satisfy the conclusion of Theorem 4.4. We have thus completed the proof of Theorem 4.4 for $n = 2$.

The generalization of the results of Section 3, and therefore of Theorem 4.4, to arbitrary dimension n, requires a rather straightforward adaptation of the results from [5, 13, 16, 20] to the general case of a non-reversible, geodesically-convex Finsler manifold of arbitrary dimension.

Theorem 4.5. *Let (M, Γ) be a path space with $\dim(M) = 2$, and let μ be an absolutely-continuous measure on M with a smooth density. Let $\rho_1, \rho_2 : M \to [0, \infty)$ be μ-integrable, compactly-supported functions with*

$$\int_M \rho_1 d\mu = \int_M \rho_2 d\mu.$$

Then there is a subcollection $\Lambda \subseteq \Gamma$ of disjoint curves of Γ, a measure ν on Λ and a family $\{\mu_\gamma\}_{\gamma \in \Lambda}$ of Borel measures on M such that the following hold:

(i) For all $\gamma \in \Lambda$, the measure μ_γ is supported on γ.

(ii) For any measurable set $S \subseteq M$,

$$\mu(S) = \int_\Lambda \mu_\gamma(S) d\nu(\gamma). \tag{14}$$

(iii) For ν-almost any $\gamma \in \Lambda$,

$$\int_M \rho_1 d\mu_\gamma = \int_M \rho_2 d\mu_\gamma. \tag{15}$$

The proof, which is included here for completeness in the two-dimensional case, is almost the same as in [5, 16]. The proof also includes expressions for the density of μ_γ with respect to the parametrization of γ which will be useful in the next sections.

Proof. Let μ_1, μ_2 denote the measures whose densities with respect to μ are ρ_1, ρ_2 respectively. Write X for the support of $\mu_1 - \mu_2$. Theorem 4.4 provides us with a collection $(A_k)_{k\geq 1}$ of disjoint Borel subsets of X such that

$$\mu\left(X \setminus \bigsqcup_k A_k\right) = 0, \tag{16}$$

and regular partitions $F_k : B_k \to A_k$ such that (13) holds true for every transport set S. We represent the parallel line-cluster B_k in the form (11) with $Y = Y_k$. Denote

$$A := \bigsqcup_k A_k. \tag{17}$$

Since Theorem 4.5 is indifferent to altering ρ_i on negligible sets, by (16) we may assume that

$$\rho_1 \equiv \rho_2 \qquad \text{on } M \setminus A. \tag{18}$$

We now decompose M into a disjoint union of curves in Γ. Let

$$\Lambda = \bigsqcup_{k=0}^{\infty} \Lambda_k \subseteq \Gamma, \qquad \text{where} \qquad \begin{array}{ll} \Lambda_k := \{F_k(y,\cdot)\,;\, y \in Y_k\} & k \geq 1, \\ \Lambda_0 := \{\{p\}\,;\, p \in M \setminus A\}. & \end{array}$$

By Definition 4.1 and Remark 4.2, the curves $F_k(y,\cdot)$ and the singletons $\{p\} \subseteq M \setminus A$ are indeed members of Γ. Since the sets $A_k = F_k(B_k)$ cover A, we see that M is the disjoint union of the curves in Λ, and we have a map

$$\pi : M \to \Lambda$$

assigning to each $p \in M$ the unique member of Λ passing through p.

Let $k \geq 1$. The map F_k is locally-Lipschitz, and therefore the function

$$J_k = |\det dF_k| \tag{19}$$

is defined almost-everywhere on B_k. The determinant here is understood using the Euclidean volume form on B_k and using the density μ on the manifold M. It follows from Fubini's theorem and the change of variables formula, see e.g. Evans and Gariepy [12], that for every μ-measurable function $\psi : A_k \to \mathbb{R}$,

$$\int_{A_k} \psi d\mu = \int_{B_k} (\psi \circ F_k) J_k dt dy = \int_{Y_k} \int_{a_y}^{b_y} \psi(F_k(y,t)) J_k(y,t) dt dy, \tag{20}$$

where we recall that the parallel line-cluster B_k satisfies (11) with $Y = Y_k$, and for ease of reading we suppress the dependence on k in a_y and b_y.

The measures μ_γ are defined as follows:

- If $\gamma = F_k(y, \cdot) \in \Lambda_k$ for some $k \geq 1$ and some $y \in Y_k$, then we set

$$\mu_\gamma = \gamma_\#(e^{|y|} J_k(y,t)dt), \tag{21}$$

  i.e. μ_γ is the pushforward of the measure $e^{|y|} J_k(y,t)dt$ on (a_y, b_y) via the map $\gamma : (a_y, b_y) \to M$. The factor $e^{|y|}$ is there so that the measure ν_k defined below will be finite.

- If $\gamma = \{p\} \in \Lambda_0$ for some $p \in M \setminus A$, then we set

$$\mu_\gamma = \delta_p,$$

  where δ denotes a Dirac measure.

Next, we define the measure ν on Λ.

- For each $k \geq 1$ define $f_k : Y_k \to M$ by

$$f_k(y) = F_k(y, 0),$$

  and define a measure ν_k on Λ_k by

$$\nu_k := (\pi \circ f_k)_\#(e^{-|y|}dy).$$

  Note that here $e^{-|y|}dy$ is a measure on $Y_k \subseteq \mathbb{R}$. (When we push-forward a measure, we push-forward its σ-algebra as well.)

- On Λ_0 define the measure

$$\nu_0 := \pi_\#(\mu|_{M \setminus A}).$$

Finally, let ν be the measure on Λ satisfying

$$\nu|_{\Lambda_k} = \nu_k \qquad \text{for all } k \geq 0.$$

Then for every measurable set $S \subseteq M$, and for every $k \geq 1$, by virtue of (20),

$$\begin{aligned}
\mu(S \cap A_k) &= \int_{Y_k} \int_{a_y}^{b_y} \chi_S(F_k(y,t)) J_k(y,t) dt dy \\
&= \int_{Y_k} \int_{\pi(f_k(y))} \chi_S(x) e^{-|y|} d\mu_{\pi(f_k(y))}(x) dy \\
&= \int_{Y_k} \mu_{\pi(f_k(y))}(S) e^{-|y|} dy \\
&= \int_{\Lambda_k} \mu_\gamma(S) d\nu_k(\gamma).
\end{aligned}$$

Summing this over k gives

$$\mu(S \cap A) = \sum_{k=1}^{\infty} \int_{\Lambda_k} \mu_\gamma(S) d\nu_k(\gamma). \tag{22}$$

Moreover, by our definition of ν_0,

$$\mu(S \setminus A) = \int_{M\setminus A} \chi_S(p) d\mu(p) = \int_{M\setminus A} \delta_p(S) d\mu(p) = \int_{\Lambda_0} \mu_\gamma(S) d\nu_0(\gamma)$$

which together with (22) and the definition of ν implies (14). In order to prove (15), it suffices to prove that for every ν-measurable subset $S \subseteq \Lambda$,

$$\int_S \left(\int_M \rho_1 d\mu_\gamma \right) d\nu(\gamma) = \int_S \left(\int_M \rho_2 d\mu_\gamma \right) d\nu(\gamma). \tag{23}$$

To this end, we observe that for $i = 1, 2$,

$$\int_S \left(\int_M \rho_i d\mu_\gamma \right) d\nu(\gamma) = \int_\Lambda \left(\int_M \chi_{\pi^{-1}(S)} \rho_i d\mu_\gamma \right) d\nu(\gamma) = \int_{\pi^{-1}(S)} \rho_i d\mu, \tag{24}$$

where the second equality follows from (14). Now if $S \cap \Lambda_0 = \emptyset$, then $\pi^{-1}(S)$ is a transport set, and (23) follows from (13) and (24), while if $S \subseteq \Lambda_0$ then $\pi^{-1}(S) \subseteq M \setminus A$, so (23) follows from (18) and (24). By writing $\pi^{-1}(S) = \pi^{-1}(S \setminus \Lambda_0) \sqcup \pi^{-1}(S \cap \Lambda_0)$, we are done. □

The density J_k in formula (19) is the absolute value of a determinant. The following lemma will be used later in order to eliminate the absolute value.

