J. Eur. Math. Soc. 26, 985–1029 (2024) DOI 10.4171/JEMS/1339

Marcos Cossarini · Stéphane Sabourau

Minimal area of Finsler disks with minimizing geodesics

Received October 1, 2020; revised February 10, 2023

Abstract. We show that the Holmes–Thompson area of every Finsler disk of radius r whose interior geodesics are length-minimizing is at least $\frac{6}{\pi} r^2$. Furthermore, we construct examples showing that the inequality is sharp and observe that equality is attained by a non-rotationally-symmetric metric. This contrasts with Berger's conjecture in the Riemannian case, which asserts that the round hemisphere is extremal. To prove our theorem we discretize the Finsler metric using random geodesics. As an auxiliary result, we include a proof of the integral geometry formulas of Blaschke and Santaló for Finsler manifolds with almost no trapped geodesics.

Keywords. Finsler metrics, Holmes–Thompson area, Berger's conjecture, integral geometry, discrete geometry

1. Introduction

Isoembolic inequalities on Riemannian manifolds are curvature-free volume estimates in terms of the injectivity radius. The first sharp isoembolic inequality valid in all dimension is due to Berger [12] who showed that the volume of every closed Riemannian n-manifold M satisfies

$$\operatorname{vol}(M) \geq \alpha_n \left(\frac{\operatorname{inj}(M)}{\pi} \right)^n, \tag{1.1}$$

where α_n is the volume of the canonical n-sphere. Furthermore, equality holds if and only if M is isometric to a round sphere. The two-dimensional case was proved earlier in [10].

A long standing conjecture in Riemannian geometry also due to Berger asserts that every ball $B(r)$ of radius $r \leq \frac{1}{2} \operatorname{inj}(M)$ in a closed Riemannian n-manifold M satisfies

$$\operatorname{vol} B(r) \geq \frac{\alpha_n}{2} \left(\frac{2r}{\pi} \right)^n, \tag{1.2}$$

Marcos Cossarini: Univ Paris Est Creteil, CNRS, LAMA, 94010 Creteil; and
Univ Gustave Eiffel, LAMA, 77447 Marne-la-Vallée, France; marcos.cossarini@u-pec.fr

Stéphane Sabourau: Univ Paris Est Creteil, CNRS, LAMA, 94010 Creteil; and
Univ Gustave Eiffel, LAMA, 77447 Marne-la-Vallée, France; stephane.sabourau@u-pec.fr

Mathematics Subject Classification (2020): Primary 53C23; Secondary 53C60, 53C65

with equality if and only if $B(r)$ is isometric to a round hemisphere of (intrinsic) radius r. This can be viewed as a local version of the sharp isoembolic inequality (1.1). This conjecture is open even in the two-dimensional case where the previous inequality can be written as

$$\text{area}\, D(r) \geq \frac{8}{\pi} r^2. \tag{1.3}$$

A survey on isoembolic inequalities and Berger's conjecture is given in [26, §6]. A non-sharp volume estimate $\text{vol}\, B(r) \geq c_n r^n$ was established by Berger [10, 11] for $n = 2, 3$, and by Croke [22, Proposition 14] for every n. The conjecture (with a sharp constant) is satisfied for metrics of the form $\mathrm{d}s^2 = \mathrm{d}r^2 + f(r,\theta)^2\, \mathrm{d}\theta^2$ in polar coordinates when $n \geq 3$; see [23]. In [24], Croke also showed that the optimal inequality (1.3) holds true on average over all balls $B(r)$ of M. In the two-dimensional case, the best general estimate $\text{area}\, D(r) \geq \frac{8-\pi}{2} r^n$ can be found in [25]. The lower bound (1.3) on the area of $D(r)$ has recently been obtained in [19] by Chambers–Croke–Liokumovich–Wen under the stronger hypothesis that $r \leq \frac{1}{2} \operatorname{conv}(M)$, where $\operatorname{conv}(M)$ is the convexity radius of M. (This implies that $r \leq \frac{1}{4} \operatorname{inj}(M)$, since $\operatorname{conv}(M) \leq \frac{1}{2} \operatorname{inj}(M)$.) Note, however, that this stronger condition rules out the possibility that $D(r)$ is a hemisphere of intrinsinc radius r, which is the only expected equality case of (1.3).

The condition that $r \leq \frac{1}{2} \operatorname{inj}(M)$ in Berger's conjecture (1.2) can be relaxed by requiring instead that every interior geodesic in $B(r)$ is length-minimizing. The results of [10, 11] and [22, Proposition 14], for instance, still hold under this more general condition.

In this article, we consider the case of disks with a self-reverse Finsler metric whose interior geodesics are length-minimizing. (A precise definition of Finsler metrics and area can be found in Section 2.) It is natural to expect that inequality (1.3) holds in this setting. This is the case for isosystolic inequalities on the projective plane, where the canonical round metric minimizes the systolic area among both Riemannian and Finsler metrics; see [33, 34]. However, we show that the round hemisphere is not area-minimizing among Finsler metric disks of the same radius whose interior geodesics are length-minimizing. More precisely, we establish a sharp isoembolic inequality for Finsler metrics in the two-dimensional case under the assumption that every interior geodesic is length-minimizing. We observe that the extremal metric is not Riemannian and, surprisingly, not even rotationally symmetric.

Before stating our main result, let us introduce the following definition.

Definition 1.1. A *Finsler disk D of radius r with minimizing interior geodesics* is a disk with a Finsler metric such that

- every interior point of D is at distance less than r from a specified center point O;
- every point of ∂D is at distance exactly r from O;
- every interior geodesic of D is length-minimizing.

For instance, a ball of radius r on a complete Finsler plane with no conjugate points is a Finsler disk of radius r with minimizing interior geodesics.

The optimal version of Berger's conjecture for Finsler surfaces with self-reverse metric is given by the following result. We emphasize that we make no assumptions on the convexity radius.

Theorem 1.2. *Let D be a self-reverse Finsler metric disk D of radius r with minimizing interior geodesics. Then the Holmes–Thompson area of D satisfies*

$$\operatorname{area}(D) \geq \frac{6}{\pi} r^2.$$

Furthermore, the inequality is optimal.

The lower bound is attained by a non-smooth space consisting of a disk of radius r centered at the tip of the cone obtained by gluing together three copies of a quadrant of the ℓ^1-plane. (Recall that the ℓ^1-plane is the normed plane where unit balls have the least possible area, according to Mahler's theorem on convex bodies in the plane.) Note that this disk is not rotationally symmetric. In Section 11 we use Busemann's construction of projective metrics (developed in relation with Hilbert's fourth problem) to give another description of this non-smooth extremal metric. More precisely, we define a non-smooth projective metric on the plane where the disk of radius r centered at the origin has area $\frac{6}{\pi} r^2$. Then, we approximate this non-smooth projective metric by smooth projective metrics (which are therefore Finsler and have minimizing interior geodesics) where the area of the disk converges to $\frac{6}{\pi} r^2$, proving that the inequality of Theorem 1.2 is sharp.