Lemma 4.6. *We work under the notation introduced in the proof of Theorem 4.5. Assume that M is orientable, that $n = 2$ and let $k \geq 1$. Then for almost any $y_0 \in Y_k$, the function $t \mapsto \det dF_k(y_0, t)$ does not change sign in the interval (a_{y_0}, b_{y_0}).*

Proof. We may assume that y_0 is a Lebesgue density point of $Y := Y_k$ and is a Lebesgue point of the functions a_y and b_y. Thus $\liminf_{Y \ni y \to y_0} a_y \leq a_{y_0}$ while $\limsup_{Y \ni y \to y_0} b_y \geq b_{y_0}$. Assume by contradiction that there exist $t_1 < t_2$ in the interval (a_{y_0}, b_{y_0}) such that

$$\det dF(y_0, t_1) \cdot \det dF(y_0, t_2) < 0, \tag{25}$$

i.e., the linear basis

$$\frac{\partial F}{\partial t}(y_0, t), \frac{\partial F}{\partial y}(y_0, t) \in T_{F(y_0,t)}M$$

has a different orientation for $t = t_1$ and for $t = t_2$. The curve $t \mapsto F(y_0, t)$ is a simple regular curve defined for $t \in (a_{y_0}, b_{y_0})$, and the vector $\frac{\partial F}{\partial t}(y_0, t)$ is a smooth, non-zero tangent to this curve. Hence, by (25), the two vectors $(\partial F/\partial y)(y_0, t_1)$ and $(\partial F/\partial y)(y_0, t_2)$ point at different sides of the curve γ. This implies that the curves $F(y, \cdot)$ and $F(y_0, \cdot)$ must cross each other for some y sufficiently close to y_0, in contradiction to Definition 4.3. □

5 Horocycles in the hyperbolic plane

In this section, we take M to be the hyperbolic plane $\mathbb{H}^2$, with some fixed orientation, and Γ to be the collection of unit-speed oriented horocycles, which are the unit-speed curves of constant signed geodesic curvature 1. This means that the ordered pair $(\dot\gamma, \nabla_{\dot\gamma}\dot\gamma)$ is an oriented orthonormal frame along γ. Each pair of points $p, q \in \mathbb{H}^2$ are joined by a unique oriented horocycle, whose length is $2\sinh(d(p,q)/2)$, where d is hyperbolic distance.

It is observed in Crampin and Mestdag [9] that the collection of oriented horocycles is projectively Finsler-metrizable (though [9] does not prove geodesic convexity, which we establish below). In the Poincaré disc model, the relevant Finsler metric is given by:

$$\Phi(x,y,u,v) = 2 \cdot \frac{\sqrt{u^2+v^2} + uy - xv}{1-x^2-y^2} \qquad (u,v) \in T_{(x,y)}D,\ (x,y) \in D \tag{26}$$

where $D = \{z \in \mathbb{C}\,;\, |z| < 1\} \cong \{(x,y) \in \mathbb{R}^2\,;\, x^2+y^2 < 1\}$ and the tangent plane to $\mathbb{R}^2$ at (x,y) is identified with $\mathbb{R}^2$ in the natural manner, with $(u,v) \in \mathbb{R}^2$ being the coordinates in $T_{(x,y)}D$. Let us explain the computations in [9] by proving a slightly more general fact.

Lemma 5.1. *Let (M,g) be an oriented Riemannian surface and let η be a 1-form on M with $|\eta|_g < 1$ such that $d\eta$ is the Riemannian area form. Let $0 \le \kappa < 1$ and define $F : TM \to \mathbb{R}$ via*

$$F(v) = |v| - \kappa \cdot \eta(v) \qquad\qquad (v \in TM), \tag{27}$$

where $|v| = |v|_g$ is the Riemannian norm of the tangent vector v. Then F determines a Finsler structure on M whose geodesics are curves of constant geodesic curvature κ relative to the oriented Riemannian surface (M,g).

(The parametrization of these Finsler geodesics is not necessarily by Riemannian arclength.)

Proof. Since $|\eta| < 1$, the function F is non-negative, fiberwise convex, and smooth outside the zero section of the tangent bundle. Hence it determines a Finsler structure (Finsler metrics of this form are called *Randers metrics*). We claim that the geodesic equation of this Finsler structure is

$$\nabla_{\dot\gamma}\dot\gamma + \kappa|\dot\gamma|\dot\gamma^{\perp} = 0, \tag{28}$$

where ∇ is the Levi-Civita connection of (M,g). Here, $\perp$ is rotation by $\pi/2$, whose direction is determined by the orientation of M. Equation (28) is the equation of constant geodesic curvature κ. We will derive the geodesic equation (28) following the notation in Cheeger and Ebin [7, Page 4].

Let $p, q \in M$ be two points. Suppose that $\alpha_s(t) : [0,1] \to M$, defined for $s \in (-\varepsilon, \varepsilon)$, is a family of curves in M connecting the point $p = \alpha_s(0)$ to the point $q = \alpha_s(1)$. We assume that $\alpha(s,t) = \alpha_s(t)$ varies smoothly in s and t, and denote $T = \partial\alpha/\partial t$ and $S = \partial\alpha/\partial s$. Note that $[S,T] = 0$. The Finsler length

$$\text{Length}(\alpha_s) = \int_0^1 F(\alpha_s'(t))dt = \int_0^1 \left[|T| - \kappa \cdot \eta(T)\right] dt$$

does not depend on the parametrization of the curve α_s, since F is 1-homogenous. Hence we may assume that α_s is of constant speed for $s = 0$, i.e., $|T|$ is constant in t, without changing the lengths of the curves α_s. Consequently,

$$\begin{aligned}\frac{d}{ds}\int_0^1 F(\alpha_s'(t))dt\bigg|_{s=0} &= \int_0^1 S\left[|T| - \kappa\cdot\eta(T)\right]dt \\ &= \frac{1}{|T|}\int_0^1 \langle \nabla_S T, T\rangle dt - \kappa\int_0^1 [\nabla_S\eta(T) + \eta(\nabla_S T)]dt.\end{aligned}$$

Since $\nabla_S T - \nabla_T S = [T, S] = 0$ and since $S|_{t=0,1} = 0$ we may integrate by parts without boundary terms as in [7] and rewrite this as

$$\frac{1}{|T|}\int_0^1 \langle \nabla_T S, T\rangle dt - \kappa\int_0^1 [\nabla_S\eta(T) + \eta(\nabla_T S)]dt = -\frac{1}{|T|}\int_0^1 \langle S, \nabla_T T\rangle dt + \kappa\int_0^1 [\nabla_T\eta(S) - \nabla_S\eta(T)]dt.$$

Since $d\eta(X, Y) = (\nabla_X\eta)(Y) - (\nabla_Y\eta)(X)$, the curve α_0 is critical for the Finsler length if and only if

$$-\frac{\nabla_T T}{|T|} + \kappa(d\eta)^{\#}(T) = 0,$$

where $(d\eta)^{\#}(X) = Y$ if $d\eta(X, V) = \langle Y, V\rangle$ for all V. Since $d\eta$ is the Riemannian area form, $(d\eta)^{\#}$ is simply a rotation by $\pi/2$, and (28) is proven. □

Remark 5.2. Lemma 5.1 remains valid in n dimensions, for even n, if instead of requiring that $d\eta$ is the Riemannian area form we require that the Riemannian metric g and the symplectic form $d\eta$ yield a Kähler structure.

Remark 5.3. Formula (27) is reminiscent of the magnetic Lagrangian $L(v) = |v|^2 - \eta_x(v)$, whose critical curves are again curves of constant geodesic curvature in the Riemannian manifold (M, g). However, in the magnetic case the constant geodesic curvature is *not* the same for different critical curves: a charged particle of very high speed barely rotates when it enters a magnetic field, unlike a slower particle of the same mass and charge. In contrast, the Lagrangian associated with (27) yields critical curves whose constant geodesic curvature is always κ.

In the Poincaré disc model, we set $\eta = 2(ydx - xdy)/(1 - x^2 - y^2)$ so that $|\eta| < 1$ while $d\eta$ is the hyperbolic area form, and observe that the Finsler structure Φ from (26) coincides with F from Lemma 5.1 for $\kappa = 1$.

Lemma 5.4. *Every geodesic of Φ is uniquely minimizing.*

Proof. Let $p, q \in \mathbb{H}^2$. Let γ be a curve joining p to q, and let $\tilde{\gamma}$ be the oriented horocycle joining p to q. We need to show that

$$\int_\gamma \Phi \geq \int_{\tilde{\gamma}} \Phi. \tag{29}$$

Recall that

$$\Phi = \sqrt{g} - \eta,$$

where g is the hyperbolic metric and η is a 1-form such that $d\eta$ is the hyperbolic area form.

We can work in the upper half plane model of the hyperbolic plane, and thanks to the symmetries of the hyperbolic plane, we may assume that p, q both lie on the horizontal line $y = 1$, and that the horocycle $\tilde{\gamma}$ is of the form $t \mapsto (t, 1)$. We may also replace η by any primitive of the hyperbolic area form, as this will only add the same constant to both sides of (29). On the upper half plane, the 1-form dx/y is a primitive of the hyperbolic area form $(dx \wedge dy)/y^2$.

The 1-form dy vanishes on the horocycle $\tilde{\gamma}$. Moreover, we note that $dx \geq 0$ on $\tilde{\gamma}$ in the sense that $dx(v) \geq 0$ for $v = \tilde{\gamma}'(t)$. Therefore,

$$\int_\gamma \frac{\sqrt{dx^2 + dy^2}}{y} - \frac{dx}{y} \geq \int_\gamma \frac{|dx|}{y} - \frac{dx}{y} \geq 0 = \int_{\tilde{\gamma}} \frac{\sqrt{dx^2 + dy^2}}{y} - \frac{dx}{y}.$$

Equality in the above inequality holds if and only if $dy \equiv 0$ and $dx \geq 0$ on γ, which means that γ is an orientation-preserving reparametrization of $\tilde{\gamma}$. ◻

The distance function induced by Φ has the following simple description in the Poincaré disc model. We will not make use of this fact, but we find it interesting enough to mention.

Proposition 5.5. *If z, w are two points in the unit disc, then $d_\Phi(z, w) = \angle(z, p, w)$, where p is the* Euclidean *center of the oriented horocycle joining z to w, and the angle is a Euclidean angle. In particular, the distance between any two points is less than 2π, as well as the distance from any point to the boundary.*

Proof. Consider the parametrized circle $(1-r+r\cos t, r\sin t)$, $t \in [-\pi, \pi]$. A direct calculation shows that its Finslerian speed is 1 for any $0 < r < 1$. Since it is tangent to the boundary of the unit disc at the point $(1, 0)$, it follows that it is a unit-speed geodesic with respect to the metric Φ. By symmetry and homogeneity it follows that the Finslerian arclength along any horocycle is the same as *angular speed* with respect to the Euclidean center of the horocycle. Since each horocycle is a minimizing geodesic with respect to Φ, the proposition follows. ◻

Horocycles are thus geodesics of a Finsler metric on the unit disc; albeit not complete, this Finsler metric is *strongly convex*, as every two points on the hyperbolic plane are joined by a unique oriented horocycle, which is a uniquely minimizing geodesic by Lemma 5.4. Denoting the collection of oriented horocycles by Γ, we see that Γ is a path space according to Definition 4.1, where we choose the parametrization of the horocycles to be hyperbolic (not Finslerian!) arclength parametrization. The key to the horocyclic Brunn-Minkowski inequality (4) lies in the following lemma.

Lemma 5.6. *Let $M = \mathbb{H}^2$, let Γ be the collection of oriented horocycles with hyperbolic arclength parametrization, and let μ be the hyperbolic area measure. Then the decomposition described in Theorem 4.5 can be chosen so that ν-almost every measure μ_γ which is not a Dirac measure has a density with respect to arclength which is affine-linear.*

Proof. We work under the notations introduced in the proof of Theorem 4.5. When μ_γ is not a Dirac measure, it is given by formula (21). Thus, by the definition of ν, it suffices to prove that for any of the partitions $F_k : B_k \to A_k$, the Jacobian $J_k(y,t) = |\det dF_k(y,t)|$ is affine-linear in the variable t, for almost every $y \in Y_k$.

We consider the Poincaré disc model of the hyperbolic plane, identifying $\mathbb{H}^2$ with the unit disc of the complex plane $\{z \in \mathbb{C}\,;\, |z| < 1\}$ endowed with the Poincaré metric $2|dz|/(1-|z|^2)$. The tangent space at any point of the unit disc will be naturally identified with $\mathbb{C}$. The collection Γ of constant-speed, oriented horocycles in the Poincaré disc is exactly the set of curves of the form

$$\alpha_{\lambda,t_0,\omega}(t) = \omega \cdot \frac{t - t_0 + (1-\lambda)i}{t - t_0 + (1+\lambda)i} \qquad (t \in \mathbb{R}) \tag{30}$$

for $\lambda > 0$, $t_0 \in \mathbb{R}$ and ω on the unit circle. From Lemma 5.1 and Lemma 5.4 we conclude that (M, Γ) is indeed a path space in the sense of Definition 4.1.

Given a point z in the unit disc and a unit tangent vector v at z of unit Euclidean norm, the unique oriented horocycle which, at time t, visits z with velocity in the direction v, is represented by the parameters

$$\begin{aligned} \lambda &= \frac{1-|z|^2}{|v - iz|^2} \\ t_0 &= t - \lambda a \\ \omega &= \frac{(t - t_0 + i(1+\lambda))^2}{2\lambda i} v. \end{aligned}$$

In particular, ω, t_0, λ are smooth functions of z, v and t.

Recall the notation from the proof of Theorem 4.5. Fix $k \geq 1$ and denote $A = A_k$ and $B = B_k$, so that $A = F(B)$. Recall from (11) that the parallel line cluster B is written as

$$B = \{(y,t)\,;\, y \in Y,\, a_y < t < b_y\}\,.$$

Since $F(y, \cdot) \in \Gamma$ for every $y \in Y$, we can write for any $y \in Y$,

$$F(y,t) = \omega(y) \cdot \frac{t - t_0(y) + (1-\lambda(y))i}{t - t_0(y) + (1+\lambda(y))i} \qquad (a_y < t < b_y) \tag{31}$$

for some functions $\lambda, t_0, \omega : Y \to (0,\infty) \times \mathbb{R} \times S^1$, where $S^1 = \{z \in \mathbb{C}\,;\, |z| = 1\}$. The map

$$y \mapsto (\lambda(y), t_0(y), \omega(y)) \tag{32}$$

is the composition of the map $y \mapsto (\partial F/\partial t)(y, 0)$ with the map sending a unit tangent vector to the triple (λ, t_0, ω) corresponding to the horocycle passing through that vector. The first map is locally Lipschitz by the definition of a regular partition, and second is smooth by the discussion above. Therefore, the map in (32) is differentiable at almost every $y \in Y$.