Let us further comment on the result proved in [19] for Riemannian disks $D(r) \subseteq M$ of radius $r \leq \frac{1}{2} \operatorname{conv}(M)$. As previously mentioned, this excludes the possibility that $D(r)$ is a hemisphere of intrinsinc radius r. Still, the argument in [19] is valid for Finsler surfaces with self-reverse metric, except for the proof of their Lemma 2.1, which is purely Riemannian. Therefore, a self-reverse Finsler metric disk of radius r in which the distance function from each given point is convex along all geodesics satisfies (1.3). The extremal surfaces that we construct in this paper violate this inequality, but this is no contradiction because they have a vanishing convexity radius.

Instead of the Holmes–Thompson area, one could consider the Busemann–Hausdorff area, which, in general, is bounded below by the former; see [29]. However, the Busemann–Hausdorff area of the extremal metric in Theorem 1.2 is equal to $\frac{3}{4}\pi r^2$, which is greater than the area of the round hemisphere of intrinsic radius r that is conjectured to be minimal.

The proof of Theorem 1.2 and the construction of extremal and almost extremal metrics occupy the whole article. The approach, based on a discretization of the metric (cf. [20]), is fairly robust and new in this context.

The article is organized as follows.

In Section 2, we recall the notions of Finsler manifolds, their Holmes–Thompson measure, and their geodesics described from the Hamiltonian point of view.

In Section 3, we go over the standard proofs of the integral geometry formulas of Blaschke and Santaló, showing that they are valid for Finsler manifolds with almost no trapped geodesics. In the case of a disk as in Theorem 1.2, the formulas say that the length of a curve in the disk is proportional to the expected number of intersections with a random geodesic, and the area of a region is proportional to the expected length of the intersection with a random geodesic.

In Section 4, we introduce the notion of a quasi wall system on a surface, generalizing the wall systems studied in [20]. A quasi wall system on a surface is a 1-dimensional submanifold satisfying certain conditions. It determines a discrete metric, according to which the length of a curve is its number of intersections with the quasi wall system, and the area of the surface is the number of self-intersections of the quasi wall system. We show how to approximate a self-reverse Finsler metric with minimizing geodesics by a quasi wall system consisting of random geodesics. To prove the approximation properties we use the integral geometry formulas to compute the expected values of discrete length and area, and then we apply the law of large numbers.

In Section 5, we use this approximation result to show that Theorem 1.2 follows from an analogous theorem on simple discrete metric disks.

Sections 6–9 are devoted to the proof of this discrete theorem. The proof is based on identifying certain configurations on a quasi wall system and operating on these configurations in order to transform a simple discrete disk into a new one of less area. When the disk has minimum area, none of these configurations is present, and this implies that the quasi wall system is of a special kind where we can compute a lower bound for the area.

In Section 10, we construct a simple discrete disk of minimal discrete area and show that it is unique up to isotopy.

In Section 11, we use Busemann's construction of projective metrics to obtain continuous versions of our discrete area-minimizing disk.

Finally, Section 12 is an appendix where we show that on a Finsler surface with boundary, distance-realizing curves are C^1.

2. Finsler metrics and Holmes–Thompson volume

In this section, we recall basic definitions of Finsler geometry.

2.1. Finsler metrics

Let us recall the definition of a Finsler metric.

Definition 2.1. A *Finsler metric* on a smooth manifold M is a continuous function $F : TM \to [0, +\infty)$ on the tangent bundle TM of M satisfying the following properties (here, $F_x := F|_{T_x M}$ for short):

(1) *Positive homogeneity:* $F_x(tv) = t\, F_x(v)$ for all $v \in T_x M$ and $t \geq 0$.

(2) *Subadditivity:* $F_x(v + w) \leq F_x(v) + F_x(w)$ for all $v, w \in T_x M$.

(3) *Positive definiteness:* $F_x(v) > 0$ for every nonzero $v \in T_x M$.

(4) *Smoothness:* F is smooth outside the zero section.

(5) *Strong convexity:* for any two linearly independent vectors $v, w \in T_x M$, the Hessian value $q_v(w) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\big|_{t=0} F_x(v + tw)$ is strictly positive.

Additionally, a Finsler metric F may or be not be

(6) *Self-reverse:* $F_x(-v) = F_x(v)$ for every $v \in T_x M$.

Equivalently, one could define a Finsler metric by replacing (3) and (5) with the condition that for every nonzero vector $v \in TM$, the Hessian of F^2 at v is positive definite; see [21].

In each tangent space $T_x M$, the unit ball and unit sphere determined by the norm F_x are

$$B_x M = \{v \in T_x M \mid F_x(v) \leq 1\}, \quad U_x M = \{v \in T_x M \mid F_x(v) = 1\}.$$

Similarly, in the cotangent space $T_x^* M$, the norm F_x^* dual to F_x determines the unit coball $B_x^* M$ and the unit cosphere $U_x^* M$.

Remark 2.2. To handle technical details in case M has nonempty boundary, we extend the metric F to a manifold $M^+ \supseteq M$, of the same dimension as M but without boundary.

2.2. *Length, geodesics and distance-realizing arcs*

Definition 2.3. Let M be a manifold with a Finsler metric F. The *length* of a piecewise-C^1 curve $\gamma : I \to M$ is defined as the integral of its *speed* $F(\gamma'(t))$, that is,

$$\operatorname{length}(\gamma) = \int_I F(\gamma'(t))\,\mathrm{d}t \tag{2.1}$$

and the *distance* $d_F(x, y)$ between two points x and y in M is the infimum length of a curve γ in M joining x to y.

A *distance-realizing curve* is a curve $\gamma : I \to M$ such that

$$d_F(\gamma(t), \gamma(t')) = t' - t$$

for all $t < t'$.

A *geodesic* of M is a smooth, unit-speed curve $\gamma : I \to M$ that is extremal for the length functional. In case M has boundary, the extremality is defined by considering variations in M^+; see Remark 2.2. Thus the geodesics of M are the geodesics of M^+ that are contained in M. Equivalently, the geodesics of M are the unit-speed curves that satisfy the Euler–Lagrange equation for the length functional; see Definition 2.6 below for an explicit equation in terms of momentum.

In a compact connected Finsler manifold, any two points can be joined by a distance-realizing arc.[2] A distance-realizing arc contained in the interior of M is necessarily

[2] A proof for more general, complete self-reverse metrics is given [32, §1.12]; see also [36, Theorem 9.1] for directed metrics.

a geodesic and is therefore smooth. However, a distance-realizing arc of M does not necessarily lie in the interior of M, even if its endpoints do. Still, if the manifold is two-dimensional, then every distance-realizing arc is C^1 and has unit speed; see Theorem 12.1. Thus, in a compact Finsler surface, any two points x, y can be joined by a C^1 arc of length $d_F(x, y)$.

2.3. Symplectic structure on the cotangent bundle

Recall also some definitions about the geodesic flow of a Finsler manifold from the Hamiltonian viewpoint; see [5, 8, Chaps. 7–9] and [21].