Pick such a point of differentiability $y \in Y$. Then for all $a_y < t < b_y$, we see from (31) that the map F is differentiable at (y, t). The determinant of dF with respect to the *Euclidean* area forms on $B \subseteq \mathbb{R}^2$ and on $A \subseteq \mathbb{C}$, is given by

$$\operatorname{Im}\left(\overline{\partial_t F(y,t)}\partial_y F(y,t)\right). \tag{33}$$

Write $\omega = e^{i\phi}$. Then by (31),

$$\partial_t F = \frac{2\lambda i e^{i\phi}}{\left((t-t_0)+i(1+\lambda)\right)^2}, \qquad \partial_y F = \frac{ie^{i\phi}\Big(\phi'\left((t-t_0+i)^2+\lambda^2\right) - 2(t-t_0+i)\lambda' - 2t_0'\lambda\Big)}{\left((t-t_0)+i(1+\lambda)\right)^2}$$

and

$$|F|^2 = \frac{(t-t_0)^2+(1-\lambda)^2}{(t-t_0)^2+(1+\lambda)^2} = 1 - \frac{4\lambda}{(t-t_0)^2+(1+\lambda)^2},$$

whence

$$\begin{aligned}\operatorname{Im}\left(\overline{\partial_t F}\partial_y F\right) &= \frac{4\lambda\phi'(t-t_0)-4\lambda\lambda'}{\left((t-t_0)^2+(1+\lambda)^2\right)^2} \\ &= \frac{(1-|F|^2)^2}{4\lambda}\cdot\big((t-t_0)\phi'-\lambda'\big).\end{aligned}$$

In the Poincaré disc model, the hyperbolic area density form is given by $4|dz|^2/(1-|z|^2)^2$. It follows that, this time with respect to the Euclidean area form on B and the hyperbolic area form on A,

$$J = |\det dF| = |\lambda^{-1}((t-t_0)\phi' - \lambda')| \tag{34}$$

at any point (y,t) such that λ, t_0, ϕ are differentiable at y. Note that the expression in absolute value in (34) is affine-linear in t. Thus, in order to conclude that J is affine-linear and finish the proof, it remains to prove that for almost every $y \in Y$, this expression $\det dF$ does not change sign for $t \in (a_y, b_y)$. This is precisely the content of Lemma 4.6, and the proof is complete. □

Remark 5.7. One might ask whether it is possible to find a Finsler structure $\tilde{\Phi}$ whose geodesics are oriented horocycles in their *hyperbolic arc-length* parametrization. The answer is negative. Indeed, such a metric would satisfy

$$d_{\tilde{\Phi}}(z,w) = 2\sinh(d_g(z,w)/2), \qquad z, w \in \mathbb{H}^2,$$

where g denotes the hyperbolic metric. By differentiation it would then follow that $\tilde{\Phi}$ and g coincide, which is a contradiction. We did not rule out the possibility that a *Lagrangian* exists whose minimizers are the set of unit-speed oriented horocycles. Such a Lagrangian cannot be homogeneous by the above remark, and apriori it might be time-dependent.

6 A Brunn-Minkowski inequality for horocycles

In this section we prove a generalized, functional version of Theorem 1.1, in spirit of the Prékopa-Leindler and Borell-Brascamp-Lieb inequalities. Given $x, y \in \mathbb{H}^2$ and $0 < \lambda < 1$ we write

$$[x : y]_\lambda = \gamma(\lambda) \tag{35}$$

where $\gamma : [0,1] \to \mathbb{H}^2$ is a constant-speed oriented horocycle with $\gamma(0) = x$ and $\gamma(1) = y$. As explained in the Introduction, the point $[x : y]_\lambda \in \mathbb{H}^2$ is well-defined.

Given two numbers $a, b > 0$ and parameters $0 < \lambda < 1, p \in [-\infty, +\infty]$, we define the λ-weighted p-mean of a and b via

$$M_p(a, b; \lambda) = [(1-\lambda)a^p + \lambda b^p]^{1/p},$$

in the case where $p \in \mathbb{R} \setminus \{0\}$. In the remaining cases we interpret $M_p(a, b; \lambda)$ by continuity as follows:

$$M_p(a, b; \lambda) = \begin{cases} a^{1-\lambda} b^\lambda & p = 0 \\ \max\{a, b\} & p = +\infty \\ \min\{a, b\} & p = -\infty \end{cases}$$

Thus $M_p(a, b; \lambda)$ is a continuous, non-decreasing, bounded function of $p \in [-\infty, +\infty]$. The quantity $M_p(a, b; \lambda)$ is 1-homogeneous in the pair $(a, b) \in (0, \infty) \times (0, \infty)$. A useful property of p-means, which follows from Hölder's inequality, is that if $1/p_1 + 1/p_2 = 1/p$ then

$$M_{p_1}(a_1, b_1; \lambda) \cdot M_{p_2}(a_2, b_2; \lambda) \geq M_p(a_1 a_2, b_1 b_2; \lambda), \tag{36}$$

provided that at least one of p_1, p_2 belongs to $[0, +\infty]$, and provided that both belong to $[0, +\infty]$ in the case where $p \in [0, +\infty]$. Here we interpret $1/p$ as $+\infty$ when $p = 0$ and as 0 when $p = \pm\infty$, and similarly for $1/p_1$ and $1/p_2$. All integrals in the following theorem are carried out with respect to the hyperbolic area measure.

Theorem 6.1. *Let $f, g, h : \mathbb{H}^2 \to [0, \infty)$ be three measurable functions, and assume that f and g are integrable with a non-zero integral. Let $0 < \lambda < 1$ and $p \in [-1/2, +\infty]$. Assume that for any $x, y \in \mathbb{H}^2$ with $f(x)g(y) > 0$,*

$$h([x : y]_\lambda) \geq M_p(f(x), g(y); \lambda). \tag{37}$$

Define $q = p/(1+2p)$, where by continuity $q = -\infty$ if $p = -1/2$ and $q = 1/2$ if $p = +\infty$. Then,

$$\int_{\mathbb{H}^2} h \geq M_q\left(\int_{\mathbb{H}^2} f, \int_{\mathbb{H}^2} g; \lambda\right). \tag{38}$$

The case where $p = q = 0$ in Theorem 6.1 is particularly simple and is referred to as the *horocyclic Prekópa-Leindler inequality*, since in this case the means M_p and M_q are just geometric averages. We proceed with a proof of Theorem 6.1. Let f, g, h satisfy the requirements of Theorem 6.1. By a standard approximation argument, we may assume that f and g are compactly-supported. Let μ denote the hyperbolic area measure in $\mathbb{H}^2$. Denote

$$\rho_1 = \frac{f}{\int_{\mathbb{H}^2} f d\mu} \qquad \text{and} \qquad \rho_2 = \frac{g}{\int_{\mathbb{H}^2} g d\mu},$$

where the integrals are finite and positive by our assumptions. Theorem 4.5 provides us with a family Λ of unit-speed oriented horocycles (possibly singletons), a measure ν on Λ, and a family $\{\mu_\gamma\}_{\gamma \in \Lambda}$ of Borel measures, where μ_γ is supported on γ for each $\gamma \in \Lambda$, such that (14) and (15) hold. By (15) and the definition of ρ_i, for ν-almost any $\gamma \in \Lambda$,

$$\int_{\mathbb{H}^2} f d\mu_\gamma = \int_{\mathbb{H}^2} f d\mu \int_{\mathbb{H}^2} \rho_1 d\mu_\gamma = \int_{\mathbb{H}^2} f d\mu \int_{\mathbb{H}^2} \rho_2 d\mu_\gamma = \frac{\int_{\mathbb{H}^2} f d\mu}{\int_{\mathbb{H}^2} g d\mu} \cdot \int_{\mathbb{H}^2} g d\mu_\gamma. \tag{39}$$

Lemma 6.2. *For ν-almost any $\gamma \in \Lambda$, if $0 < \int f d\mu_p < \infty$ then*

$$\int_{\mathbb{H}^2} h d\mu_\gamma \geq M_q \left(\int_{\mathbb{H}^2} f d\mu_\gamma, \int_{\mathbb{H}^2} g d\mu_\gamma \, ; \lambda \right). \tag{40}$$

Proof. If the measure μ_γ is a Dirac mass at the point $x \in \mathbb{H}^2$, then by applying (37) with $x = y = [x : y]_\lambda$,

$$\begin{aligned} \int_{\mathbb{H}^2} h d\mu_\gamma &= h(x) \geq M_p \left(f(x), g(x) \, ; \lambda \right) \\ &= M_p \left(\int_{\mathbb{H}^2} f d\mu_\gamma, \int_{\mathbb{H}^2} g d\mu_\gamma \, ; \lambda \right) \geq M_q \left(\int_{\mathbb{H}^2} f d\mu_\gamma, \int_{\mathbb{H}^2} g d\mu_\gamma \, ; \lambda \right), \end{aligned}$$

where we used $p \geq q$ in the last passage. Hence (40) holds true when μ_γ is a Dirac mass. We may thus restrict attention to the case where μ_γ is not a Dirac mass, whence $\gamma : I \to \mathbb{H}^2$ is a unit-speed horocycle arc. Here $I \subseteq \mathbb{R}$ is some interval, which is possibly a ray or the full line. Denote

$$\tilde{f} = f \circ \gamma, \qquad \tilde{g} = g \circ \gamma, \qquad \tilde{h} = h \circ \gamma.$$

According to (37), for any $x, y \in I$ with $\tilde{f}(x)\tilde{g}(y) > 0$,

$$\tilde{h}((1-\lambda)x + \lambda y) \geq M_p \left(\tilde{f}(x), \tilde{g}(y) \, ; \lambda \right). \tag{41}$$

Lemma 5.6 tells us that the density of μ_γ with respect to arclength, denoted by ρ, is affine linear. Denote $\tilde{\rho} = \rho \circ \gamma$. Then $\tilde{\rho} : I \to [0, \infty)$ is affine-linear and in particular concave. Hence, for $x, y \in I$,

$$\tilde{\rho}((1-\lambda)x + \lambda y) \geq M_1 \left(\tilde{\rho}(x), \tilde{\rho}(y) \, ; \lambda \right). \tag{42}$$

Write $F = \tilde{f}\tilde{\rho}$ and $G = \tilde{g}\tilde{\rho}$ and also $H = \tilde{h}\tilde{\rho}$. It follows from (36), (41) and (42) that

$$H((1-\lambda)x + \lambda y) \geq M_{\tilde{p}} \left(F(x), G(y) \, ; \lambda \right) \tag{43}$$

for any $x, y \in I$ with $\tilde{f}(x)\tilde{g}(y) > 0$. Here $\tilde{p} = p/(p+1)$ and we used that $p > -1$. By the Borell-Brascamp-Lieb inequality in one dimension (see e.g. [14, Section 10]), we conclude from (43) that

$$\int_I H \geq M_q\left(\int_I F, \int_I G\,; \lambda\right) \tag{44}$$

where $q = \tilde{p}/(\tilde{p}+1) = p/(2p+1)$ and we used that $p \geq -1/2$. Since $\int H = \int h d\mu_\gamma$ and similarly for f and g, the desired inequality (40) is equivalent to (44). □

Proof of Theorem 6.1. Since f is μ-integrable, by (14) we know that $\int f d\mu_\gamma < \infty$ for ν-almost any $\gamma \in \Lambda$. Thanks to (39) and the 1-homogeneity of M_q, we can rewrite (40) as follows: for ν-almost any $\gamma \in \Lambda$,

$$\int_{\mathbb{H}^2} h d\mu_\gamma \geq \int_{\mathbb{H}^2} f d\mu_\gamma \cdot M_q\left(1, \frac{\int_{\mathbb{H}^2} g d\mu}{\int_{\mathbb{H}^2} f d\mu}\,; \lambda\right) = \int_{\mathbb{H}^2} f d\mu_\gamma \cdot A,$$

where we define A as the λ-weighted q-mean of 1 and the ratio $\int_{\mathbb{H}^2} g d\mu / \int_{\mathbb{H}^2} f d\mu$. By (14),

$$\begin{aligned}
\int_{\mathbb{H}^2} h d\mu &= \int_\Lambda \left(\int_{\mathbb{H}^2} h d\mu_\gamma\right) d\nu(\gamma) \\
&\geq A \cdot \int_\Lambda \left(\int_{\mathbb{H}^2} f d\mu_\gamma\right) d\nu(\gamma) \\
&= A \cdot \int_{\mathbb{H}^2} f d\mu \\
&= M_q\left(\int_{\mathbb{H}^2} f d\mu, \int_{\mathbb{H}^2} g d\mu\,; \lambda\right),
\end{aligned}$$

and (38) is proved. □

Corollary 6.3. *Let $A, B \subseteq \mathbb{H}^2$ be Borel measurable sets of positive (or infinite) hyperbolic area. Then for any $0 < \lambda < 1$,*

$$\mu([A : B]_\lambda])^{1/2} \geq (1-\lambda) \cdot \mu(A)^{1/2} + \lambda \cdot \mu(B)^{1/2}. \tag{45}$$

Proof. By monotonicity, it suffices to prove (45) under the additional assumptions that $\mu(A) < \infty$ and $\mu(B) < \infty$. Set $f = \chi_A, g = \chi_B$ and $h = \chi_{[A:B]_\lambda}$. Observe that the requirements of Theorem 6.1 are satisfied with $p = +\infty$. Thus $q = 1/2$, and (45) follows from conclusion (38) of Theorem 6.1. □

All that remains in order to complete the proof of Theorem 1.1 is to show that equality is attained for concentric circles, and to deal with non-empty sets $A, B \subseteq \mathbb{H}^2$ whose hyperbolic area might be zero.

Let us analyze the case of concentric circles. Let $p \in \mathbb{H}^2$ and $0 < r_0 < r_1$. Let A, B denote the closed discs with center p and radii r_0, r_1 respectively. The set $[A : B]_\lambda$ is connected, being the image of the connected set $A \times B$ under a continuous map. By symmetry, $[A : B]_\lambda$ is a disc centered at the point p, and we denote its radius by r_λ. The area of a hyperbolic disc of radius r is $4\pi \sinh^2(r/2)$, and therefore, in order to prove that equality holds in (45), we should prove that

$$\sinh(r_\lambda/2) = (1-\lambda) \cdot \sinh(r_0/2) + \lambda \cdot \sinh(r_1/2). \tag{46}$$

By definition,

$$r_\lambda = \max_\gamma d(p, \gamma(\lambda)),$$

where the maximum is over constant-speed oriented horocycles γ with $\gamma(0) \in A$ and $\gamma(1) \in B$, i.e.,

$$d(p, \gamma(0)) \leq r_0 \qquad \text{and} \qquad d(p, \gamma(1)) \leq r_1. \tag{47}$$

Thus to find r_λ we need to maximize $d(p, \gamma(\lambda))$ over all constant-speed oriented horocycles γ satisfying (47). If we let

$$h(p, q) := 2\sinh(d(p, q)/2)$$

denote the length of a horocycle joining p and q, then by monotonicity the same horocycle γ would also maximize $h(p, \gamma(\lambda))$.

Lemma 6.4. *The function $t \mapsto h(p, \gamma(t))$, for γ a constant-speed horocycle, is convex.*

Proof. By the symmetries of the hyperbolic plane, it suffices to consider the point $p = i$ in the upper half plane and the horocycle $\gamma(t) = at + yi$ where $a, y > 0$. In this case

$$h(i, at + yi) = \sqrt{a^2t^2 + (1-y)^2}/\sqrt{y},$$

which is a convex function of t. □

By Lemma 6.4, for every constant-speed horocycle γ satisfying (47),

$$\begin{aligned} h(p, \gamma(\lambda)) &\leq (1-\lambda) \cdot h(p, \gamma(0)) + \lambda \cdot h(p, \gamma(1)) \\ &\leq (1-\lambda) \cdot 2\sinh(r_0/2) + \lambda \cdot 2\sinh(r_1/2). \end{aligned}$$

Equality is attained when γ is a horocycle through p satisfying $\gamma(0) \in \partial A$ and $\gamma(1) \in \partial B$, because then equality holds in (47) and the function $t \mapsto h(p, \gamma(t))$ is affine-linear. Therefore,

$$2\sinh(r_\lambda/2) = \max_\gamma h(p, \gamma(\lambda)) = (1-\lambda) \cdot 2\sinh(r_0/2) + \lambda \cdot 2\sinh(r_1/2),$$

which gives (46). We have thus shown that equality in (45) is attained for concentric discs.