Definition 2.4. Let M be a manifold. The *tautological* 1*-form* α_M on T^*M is defined as

$$\alpha_M|_\xi(V) = \xi(\mathrm{d}\pi_\xi(V))$$

for all $\xi \in T^*M$ and $V \in T_\xi T^*M$, where $\pi : T^*M \to M$ is the canonical projection. The *standard symplectic form* ω_M on T^*M is given by

$$\omega_M = \mathrm{d}\alpha_M.$$

Using canonical coordinates (x_i, ξ_i) on T^*M, these forms can be expressed as

$$\alpha_M = \sum_i \xi_i \mathrm{d}x_i, \quad \omega_M = \sum_i \mathrm{d}\xi_i \wedge \mathrm{d}x_i. \tag{2.2}$$

Definition 2.5. Let (M, F) be a Finsler manifold. The *Legendre map*

$$\mathcal{L} : UM \to U^*M$$

is defined as follows: the image of a unit vector $v \in U_xM$ is the unique unit covector $\xi \in U_x^*M$ such that $\xi(v) = 1$. Since F is strongly convex, the Legendre map is a diffeomorphism. Its inverse is the Legendre map associated to the dual metric F^* on T^*M, which is also strongly convex. The unit covectors will also be referred to as *momentums*. The *Hamiltonian lift* of a unit-speed curve γ in M is the curve $t \mapsto \mathcal{L}(\gamma'(t))$ in U^*M.

Definition 2.6. The *cogeodesic vector field* of a Finsler manifold M is the vector field Z on U^*M given by the equations

$$\iota_Z(\omega_M|_{U^*M}) = 0, \quad \iota_Z(\alpha_M) = 1,$$

where ι_Z is the operator that contracts a differential form with the vector field Z. The integral curves of Z are the Hamiltonian lifts of the geodesics in M; see [21].

It follows from the Cartan formula that the forms α_M and ω_M restricted to U^*M are invariant under the cogeodesic flow.

2.4. Holmes–Thompson volume

We will consider the following notion of volume.

Definition 2.7. The *Holmes–Thompson volume* of a Finsler n-manifold M is defined as the symplectic volume of its unit coball bundle $B^*M \subseteq T^*M$, divided by the volume ϵ_n of the Euclidean unit ball in $\mathbb{R}^n$. That is,

$$\mathrm{vol}(M) = \frac{1}{\epsilon_n} \int_{B^*M} \frac{1}{n!} \omega_M^n, \tag{2.3}$$

where ω_M is the standard symplectic form on T^*M and $\frac{1}{n!}\omega_M^n = \frac{1}{n!}\,\omega_M \wedge \cdots \wedge \omega_M$ is the corresponding volume form. Equivalently, the Holmes–Thompson volume is given as an integral over the unit sphere bundle by the formula

$$\mathrm{vol}(M) = \frac{1}{\epsilon_n n!} \int_{U^*M} \alpha_M \wedge \omega_M^{n-1}. \tag{2.4}$$

The factor $\frac{1}{\epsilon_n}$ ensures that for Riemannian metrics, the Holmes–Thompson definition of volume agrees with the conventional Riemannian definition.

3. Integral geometry in Finsler manifolds with almost no trapped geodesics

The goal of this section is to present versions of two classical formulas in integral geometry, due to Blaschke [15] and Santaló [39, 40], which are in turn generalizations to manifolds of the classical Crofton formulas on the Euclidean plane. In [6], Blaschke's formula is proved for Finsler manifolds whose space of geodesics is a smooth manifold. Here, we give slightly more general versions which hold for Finsler manifolds with almost no trapped geodesics (and, in particular, for compact Finsler manifolds with minimizing interior geodesics). The proofs mimick those given by Blaschke, Santaló, and Álvarez-Paiva–Berck. However, we give them in full in order to provide additional details and introduce the few extra steps needed for the generalization.

Definition 3.1. Let M be a Finsler n-manifold with nonempty boundary. A *traversing geodesic* of M is a maximal geodesic $\gamma : [0, \ell(\gamma)] \to M$ which does not intersect ∂M, except at its endpoints where it meets the boundary transversely. The Finsler manifold M has *almost no trapped geodesics* if for almost every unit tangent vector $v \in UM$, the maximal geodesic γ_v defined by $\gamma_v'(0) = v$ reaches the boundary of M in the future and in the past, that is, $\gamma_v(t) \in \partial M$ for some $t \geq 0$ and some $t \leq 0$.

For instance, a compact Finsler manifold with minimizing interior geodesics has almost no trapped geodesics. Another example is obtained by taking a closed Finsler manifold with ergodic geodesic flow and removing a smoothly bounded nonempty open set.

As we will explain below, the space Γ of traversing geodesics of M is a $(2n-2)$-dimensional manifold admitting a natural symplectic structure, whose corresponding natural volume measure is denoted by μ_Γ; see Definition 3.6.

Theorem 3.2 (Blaschke's formula). *Let M be a Finsler n-manifold with almost no trapped geodesics. Then the Holmes–Thompson volume of an immersed hypersurface $N \subseteq M$ is equal to*

$$\mathrm{vol}_{n-1}(N) = \frac{1}{2\epsilon_{n-1}} \int_{\gamma\in\Gamma} \#(\gamma \cap N)\, \mathrm{d}\mu_\Gamma(\gamma), \tag{3.1}$$

where $\#(\gamma \cap N)$ is the number of times that γ intersects N.

Similarly, the Holmes–Thompson volume of a cooriented immersed hypersurface $N \subseteq M$ is equal to

$$\mathrm{vol}_{n-1}(N) = \frac{1}{\epsilon_{n-1}} \int_{\gamma\in\Gamma} \#(\gamma \cap^+ N)\, \mathrm{d}\mu_\Gamma(\gamma), \tag{3.2}$$

where $\#(\gamma \cap^+ N)$ is the number of times that γ intersects N transversely in the positive direction.

In (3.1), we can restrict the integral to geodesics $\gamma \in \Gamma$ which are transverse to the hypersurface N since the geodesics $\gamma \in \Gamma$ which are tangent to N form a subset of zero measure; see Proposition 3.7 (3).

Remark 3.3. Since every traversing geodesic intersects ∂M positively exactly once, we deduce from (3.2) that the total measure of the space Γ is

$$\mu_\Gamma(\Gamma) = \epsilon_{n-1}\, \mathrm{vol}_{n-1}(\partial M).$$

In particular, if M is compact, then $\mu_\Gamma(\Gamma) < \infty$.

Theorem 3.4 (Santaló's formula). *Let M be a Finsler n-manifold with almost no trapped geodesics. Then the Holmes–Thompson volume of a smoothly bounded domain $D \subseteq M$ is equal to*

$$\mathrm{vol}_n(D) = \frac{1}{n\epsilon_n} \int_{\gamma\in\Gamma} \mathrm{length}(\gamma \cap D)\, \mathrm{d}\mu_\Gamma(\gamma). \tag{3.3}$$

In the case of Finsler surfaces with self-reverse metric, the Blaschke and Santaló formulas specialize as follows.

Corollary 3.5. *Let M be a self-reverse Finsler metric surface with almost no trapped geodesics. Then the length of any immersed curve c in M is*

$$\mathrm{length}(c) = \frac{1}{4} \int_{\gamma\in\Gamma} \#(\gamma \cap c)\, \mathrm{d}\mu_\Gamma(\gamma), \tag{3.4}$$

and the Holmes–Thompson area of any smoothly bounded domain $D \subseteq M$ is

$$\mathrm{area}(D) = \frac{1}{2\pi} \int_{\gamma\in\Gamma} \mathrm{length}(\gamma \cap D)\, \mathrm{d}\mu_\Gamma(\gamma) \tag{3.5}$$

$$= \frac{1}{8\pi} \iint_{(\gamma_0,\gamma_1)\in\Gamma\times\Gamma} \#(\gamma_0 \cap \gamma_1 \cap D)\, \mathrm{d}\mu_\Gamma(\gamma_0)\, \mathrm{d}\mu_\Gamma(\gamma_1). \tag{3.6}$$

Equation (3.6), obtained from (3.5) and (3.4), will be called the *Santaló+Blaschke formula*. In deducing this formula, we use the hypothesis that the metric is self-reverse when we equate the length of a geodesic with its Holmes–Thompson measure. In general, the Holmes–Thompson measure of a curve is the average of its forward and backward lengths.

The rest of this section is dedicated to describing the symplectic structure on Γ and proving Theorems 3.2 and 3.4.

3.1. *Symplectic manifold of traversing geodesics*

Let M be a Finsler n-manifold. Recall that Γ is the space of traversing geodesics of M. The space Γ is a $(2n-2)$-dimensional manifold parameterized by the initial vectors $\gamma'(0) \in UM|_{\partial M}$ of the geodesics $\gamma : [0, \ell(\gamma)] \to M$ of Γ. Note that the length $\ell(\gamma)$ depends smoothly on $\gamma \in \Gamma$.

Definition 3.6. Define the open subset of U^*M,

$$U^*_\Gamma M = \{\mathcal{L}(\gamma'(t)) \in U^*M \mid \gamma \in \Gamma,\ t \in [0, \ell(\gamma)]\},$$

consisting of the momentums of the traversing geodesics of M. Note that $U^*_\Gamma M$ is a Z-invariant open subset of U^*M.

Consider the surjective submersion

$$\pi_\Gamma : U^*_\Gamma M \to \Gamma \tag{3.7}$$

taking any momentum $\xi \in U^*_\Gamma M$ to the geodesic $\gamma \in \Gamma$ it generates. The fibers of π_Γ are the Z-orbits corresponding to the traversing geodesics.

There exists a unique 2-form ω_Γ on Γ such that

$$\pi^*_\Gamma \omega_\Gamma = \omega_M|_{U^*_\Gamma M}.$$

This follows from the invariance of the 2-form $\omega_M|_{U^*_\Gamma M}$ under the cogeodesic flow, and the fact that this form vanishes in the direction of Z according to Definition 2.6. (See also [21] for details, or [1, §4.3] for a general account of symplectic reduction.) The form ω_Γ is symplectic, thus it determines on Γ a smooth volume measure μ_Γ given by

$$d\mu_\Gamma = \frac{1}{(n-1)!}|\omega^{n-1}_\Gamma|. \tag{3.8}$$

3.2. *Non-traversing geodesics are negligible*

We will need the following result in order to establish our versions of Blaschke's and Santaló's formulas. This feature is not required in the previous versions and necessitates the manifold having almost no trapped geodesics.

Recall that a subset A of a manifold X is *negligible* in X if the image of A in any local chart of X has zero measure.

Proposition 3.7. (1) *The complement of the open subset $U^*_\Gamma M \subseteq U^*M$ is negligible in U^*M.*

(2) *Given a hypersurface $H \subseteq U^*M$ transverse to Z, the complement of $H \cap U^*_\Gamma M$ is negligible in H.*

(3) *The set of geodesics $\gamma \in \Gamma$ tangent to an immersed hypersurface of M or passing through an immersed submanifold of M of codimension > 1 has zero measure in Γ.*

Proof. To avoid technical problems we extend the Finsler metric to an open manifold M^+; see Remark 2.2. This ensures the cogeodesic flow $(\xi, t) \mapsto Z^t(\xi)$ is defined on an open domain.

By definition of Γ, the complement $U^*M \setminus U^*_\Gamma M$ is formed by momentums of two types. First, momentums of U^*M that correspond to geodesics of M with at least one end trapped in M. These momentums form a negligible set since M has almost no trapped geodesics; see Definition 3.1. Second, momentums of U^*M corresponding to geodesics tangent to ∂M. These momentums are of the form $Z^t(\xi)$, where Z^t is the cogeodesic flow and ξ is the Legendre image of a unit vector v tangent to ∂M. These unit vectors form a manifold $U\partial M$ of dimension $2n-3$. Thus, by Sard's theorem, the map from an open subset of $U\partial M \times \mathbb{R}$ to U^*M defined by $(v,t) \mapsto Z^t(\mathcal{L}(v))$ has negligible image in U^*M. Having considered both types of momentums, we conclude that the complement $U^*M \setminus U^*_\Gamma M$ is negligible in U^*M.

For the second point, simply observe that if A is a Z-invariant negligible subset of U^*M and H is a hypersurface of U^*M transverse to Z, then $A \cap H$ is negligible in H. Apply this property to $A = U^*M \setminus U^*_\Gamma M$ to conclude.

The proof of the third point is similar to the proof of the first point and relies on Sard's theorem. ∎

3.3. Manifold of positive momentums across a hypersurface

We will need the following notion in the proof of Blaschke's formula.

Definition 3.8. Let N be a cooriented embedded hypersurface in a Finsler manifold M. (For example, we can have $N = \partial M$ cooriented so that inwards-pointing vectors are positive.) Denote by $C^*N \subseteq U^*M|_N$ the *manifold of momentums crossing positively* the hypersurface N, that is, the momentums corresponding under the Legendre map to unit vectors transverse to N pointing in the positive direction according to the coorientation of N. Note that C^*N is an open subset of $U^*M|_N$ and therefore a differentiable manifold, with the structure of an open ball bundle over N.

Consider the *restriction map*

$$\rho_N : C^*N \to \mathrm{Int}(B^*N), \quad T^*_x M \ni \xi \mapsto \xi|_{T_x N},$$

to the interior $\mathrm{Int}(B^*N)$ of the unit coball bundle B^*N of N.

The following statement can be found in [6, Lemma 5.4]. For completeness we provide the details of the proof.

Lemma 3.9. *The space C^*N is a symplectic submanifold of T^*M and the restriction map*

$$\rho_N : (C^*N, \omega_M) \to (\mathrm{Int}(B^*N), \omega_N)$$

is a symplectomorphism. Thus,

$$\rho_N^* \, \omega_N = \omega_M|_{C^*N}.$$

Proof. Let $\xi \in C^*N$ with basepoint $x \in N$. By definition, the norm of ξ is 1, so the norm of its restriction ξ' to T_xN is at most 1. Furthermore, by strong convexity of F_x^*, the linear form ξ attains its maximum only at its Legendre-dual unit vector, which is positive and thus not contained in T_xN. Therefore, $\|\xi'\| < 1$ and the restriction map ρ_N takes values in $\mathrm{Int}(B^*N)$.

To see that ρ_N is a diffeomorphism, we employ local coordinates $(x_i)_{1 \le i \le n}$ in M so that the hypersurface N is given by the equation $x_n = 0$. Let $(x_i, v_i)_i$ and $(x_i, \xi_i)_i$ be the corresponding coordinates in TM and T^*M. In terms of these coordinates, the operator ρ_N acts by suppressing the last coefficient, that is, if $\xi = (\xi_i)_{1 \le i \le n}$, then $\xi' = (\xi_i)_{1 \le i \le n-1}$. Hence ρ_N is smooth.