Proof of Theorem 1.1. Let $A, B \subseteq \mathbb{H}^2$ be non-empty, Borel measurable sets. In the case where $Area(A) > 0$ and $Area(B) > 0$, inequality (4) follows from Corollary 6.3, as μ is the hyperbolic area measure. In the case where $Area(A) = Area(B) = 0$, inequality (4) holds trivially. We need to prove (4) in the case where $Area(A) = 0$ but $Area(B) > 0$. Since A is non-empty, we may pick a point $O \in A$. For $t > 0$ denote by

$$t \times B := \{\gamma(t)\,;\, \gamma(0) = O, \gamma(1) \in B \text{ and } \gamma \text{ is a constant-speed oriented horocycle}\}\,, \qquad (48)$$

the horocyclic dilatation of B with respect to the point O. It is evident that

$$\lambda \times B \subseteq [A : B]_\lambda.$$

Hence inequality (4) would follow once we show that for any $t > 0$,

$$Area\,(t \times B) = t^2 \cdot Area(B), \qquad t > 0. \qquad (49)$$

To prove (49), consider "horocyclic polar coordinates" centered at O, in which the point (r, θ) corresponds to $\gamma_\theta(r)$, where γ_θ is the unit-speed oriented horocycle emanating from O at angle θ from some fixed direction. In the Poincaré disc model, if we take O to be the origin, then it follows from (30) that this coordinate map is

$$(r, \theta) \mapsto e^{i\theta} \cdot \frac{r}{r + 2i}.$$

By using formulae (31) and (34) with $\phi' = 1, t_0 = 0$ and $\lambda = 1$ (where r, θ are replaced by t, y), it follows that in this coordinate system, the volume form is given by $r dr \wedge d\theta$, which implies (49). Inequality (4) is therefore proven in all cases. Since we have already shown that equality holds in (4) when A and B are concentric discs, the proof of the theorem is complete. ☐

Remark 6.5. The succinct formulation (2) admits a horocyclic analogue, if one is willing to make some non-canonical choices. Let $O \in \mathbb{H}^2$ be a marked point viewed as "the origin", and for every $B \subseteq \mathbb{H}^2$ and $t > 0$ define $t \times B$ as in (48). Recall that the hyperbolic area of horocyclic dilatation of B with respect to the point O satisfies the scaling property (49). If we denote $[A : B] = 2 \times [A : B]_{1/2}$, then Theorem 1.1 implies that for any Borel measurable, non-empty sets $A, B \subseteq \mathbb{H}^2$,

$$Area([A : B])^{1/2} \geq Area(A)^{1/2} + Area(B)^{1/2} \qquad (50)$$

with equality when A and B are concentric discs. Indeed, (50) is a consequence of (45) with $\lambda = 1/2$, together with (49).

The main geometric property of the family of horocycles that was used in the proof of the inequality of Theorem 1.1 is Lemma 5.6, i.e. that the density of the measures μ_γ is affine-linear with respect to arclength. In fact, we only used the fact that the density of μ_γ with respect to arclength is concave. Our proof therefore yields a a more general result:

Definition 6.6. Let (M, Γ) be a path space. We can define the *Minkowski sum associated with* Γ, for any $A, B \subseteq M$ and $0 \leq \lambda \leq 1$, to be

$$[A:B]_\lambda^\Gamma := \{\gamma((1-\lambda)s + \lambda t)\,;\, \gamma \in \Gamma, \gamma(s) \in A, \gamma(t) \in B, s < t\}.$$

Theorem 6.7. *Let (M, Γ) be a two-dimensional path space and let μ be an absolutely continuous measure on M with a smooth density. Assume that for every regular partition $F : B \to M$ associated with (M, Γ) (see Definition 4.3), the function $J(y,t) = |\det dF(y,t)|$ is well-defined and concave in t for almost every $y \in Y$. Then for every Borel sets $A, B \subseteq M$ of finite, positive μ-measure and for every $0 < \lambda < 1$,*

$$\mu\left([A:B]_\lambda^\Gamma\right)^{1/2} \geq \lambda \cdot \mu(A)^{1/2} + (1-\lambda) \cdot \mu(B)^{1/2}.$$

Remark 6.8. Consider the n-dimensional hyperbolic space $\mathbb{H}^n$. Suppose that $A, B \subseteq \mathbb{H}^n$ are two concentric balls, with radii r_0, r_1 respectively. Write $C \subseteq \mathbb{H}^n$ for the collection of all midpoints of constant-speed curves of geodesic-curvature 1 connecting a point in A with a point in B. Then $C \subseteq \mathbb{H}^n$ is a ball of radius $r_{1/2} > 0$, where relation (46) holds true with $\lambda = 1/2$. This follows from an n-dimensional generalization of Lemma 6.4. Hence,

$$f_n(Vol_n(C)) = \frac{f_n(Vol_n(A)) + f_n(Vol_n(B))}{2},$$

where Vol_n is hyperbolic volume, and $f_n(V) := \sinh(R_n(V)/2)$ where $R_n(V)$ is the radius of a hyperbolic ball with volume V. In the two-dimensional case, $f_2(t) = c\sqrt{t}$ for some universal constant $c > 0$, but for $n \geq 3$ the function f_n is not so simple. Thus, if concentric balls attain equality for all n, then an n-dimensional version of theorem 1.1 would be more complicated.

7 Appendix: Proof of Lemma 3.3

Let (M, Φ) be a two-dimensional Finsler manifold. Let $\pi : TM \to M$ denote the tangent bundle map of M, and let $SM \subseteq TM$ denote the unit tangent bundle. For each $v \in SM$, let γ_v denote the unit-speed geodesic satisfying $\dot{\gamma}_v(0) = v$, defined on its maximal domain of definition. As in [2, Section 2.1] at each $p \in M$ we consider the *Hessian metric* induced by $\Phi^2/2$ on the fiber T_pM. That is, we consider the strongly convex function $\Phi^2/2$ in the linear space T_pM, and for $u, v, w \in T_pM$ with $u \neq 0$ we set

$$g_u(v, w) = \frac{1}{2}\partial_v \partial_w \Phi^2(u),$$

where ∂_v is directional derivative in the direction of the vector v in the linear space T_pM. Since Φ is 1-homogeneous in T_pM, from the Euler identity we know that

$$g_u(u, v) = \partial_v(\Phi^2/2).$$

If $U \subseteq M$ is an orientable domain, then there exists a smooth map $J : SU \to SU$ (here SU is the unit tangent bundle of (U, Φ)) satisfying

$$g_{J(v)}(J(v), v) = 0, \qquad v \in SU. \tag{51}$$

Indeed, by orientability we can choose smoothly for each $v \in SU$ a unit covector $v_\perp \in S^*U$ in the annihilator of v. Then we take $J(v) = \mathcal{L}(v_\perp)$, where $\mathcal{L}$ is the Legendre transform (see Section 2). Now (51) follows by a standard argument in convex analysis.

The following lemma asserts the existence of "parallel geodesic coordinates" in any direction in SM. Every time we work in local coordinates, we will, as we may, assume that for every p, q in the coordinate neighbourhood, the quantities $d(p, q), d(q, p)$ and $|p - q|$ (where the latter stands for Euclidean distance in coordinates), are all comparable up to some multiplicative factor depending only on the neighbourhood. For a subset $A \subseteq M$ and a point $x \in M$ we write $d(x, A) = \inf_{y \in A} d(x, y)$ and $d(A, x) = \inf_{y \in A} d(y, x)$.