To prove that ρ_N is bijective, consider a covector $\xi' = (\xi_i)_{1 \le i \le n-1} \in \mathrm{Int}(B_x^*N)$ and denote its norm $\lambda = \|\xi'\| < 1$. The covectors $\xi \in T_x^*M$ such that $\xi|_{T_xN} = \xi'$ are of the form $\xi^t = (\xi_1, \ldots, \xi_{n-1}, t)$ with $t \in \mathbb{R}$. Consider the function $t \mapsto \|\xi^t\|$, where $\|\cdot\|$ is the norm F_x^* on T_x^*M that is dual to F_x. This function is bounded below by λ, and by the Hahn–Banach theorem, this lower bound is attained at some $t_0 \in \mathbb{R}$. Furthermore, since the norm F_x^* is strongly convex, the set of values of t such that $\|\xi^t\| \le 1$ is a compact interval $[t_-, t_+]$ that contains t_0 in its interior, and $\|\xi^t\| = 1$ if and only if $t = t_\pm$. Thus we are left with two candidates $\xi^{t_\pm}$ that are the only unit covectors ξ whose restriction to T_xN is ξ'.

We claim that ξ^{t_+} is positive (and ξ^{t_-} is negative). That is, the vector in Legendre correspondence with ξ^{t_+} (i.e., the unit vector where ξ^{t_+} attains its norm) is positive. Indeed, when $t = t_0$, the covector ξ^t, as a function $B_xM \to \mathbb{R}$, is bounded above by λ. As t increases towards t_+, the coefficient ξ_n increases, and thus the values of $\xi^t(v)$ for v on the negative side decrease (hence they are $< \lambda$). Thus, any functional ξ^t with $t > t_0$, restricted to the ball B_xM, must attain its maximum value $\|\xi^t\|$ (which is $> \lambda$) on a positive vector, as required. This shows that ξ^t is positive if $t > t_0$ (and similarly ξ^t is negative if $t < t_0$). We conclude that ξ^{t_+} is the only positive unit covector ξ whose restriction to T_xN is ξ'. This proves that ρ_N is bijective. Additionally, t_+ depends smoothly on ξ' by the implicit function theorem. This finishes the proof that the restriction map $\rho_N : C^*N \to \mathrm{Int}(B^*N)$ is a diffeomorphism.

Let us show that $(\rho_N)^*\alpha_N = \alpha_M|_{C^*N}$. In canonical coordinates, the tautological 1-form α_M on T^*M is written as $\alpha_M = \sum_{i=1}^n \xi_i \, \mathrm{d}x_i$. In restricting to C^*N, the last term vanishes because $x_n = 0$ on N, thus the restricted form can be written as $\alpha_M|_{C^*N} = \sum_{i=1}^{n-1} \xi_i \, \mathrm{d}x_i$. On the other hand, the tautological 1-form of N is $\alpha_N = \sum_{i=1}^{n-1} \xi_i \, \mathrm{d}x_i$, and this expression is unchanged by the pullback $(\rho_N)^*$ since the map $\rho_N : C^*N \to \mathrm{Int}(B^*N)$ acts simply by suppressing the coordinate ξ_n. We conclude that $(\rho_N)^*\alpha_N = \alpha_M|_{C^*N}$.

Taking the exterior differential of this expression, we obtain $(\rho_N)^*\omega_N = \omega_M|_{C^*N}$. This implies that C^*N is a symplectic submanifold of T^*M. ∎

3.4. *Coarea formula and fiber integration*

In the proofs of Blaschke's and Santaló's formulas, we will need the following version of the coarea formula; see [28, (16.24.8)] (see also [30, Theorem 3.2.3] and [17, Theorem 5.5.8] when $n = m$).

Lemma 3.10. *Let $\pi : X \to Y$ be a submersion between two oriented manifolds of dimension n and m with $n \geq m$. Let α and β be differential forms on X and Y of degree $n-m$ and m. Then*

$$\int_X \alpha \wedge \pi^*\beta = \int_{y\in Y} \left(\int_{\pi^{-1}(y)} \alpha \right) \beta,$$

where $\pi^{-1}(y)$ is endowed with the orientation induced by π from the orientations of X and Y. In particular, for $n = m$ and $\alpha = 1$, we have

$$\int_X \pi^*\beta = \int_{y\in Y} \#(\pi^{-1}(y))\, \beta. \tag{3.9}$$

3.5. *Proof of the Blaschke formula*

We can now proceed to the proof of Blaschke's formula (3.1).

Proof of Theorem 3.2. We will follow the proof given in [6, Theorem 5.2] under the extra assumption that the space of oriented geodesics on M is a manifold.

The Blaschke formula (3.1) for a non-cooriented hypersurface N can be deduced from the cooriented version (3.2) by taking the cooriented double cover of N. Therefore it is sufficient to prove the latter formula. Furthermore, every immersed hypersurface can be decomposed into a disjoint union of embedded hypersurfaces up to a negligible set. Therefore it is sufficient to prove (3.2) for a cooriented embedded hypersurface N.

By definition of the Holmes–Thompson volume (see (2.3)), we have

$$\mathrm{vol}_{n-1}(N) = \frac{1}{\epsilon_{n-1}(n-1)!} \int_{B^*N} \omega_N^{n-1} = \frac{1}{\epsilon_{n-1}(n-1)!} \int_{C^*N} \omega_M^{n-1},$$

where the second equality follows from Lemma 3.9.

Now, apply Proposition 3.7 (2) with $H = C^*N \subseteq U^*M$. It follows that $C^*N \cap U^*_\Gamma M$ has full measure in C^*N. Thus,

$$\int_{C^*N} \omega_M^{n-1} = \int_{C^*N \cap U^*_\Gamma M} \omega_M^{n-1}.$$

Consider the map $\pi : C^*N \cap U^*_\Gamma M \to \Gamma$ taking a unit momentum of M based at N pointing in a positive direction (with respect to the coorientation of N) to the traversing geodesic it generates. Apply the fiber integration formula (3.9) to this map with $\beta = \omega_\Gamma^{n-1}$. This yields the relation

$$\int_{C^*N \cap U_\Gamma^* M} \omega_M^{n-1} = \int_{\gamma \in \Gamma} \#(\gamma \cap^+ N)\, \omega_\Gamma^{n-1},$$

where $\#(\gamma \cap^+ N)$ is the number of times that γ crosses N transversely in the positive sense (as determined by the coorientation of N). Taking into account the definition of μ_Γ by (3.8), we get Blaschke's formula. ∎

3.6. *Proof of the Santaló formula*

Let us prove Santaló's formula (3.3).

Proof of Theorem 3.4. Recall that $\omega_M = \pi_\Gamma^* \omega_\Gamma$ (see Definition 3.6), and $U_\Gamma^* M$ has full measure in U^*M (see Proposition 3.7 (1)). By (2.4), we have

$$\mathrm{vol}_n(D) = \frac{1}{\epsilon_n n!} \int_{U^*D \cap U_\Gamma^* M} \alpha_M \wedge \pi_\Gamma^* \omega_\Gamma^{n-1}.$$

By Lemma 3.10, integrating along the fibers of the submersion $\pi : U^*D \cap U_\Gamma^* M \to \Gamma$ induced by π_Γ (see (3.7)), we obtain

$$\mathrm{vol}_n(D) = \frac{1}{\epsilon_n n!} \int_{\gamma \in \Gamma} \left(\int_{\pi^{-1}(\gamma)} \alpha_M \right) \omega_\Gamma^{n-1}.$$

Since all the fibers $\pi^{-1}(\gamma) = \{\mathcal{L}(\gamma'(t)) \in U^*M \mid t \in [0, \ell(\gamma)]\} \cap U^*D$ are tangent to the cogeodesic vector field Z on U^*M and $\alpha_M(Z) = 1$, we derive

$$\int_{\pi^{-1}(\gamma)} \alpha_M = \mathrm{length}(\gamma \cap D).$$

Hence,

$$\mathrm{vol}_n(D) = \frac{1}{\epsilon_n\, n!} \int_{\gamma \in \Gamma} \mathrm{length}(\gamma \cap D) \omega_\Gamma^{n-1} = \frac{1}{\epsilon_n\, n} \int_{\gamma \in \Gamma} \mathrm{length}(\gamma \cap D)\, d\mu_\Gamma(\gamma).$$ ∎

4. Discretization of Finsler surfaces

The goal of this section is to describe a discretization of Finsler disks with minimizing interior geodesics into simple discrete metric disks. For this, we adapt the general scheme of discretization developed in [20] in relation with the filling area conjecture. The main novelty is that, in our case, the discrete geometry is described by a system of curves (wall system) made up of geodesics.

First, we need to fix some notation regarding intersections of maps.

Definition 4.1. The *intersections* of a map $f : X \to Y$ with a map $f' : X' \to Y$ lying in a subset $A \subseteq Y$ are the ordered pairs in the set

$$I_A(f, f') = \{(x, x') \in X \times X' \mid f(x) = f'(x') \in A\}.$$

The *number of intersections* between f and f' is defined as

$$\#(f \cap f') = \#I_Y(f, f'),$$

where $\#S$ denotes the cardinality of a set S.

Similarly, the *self-intersections* of a map $f : X \to Y$ lying in a subset $A \subseteq Y$ are the *unordered* pairs in the set

$$I_A(f) = \big\{\{x, x'\} \subseteq X \mid f(x) = f(x') \in A \text{ but } x \neq x'\big\},$$

and the *multiplicity* of a point $y \in Y$ as a self-intersection of f is the number $\#I_{\{y\}}(f)$. A self-intersection is *simple* if it has multiplicity 1.

Let us introduce the notion of a wall system on a disk; see [20].

Definition 4.2. A (smooth) *wall system* on a surface M is a 1-dimensional (smooth) immersed submanifold $\mathcal{W}$ satisfying the following conditions:

(1) the immersion map is proper (that is, the preimage of any compact subset of M is compact);

(2) $\mathcal{W}$ is transverse to the boundary ∂M and satisfies $\partial\mathcal{W} = \mathcal{W} \cap \partial M$;

(3) $\mathcal{W}$ is self-transverse and has only simple self-intersections;

(4) no self-intersections of $\mathcal{W}$ lie on the boundary ∂M.

As a technical remark, we note that the symbol $\mathcal{W}$ denotes the *immersion map*, not its image $\mathrm{Im}(\mathcal{W}) \subseteq M$ or its domain. The domain is a 1-manifold, i.e., a disjoint union of countably many intervals and circles. Hence the expression $\partial\mathcal{W} \subseteq \partial M$ involves an abuse of notation and actually means $\mathrm{Im}(\partial\mathcal{W}) \subseteq \partial M$, where $\partial\mathcal{W}$ is the restriction of the map $\mathcal{W}$ to the boundary of the domain of $\mathcal{W}$. The image of $\mathcal{W}$ will also be denoted $\mathcal{W}$. Thus, the expression $M \setminus \mathcal{W}$ denotes $M \setminus \mathrm{Im}(\mathcal{W})$.

Eventually we will need to relax the definition by dropping condition (4). In this case, we say that $\mathcal{W}$ is a *quasi wall system* on M.

The curves that form a (quasi) wall system are called its *walls*. Note that if the surface M is compact, then W consists of finitely many compact walls; each of these walls is either a loop that avoids the boundary or an arc that meets the boundary only at its two endpoints.

A quasi wall system $\mathcal{W}$ on a disk D is *simple* if its walls are arcs that have no self-intersections and that meet each other at most once.[3]

[3] Simple wall systems are also called pseudoline arrangements; see [20]. However, some authors (e.g., [31]) only consider complete pseudoline arrangements, which are those where every pair of walls crosses *exactly* once.

In this paper, every quasi wall system $\mathcal{W}$ is smooth unless we make it clear that it is piecewise smooth. In that case, the non-smooth points of $\mathcal{W}$ may not coincide with the self-intersection points of $\mathcal{W}$. Note that a piecewise smooth quasi wall system can be turned into a smooth quasi wall system by an isotopic deformation.

Example 4.3. Let D be the unit disk in the Euclidean plane. A wall system made of the horizontal and vertical diameters of D has area 1. A quasi wall system made of the three sides of an inscribed triangle of D has area $3/2$.

We will also need the following definitions regarding the geometry induced by a quasi wall system.

Definition 4.4. Every quasi wall system $\mathcal{W}$ on a compact surface M determines a *discrete length*

$$\operatorname{length}_{\mathcal{W}}(c) = \#(c \cap \mathcal{W}) \tag{4.1}$$

for curves c in M. That is, the length of a curve is the number of times it intersects the quasi wall system (counted with multiplicity). Every quasi wall system $\mathcal{W}$ also induces a pseudo-distance on $M \setminus \mathcal{W}$ defined by

$$d_{\mathcal{W}}(x, y) = \inf_{c} \operatorname{length}_{\mathcal{W}}(c)$$

where the infimum is taken over all paths of M joining x to y. We will refer to the pseudo-distance $d_{\mathcal{W}}$ on M as the *discrete distance* induced by $\mathcal{W}$ on D.

The *discrete area* of $(M, \mathcal{W})$ is the number of self-crossings of $\mathcal{W}$ contained in the interior of M plus half the number of self-crossings on the boundary. That is,

$$\operatorname{area}(M, \mathcal{W}) = \#I_{\operatorname{Int} M}(\mathcal{W}) + \tfrac{1}{2}\#I_{\partial M}(\mathcal{W}) = \#I_{\operatorname{Int} M + \frac{1}{2}\partial M}(\mathcal{W})$$

where $\#I_{\operatorname{Int} M + \frac{1}{2}\partial M}$ is just an abbreviation for $\#I_{\operatorname{Int} M} + \frac{1}{2}\#I_{\partial M}$.

Note that, if $\mathcal{W}$ consists of finitely many curves γ_i, then

$$\operatorname{area}(M, \mathcal{W}) = \sum_{i<j} \#I_{\operatorname{Int} M + \frac{1}{2}\partial M}(\gamma_i, \gamma_j) + \sum_{i} \#I_{\operatorname{Int} M + \frac{1}{2}\partial M}(\gamma_i). \tag{4.2}$$

When the quasi wall system is simple, the curves of $\mathcal{W}$ have no self-intersections and the second sum vanishes.