Lemma 7.1. *("Existence of geodesic parallel coordinates") Let (M, Φ) be a Finsler surface. Then for every $p \in M$ there exist $s_0, t_0 > 0$ and a neighbourhood $U \ni p$ with the following property: for every $v \in SU$ there exists a coordinate chart $\psi_v : (-s_0, s_0) \times (-t_0, t_0) \to U_v \subseteq M$, whose image U_v contains U and such that*

$$\psi_v(s, 0) = \gamma_v(s), \qquad s \in (-s_0, s_0),$$

and

$$d_M(\psi_v(s, t), \gamma_v) = d_M(\psi_v(s, t), \gamma_v(s)) = -t, \qquad s \in (-s_0, s_0), t \in (-t_0, 0), \tag{52}$$
$$d_M(\gamma_v, \psi_v(s, t)) = d_M(\gamma_v(s), \psi_v(s, t)) = t, \qquad s \in (-s_0, s_0), t \in (0, t_0). \tag{53}$$

In particular, for all $s \in (-s_0, s_0)$, the curve $\psi_v(s, \cdot)$ is a forward-minimizing geodesic. Moreover, the transition maps $\{\psi_u^{-1} \circ \psi_v\}_{u,v \in SU}$ all have C^3 norms bounded by a constant indepenent of u, v.

Proof. Let U be a strongly convex neighbourhood of p, and let J be the map discussed above. Possibly after shrinking U, there exists a rectangle $Q = (-s_0, s_0) \times (-t_0, t_0) \subseteq \mathbb{R}^2$ such that for every $v \in SU$, the map

$$\psi_v : Q \to M, \qquad \psi_v(s, t) = \pi \circ \phi_t \circ J \circ \phi_s(v),$$

where ϕ denotes the geodesic flow on TM, is defined and smooth on all of Q. Note that for all $s \in (-s_0, s_0)$,

$$\psi_v(s, 0) = \gamma_v(s).$$

The differential of ψ_v is nonsingular at $(0, 0)$. In fact,

$$d\psi_v(0, 0) = v \otimes ds + Jv \otimes dt.$$

By the inverse function theorem, possibly after shrinking U and decreasing s_0 and t_0, we may assume that the map ψ_v is bijective from Q to $U_v := \varphi_v(Q)$, for every $v \in TU$. Moreover, we may assume that both ψ_v and ψ_v^{-1} have C^3 norm bounded uniformly over all $v \in TU$, where the C^3 norm is taken with respect to some fixed coordinate chart on U. This also implies that the image U_v contains a ball of radius $r_0 = r_0(p, t_0, s_0)$ in this coordinate chart; thus, by taking U smaller, we may assume that $U \subseteq U_v$ for every $v \in TU$.

We prove the second equality (53). First, by the triangle inequality, if we take t_0 small enough, then for any $s \in (-s_0/2, s_0/2)$ and $t \in (0, t_0)$, the intersection of γ_v with any backward ball of radius t_0 centered at $\psi_v(s,t)$ is contained in $\gamma_v|_{(-s_0,s_0)}$. Therefore, as the distance from γ_v to $\psi_v(s,t)$ is at most t_0, it must be attained at a point $\gamma_v(\hat{s})$, for some $\hat{s} \in (-s_0, s_0)$.

By (51) and the first variation formula (see [2, Section 5.1]), the geodesic from $\gamma_v(\hat{s})$ to $\psi_v(s,t)$ has initial velocity $J(\dot{\gamma}_v(\hat{s}))$ (here we use the fact that $t_0 > 0$, because then $J(\dot{\gamma}_v(\hat{s}))$ points at the side of γ_v where $\psi_v(s,t)$ lies).

By the definition of ψ_v we have $\psi_v(\hat{s}, \hat{t}) = \psi_v(s,t)$ for some $\hat{t} \in (0, t_0)$, but since ψ_v is bijective we conclude that $\hat{s} = s$ and $\hat{t} = t$, i.e., the closest point to $\psi_v(s,t)$ is $\gamma_v(s)$ and the distance is t. This proves the second equation, after replacing s_0 by $s_0/2$. The proof of (52) is similar. □

Let us recall the formulation of Lemma 3.3:

Lemma 7.2 ("Disjoint geodesics that are close to each other at one point have similar tangents"). *Let $p \in M$. Then there exist $C, c > 0, \sigma_0 \in (0,1)$ and a coordinate chart $\psi : V \to U$, where V is an open subset of $\mathbb{R}^2$ and U is a neighbourhood of p, such that the following holds.*

Suppose that $0 < \sigma \leq \sigma_0$ and $\gamma_1, \gamma_2 : (-\sigma, \sigma) \to U$ are two disjoint, unit-speed, forward geodesics. Set

$$\eta_i = \psi^{-1} \circ \gamma_i, \tag{54}$$

and assume that $|\dot{\eta}_1(0) - \dot{\eta}_2(0)| < c$. Then,

$$|\dot{\eta}_1(0) - \dot{\eta}_2(0)| \leq \frac{C}{\sigma} \cdot |\eta_1(0) - \eta_2(0)|. \tag{55}$$

For the proof of Lemma 7.2 we require the following little lemma:

Lemma 7.3. *Let $0 < \delta \leq 1, A > 0$. Let $y : (-\delta, \delta) \to \mathbb{R}$ be a positive smooth function satisfying $|y''| \leq A(|y| + |y'|)$. Then $|y'(0)| \leq By(0)/\delta$, where B is a constant depending on A alone.*

Proof of Lemma 7.2. There is no loss of generality in assuming that $y(0) = 1$ and $y'(0) \leq 0$. Since $y(\delta/2)$ is positive, there exists $x \in (0, \delta/2)$ where $y'(x) \geq -2/\delta$. Let x_0 be the minimal such point. Then for all $x \in [0, x_0]$ we have $y'(x) \leq -2/\delta$ and $0 < y(x) \leq 1$. In particular, $|y(x)| \leq |y'(x)|$, and by our assumption this implies that $y''(x) \leq 2A|y'(x)| = -2Ay'(x)$ for $x \in [0, x_0]$. It follows that $e^{2Ax}y'(x)$ is decreasing for $x \in [0, x_0]$. Thus

$$0 \geq y'(0) \geq y'(x_0)e^{2Ax_0} \geq -2e^{2A}/\delta$$

as desired. □

Proof. Let $p \in M$, and let U be the neighbourhood of p supplied by Lemma 7.1. We need to provide a coordinate chart for which the conclusion of the lemma holds true, and we arbitrarily choose any of the ψ_u supplied by Lemma 7.1. Note that U is contained in the image of ψ_u. We use the notation $A \lesssim B$ to indicate $A \leq CB$ where $C > 0$ is a constant depending only on the manifold M and the point p. We will set $\sigma_0 = \min\{s_0, t_0, r_0, 1\}$ for s_0 and t_0 from Lemma 7.1, and where $r_0 > 0$ will be described below, and will be independent of the choice of the geodesics γ_1 and γ_2.

Let γ_1, γ_2 be two disjoint geodesics contained in U, defined on some symmetric interval $(-\sigma, \sigma) \subseteq \mathbb{R}$. We need to find $C, c > 0$, independent of γ_1 and γ_2, such that (55) holds true assuming $|\dot{\eta}_1(0) - \dot{\eta}_2(0)| < c$ and assuming $\sigma \leq \sigma_0$, where η_i is defined in (54).