We will need the following result describing the intersection of two distance-realizing arcs of M. Recall that Γ is the space of traversing geodesics of M (i.e., geodesic arcs of M which do not intersect ∂M except at their endpoints, where they meet the boundary transversely).

Lemma 4.5. *Let M be a self-reverse Finsler metric disk with minimizing interior geodesics. Let $\gamma \in \Gamma$ be a traversing geodesic of M and let $[x, y]$ be a distance-realizing arc of M with endpoints x and y not lying in γ. Then*

$$\#(\gamma \cap [x, y]) = \begin{cases} 1 & \textit{if } \gamma \textit{ separates } x \textit{ and } y, \\ 0 & \textit{otherwise}. \end{cases}$$

Proof. By Theorem 12.1, the distance-realizing arc $[x, y]$ is C^1.

Suppose that the arcs γ and $[x, y]$ are tangent, either at an interior point of M or at an endpoint of γ in ∂M. In both cases, this implies that $[x, y]$ contains γ since the distance-realizing arc $[x, y]$ follows the geodesic flow in the interior of M and the endpoints x, y do not lie in γ. Now, since the interior geodesic γ is transverse to ∂M at its endpoints $\bar{x}$ and $\bar{y}$, the distance-realizing arc $[x, y]$ is not differentiable at $\bar{x}$ and $\bar{y}$. In particular, it is not C^1, which is absurd. Therefore, the arcs γ and $[x, y]$ may only have transverse intersections.

Suppose that the arcs γ and $[x, y]$ intersect at least twice, say at a and b (with a and b different from x and y). Since both arcs are distance-realizing curves, the subarcs $[a, b] \subseteq [x, y]$ and $\gamma_{ab} \subseteq \gamma$ joining a and b have the same length. Construct an arc α joining x and y by replacing the subarc $[a, b]$ of $[x, y]$ with the arc γ_{ab} of the same length. By construction, the arc α is a distance-realizing curve. But since the intersection between γ and $[x, y]$ is transverse, the arc α is not differentiable at a and b. In particular, it is not C^1, which is absurd. Therefore, the arcs γ and $[x, y]$ intersect at most once, and so exactly once if γ separates x and y.

Suppose now that γ does not separate x and y. Then the arc $[x, y]$ does not intersect γ. Otherwise, it would go from one side of γ to the other (recall that γ and $[x, y]$ have transverse intersection) and, because x and y are on the same side of γ, it would have to cross γ a second time, which is excluded. Therefore, the arcs γ and $[x, y]$ do no intersect if γ does not separate x and y. ∎

Let us compare the shortest paths for Finsler metrics and discrete metrics.

Definition 4.6. A quasi wall system is *geodesic* if its walls are geodesics.

Proposition 4.7. *Let M be a self-reverse Finsler metric disk with minimizing interior geodesics, and let $\mathcal{W}$ be a geodesic quasi wall system on M. Then every distance-realizing arc $[x, y]$ of M with endpoints x, y not lying in $\mathcal{W}$ is also length minimizing with respect to $\mathcal{W}$. Thus, for all $x, y \in M \setminus \mathcal{W}$, we have*

$$d_{\mathcal{W}}(x, y) = \operatorname{length}_{\mathcal{W}}([x, y]). \tag{4.3}$$

Proof. The quasi wall system $\mathcal{W}$ is made up of finitely many geodesics γ_i that are transverse to ∂M. By Lemma 4.5, the arc $[x, y]$ crosses only those geodesics γ_i that separate x from y, exactly once. Therefore, no curve from x to y can be shorter than $[x, y]$ with respect to $\mathcal{W}$. ∎

Before proceeding we derive a useful consequence of the last lemma.

Lemma 4.8. *Let M be a self-reverse Finsler metric disk with minimizing interior geodesics. Then*

$$d(x, y) \leq \tfrac{1}{2} \operatorname{length}(\partial M) \quad \text{for all } x, y \in M.$$

The same inequality holds if the distance and length are taken with respect to a geodesic quasi wall system $\mathcal{W}$, that is,

$$d_{\mathcal{W}}(x, y) \leq \tfrac{1}{2} \operatorname{length}_{\mathcal{W}}(\partial M) \quad \text{for all } x, y \in M \setminus \mathcal{W}.$$

Proof. Join the points $x, y \in M$ by a distance-realizing arc $[x, y]$. By Lemma 4.5, each traversing geodesic γ of M intersects $[x, y]$ at most once and meets ∂M exactly twice. Then the inequality $d(x, y) \leq \frac{1}{2}\,\mathrm{length}(\partial M)$ follows from Blaschke's formula (3.4) applied to $[x, y]$.

The claim regarding the geodesic quasi wall system $\mathcal{W}$ is proved in a similar way. By Proposition 4.7, the distance-realizing arc $[x, y]$ is also length-minimizing with respect to $\mathcal{W}$. Since each wall of $\mathcal{W}$ crosses $[x, y]$ at most once and meets ∂M exactly twice, we derive the desired second inequality from the definition of $\mathrm{length}_{\mathcal{W}}$; see (4.1). ∎

Simple wall systems can be used to discretize Finsler disks M with minimizing interior geodesics.

For any $a, b \in \mathbb{R}$ and $\varepsilon > 0$, we write

$$a \simeq b \pm \varepsilon \quad \text{if } |a - b| < \varepsilon.$$

Theorem 4.9. *Let (M, F) be a self-reverse Finsler metric disk with minimizing interior geodesics. Then, for every $\varepsilon > 0$ and every integer n large enough, there exists a wall system $\mathcal{W}$, made up of n geodesics of M, such that for all $x, y \in M \setminus \mathcal{W}$, we have*

$$\frac{1}{n} d_{\mathcal{W}}(x, y) \simeq \frac{2}{L} d_F(x, y) \pm \varepsilon, \tag{4.4}$$

$$\frac{2}{n^2 - n}\,\mathrm{area}(M, \mathcal{W}) \simeq \frac{2\pi}{L^2}\,\mathrm{area}(M, F) \pm \varepsilon, \tag{4.5}$$

where $L = \mathrm{length}_F(\partial M)$. Furthermore, the wall system $\mathcal{W}$ is necessarily simple.

Note that [20, Theorem 7.1] states the existence of a system with similar approximation properties but not necessarily made up of geodesics.

Proof of Theorem 4.9. The wall system $\mathcal{W}$ will be formed by random geodesics. Recall that Γ is the space of traversing geodesics of M (i.e., geodesic arcs of M which do not intersect ∂M except at their endpoints where they meet the boundary transversely) and has a natural measure μ_Γ; see (3.8). Furthermore, this space has finite total measure $\mu_\Gamma(\Gamma) = 2L$; see Remark 3.3. Thus we may define on Γ the probability measure $\mathbb{P} = \frac{\mu_\Gamma}{2L}$.

Take n independent identically distributed (i.i.d.) random geodesics $\gamma_1, \ldots, \gamma_n$ of Γ with probability distribution $\mathbb{P}$. Almost surely, these geodesics form a wall system $\mathcal{W}$ of M (see Definition 4.2), because they are pairwise different and have only simple crossings located in the interior of M. Moreover, this wall system is simple, since the geodesics are minimizing and therefore they cannot cross each other more than once by Lemma 4.5. Now, Theorem 4.9 follows from the next two lemmas.