Recall that the transition maps $\psi_u^{-1} \circ \psi_v$ from the conclusion of Lemma 7.1 all have C^3 norms uniformly bounded by a constant. If we define η_i via (54) with $\psi = \psi_u$ or with $\psi = \psi_v$, we may obtain different curves in $\mathbb{R}^2$. However, if we switch from ψ_u to ψ_v then both expressions

$$|\eta_1(0) - \eta_2(0)| \qquad \text{and} \qquad |\dot{\eta}_1(0) - \dot{\eta}_2(0)|$$

can change at most by a multiplicative constant independent of u and v. We may therefore switch to *any* coordinate chart ψ_v for any $v \in SU$, and find suitable constants C, c with respect to the new chart. We may even choose a chart that depends on the geodesics γ_1 and γ_2.

We choose to work with the chart $\psi := \psi_{\dot{\gamma}_1(0)}$, and define η_i as in (54). Then

$$\eta_1(t) = (t, 0) \qquad t \in (-\sigma, \sigma).$$

Thus, if we write

$$\eta_2(t) = (x(t), y(t)), \qquad t \in (-\sigma, \sigma),$$

then the inequality (55) that we need to prove reads

$$|(1,0) - (x'(0), y'(0))| \leq C \cdot \sigma^{-1} |(x(0), y(0))|, \tag{56}$$

assuming that $|(1,0) - (x'(0), y'(0))|$ is less than some positive constant and that $\sigma \leq \sigma_0$. To prove this, we first claim that if $|(1,0) - (x'(0), y'(0))|$ is less than some constant then

$$|(1,0) - (x'(t), y'(t))| \lesssim |y(t)| + |y'(t)|, \qquad t \in (-\sigma, \sigma). \tag{57}$$

Let us prove this claim. Fix $t \in (-\sigma, \sigma)$, and let $\Phi_0 = \Phi|_{(x(t),0)}$. That is, Φ_0 is the Finsler norm Φ, evaluated at $(x(t), 0)$ and viewed as a norm on $\mathbb{R}^2$ via the identification $TU \cong TV \cong V \times \mathbb{R}^2$. Note that since η_2 is unit-speed, and since Φ is locally-Lipschitz,

$$|\Phi_0(x'(t), y'(t)) - 1| = |\Phi|_{(x(t),0)}(x'(t), y'(t)) - \Phi|_{(x(t),y(t))}(x'(t), y'(t))| \lesssim |y(t)|. \tag{58}$$

On a certain neighbourhood of the point $(1,0)$ in $\mathbb{R}^2$, the map $(v^1, v^2) \mapsto (\Phi_0(v^1, v^2), v^2)$ is bi-Lipschitz. If $|(x'(0), y'(0)) - (1,0)|$ is less than some constant, then the point $(x'(0), y'(0))$

lies in this neighbourhood, and if $\sigma \leq \sigma_0$ and σ_0 is smaller than some constant $r_0 > 0$, then $(x'(t), y'(t))$ lies in this neighbourhood for all $t \in (-\sigma, \sigma)$. Therefore

$$|(x'(t), y'(t)) - (1,0)| \lesssim |\Phi_0(x'(t), y'(t)) - \Phi_0(1,0)| + |y'(t)| = |\Phi_0(x'(t), y'(t)) - 1| + |y'(t)|, \tag{59}$$

and (57) follows from (58) and (59). To prove (56) and complete the proof of the lemma, all that remains is to show that

$$|y'(0)| \lesssim |y(0)|/\sigma. \tag{60}$$

The curves $\eta_1(t) = (t, 0)$ and $\eta_2(t) = (x(t), y(t))$ are unit-speed geodesics, so by Lipschitz continuity of the Christoffel symbols in our coordinate chart,

$$|y''| \lesssim |y| + |(x', y') - (1,0)| \lesssim |y| + |y'|, \tag{61}$$

where the second inequality follows from (57). Now, if $|\eta_1(0) - \eta_2(0)| \geq c'\sigma$, then we get (56) immediately since the left hand side is bounded by a constant. Thus, we may assume that

$$|\eta_1(0) - \eta_2(0)| \leq c\sigma. \tag{62}$$

If this c is taken small enough, then (62) implies that $y(t)$ does not change sign when $t \in (-\sigma/2, \sigma/2)$, because otherwise η_2 would intersect the x axis within a distance of less than σ from the origin, which would imply that η_1 and η_2 intersect. Assume without loss of generality that $y(t)$ is positive for all $t \in (-\sigma/2, \sigma/2)$. Then (60) follows from (61) and Lemma 7.3. □

References

[1] Bakry, D., Gentil, I. & Ledoux, M, *Analysis and geometry of Markov diffusion operators.* Springer, 2014.

[2] Bao, D., Chern, S. & Shen, Z., *An introduction to Riemann-Finsler geometry.* Springer, 2000.

[3] Bucataru, I. & Muzsnay, Z., *Projective metrizability and formal integrability.* SIGMA Symmetry, Integrability And Geometry. Methods And Applications. 7 (2011), Paper 114.

[4] Burago, Y. & Zalgaller, V. Geometric inequalities. Springer, 1988.

[5] Cavalletti, F. & Mondino, A. *Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds.* Geom. Topol., 21, (2017), 603–645.

[6] Cavalletti, F. & Mondino, A., *A review of Lorentzian synthetic theory of timelike Ricci curvature bound.* arXiv:2204.13330

[7] Cheeger, J. & Ebin, D., *Comparison theorems in Riemannian geometry.* AMS Chelsea Publishing, Providence, RI,2008.

[8] Cordero-Erausquin, D., McCann, R. & Schmuckenschläger, M., *A Riemannian interpolation inequality á la Borell, Brascamp and Lieb.* Invent. Math., 146, (2001), 219-257.

[9] Crampin, M. & Mestdag, T., *A class of Finsler surfaces whose geodesics are circles.* Publ. Math. Debrecen, 84, (2014), 3–16.

[10] Eilat, M. & Klartag, B., *Rigidity of Riemannian embeddings of discrete metric spaces.* Invent. Math., 226, (2021), 349–391.

[11] Evans, L. C. & Gangbo, W., *Differential equations methods for the Monge-Kantorovich mass transfer problem.* Mem. Amer. Math. Soc., Vol. 137, no. 653, (1999).

[12] Evans, L. & Gariepy, R., *Measure theory and fine properties of functions.* CRC Press, 2015.

[13] Feldman, M. & McCann, R. *Monge's transport problem on a Riemannian manifold.* Trans. Amer. Math. Soc. 354, (2002), 1667–1697.

[14] Gardner, R. *The Brunn-Minkowski inequality.* Bull. Amer. Math. Soc., 39, (2002), 355–405.

[15] Izumiya, S., *Horospherical geometry in the hyperbolic space.* Noncommutativity And Singularities, 55, (2009), 31–49.

[16] Klartag, B., *Needle decompositions in Riemannian geometry.* Mem. Amer. Math. Soc., 249, no. 1180, 2017.

[17] Lott, J. & Vilanni, C., *Ricci curvature for metric-measure spaces via optimal transport.* Ann. of Math., Vol. 169, no. 3, (2009), 903–991.

[18] Matsumoto, M., *Every path space of dimension two is projectively related to a Finsler space.* Open Syst. Inf. Dyn., 3, (1995), 291–303.

[19] McCann, R. J., *Displacement convexity of Boltzmann's entropy characterizes the strong energy condition from general relativity.* Camb. J. Math., 8, no. 3, (2020), 609–681.

[20] Ohta, S., *Needle decompositions and isoperimetric inequalities in Finsler geometry.* J. Math. Soc. Japan, 70, (2018), 651–693.

[21] Schneider, R., *Convex Bodies: The Brunn–Minkowski Theory.* Cambridge University Press, 2013.

[22] Sturm, K.-T., *On the geometry of metric measure spaces. I.* Acta Math., Vol. 196, (2006), 65–131

[23] Sturm, K.-T., *On the geometry of metric measure spaces. II.* Acta Math., Vol. 196, (2006), 133–177.

Department of Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel.
e-mails: {rotem.assouline,boaz.klartag}@weizmann.ac.il