The first lemma is obtained by applying the weak law of large numbers to the Blaschke formula (3.4) in a uniform way.

Lemma 4.10. *With probability converging to 1 as $n \to \infty$, we have*

$$\frac{1}{n} d_{\mathcal{W}}(x, y) \simeq \frac{2}{L} d_F(x, y) \pm \varepsilon \quad \textit{for all } x, y \in M \setminus \mathcal{W}.$$

Proof. Let $\mathcal{D}$ be a finite covering of M by smoothly bounded disks D with perimeter $\text{length}_F(\partial D) < \varepsilon$. Fix a basepoint p in each disk $D \in \mathcal{D}$ and denote by P the collection of all basepoints. Almost surely, the geodesics of $\mathcal{W}$ avoid the points of P and are transverse to the boundaries of the disks $D \in \mathcal{D}$.

The following claim shows that the conclusion of the lemma holds in some finite cases.

Claim 4.11. *The following assertions hold with probability converging to* 1 *as* $n \to \infty$*:*

(1) *For all* $p, q \in P$*, we have*

$$\frac{1}{n} d_{\mathcal{W}}(p,q) \simeq \frac{2}{L} d_F(p,q) \pm \varepsilon. \tag{4.6}$$

(2) *For every disk* $D \in \mathcal{D}$ *and all* $x, y \in D \setminus \mathcal{W}$*, we have*

$$\frac{1}{n} d_{\mathcal{W}}(x,y) \leq \left(\frac{1}{L} + \frac{1}{2} \right) \varepsilon. \tag{4.7}$$

Proof. (1) Recall that the distance-realizing arc $[p,q]$ is C^1 embedded in M (see Theorem 12.1).

The intersection function $f = f_{p,q} : \Gamma \to \mathbb{N}$ defined by

$$f(\gamma) = \#(\gamma \cap [p,q])$$

is a nonnegative measurable function. By Blaschke's formula (3.4), the random variables $X_i = f(\gamma_i)$ with $1 \leq i \leq n$ are i.i.d. with finite expected value

$$\mathbb{E}(X_i) = \int_\Gamma \#(\gamma \cap [p,q]) \, d\mathbb{P} = \frac{2}{L} d_F(p,q).$$

Note that $\mathbb{E}(|X_i|) = \mathbb{E}(X_i) < \infty$. By the weak law of large numbers applied to $\{X_i\}$ (see e.g. [42]), we derive

$$\left| \frac{1}{n} \sum_{i=1}^{n} \#(\gamma_i \cap [p,q]) - \frac{2}{L} d_F(p,q) \right| < \varepsilon$$

with probability converging to 1 as $n \to \infty$. By Proposition 4.7, we have

$$d_{\mathcal{W}}(p,q) = \text{length}_{\mathcal{W}}([p,q]) = \sum_{i=1}^{n} \#(\gamma_i \cap [p,q]),$$

hence (1) follows.

(2) The proof of the second assertion is similar. For a disk $D \in \mathcal{D}$, the intersection function $f(\gamma) = \#(\gamma \cap \partial D)$ has expected value $\frac{2}{L} \text{length}_F(\partial D)$ by Blaschke's formula (3.4). Applying the weak law of large numbers to the random variables $X_i = f(\gamma_i)$ as previously, we derive

$$\left| \frac{1}{n} \sum_{i=1}^{n} \#(\gamma_i \cap \partial D) - \frac{2}{L} \text{length}_F(\partial D) \right| < \varepsilon$$

with probability converging to 1 as $n \to \infty$. Thus,

$$\begin{aligned}\frac{1}{n}\,\mathrm{length}_{\mathcal{W}}(D) &\simeq \frac{2}{L}\,\mathrm{length}_F(D) \pm \varepsilon \\ &\leq \left(\frac{2}{L}+1\right)\varepsilon.\end{aligned}$$

Since D is a disk with minimizing interior geodesics, the discrete part of Lemma 4.8 yields (2). ■

Without loss of generality, we can assume that the conclusion of Claim 4.11 is satisfied. Let $x, y \in M \setminus \mathcal{W}$. The points x and y lie in some disks D_x and D_y of $\mathcal{D}$. Denote by p_x and p_y the basepoints of D_x and D_y. Since D_x is a disk with minimizing interior geodesics, by Lemma 4.8 we have

$$d_F(x, p_x) \leq \tfrac{1}{2}\,\mathrm{length}_F(\partial D_x) < \tfrac{1}{2}\varepsilon,$$

thus by the triangle inequality,

$$|d_F(x,y) - d_F(p_x, p_y)| \leq d_F(x, p_x) + d_F(y, p_y) < \varepsilon. \tag{4.8}$$

Combining the triangle inequality with (4.7), we obtain

$$\begin{aligned}\left|\frac{1}{n}d_{\mathcal{W}}(x,y) - \frac{1}{n}d_{\mathcal{W}}(p_x,p_y)\right| &\leq \frac{1}{n}d_{\mathcal{W}}(x,p_x) + \frac{1}{n}d_{\mathcal{W}}(y,p_y) \\ &\leq \left(\frac{2}{L}+1\right)\varepsilon.\end{aligned} \tag{4.9}$$

Thus,

$$\frac{1}{n}\,d_{\mathcal{W}}(x,y) \underset{(4.9)}{\simeq} \frac{1}{n}\,d_{\mathcal{W}}(p_x,p_y) \underset{(4.6)}{\simeq} \frac{2}{L}\,d_F(p_x,p_y) \underset{(4.8)}{\simeq} \frac{2}{L}\,d_F(x,y)$$

up to additive constants which are universal multiples of ε (namely, $(2/L+1)\varepsilon$ for the first one, ε for the second and $(2/L)\varepsilon$ for the third one). Therefore,

$$\left|\frac{1}{n}\,d_{\mathcal{W}}(x,y) - \frac{2}{L}\,d_F(x,y)\right| < C_0\varepsilon$$

where $C_0 = 4/L + 2$. This proves Lemma 4.10. ■

The second lemma is obtained by applying a (slightly generalized) weak law of large numbers to the Santaló+Blaschke formula (3.6).

Lemma 4.12. *With probability converging to 1 as $n \to \infty$, we have*

$$\frac{2}{n^2-n}\,\mathrm{area}(M,\mathcal{W}) \simeq \frac{2\pi}{L^2}\,\mathrm{area}(M) \pm \varepsilon.$$

Proof. The intersection counting function $f : \Gamma \times \Gamma \to \mathbb{N}$ defined by

$$f(\gamma,\gamma') = \#(\gamma \cap \gamma')$$

is a measurable function that takes value 0 or 1 almost surely. The $n(n-1)/2$ random variables $X_{i,j} = f(\gamma_i, \gamma_j)$ with $i < j$ are identically distributed but not completely independent. In fact, $X_{i,j}$ is independent of $X_{k,l}$ if and only if $\{i,j\} \cap \{k,l\} = \emptyset$. To apply the generalized weak law of large numbers, Theorem 4.13 below, we must check that the variables $X_{i,j}$ are sufficiently independent. There are $n(n-1)/2 \sim n^2$ variables $X_{i,j}$, which yield $\sim n^4$ pairs $(X_{i,j}, X_{k,l})$, of which only $\sim n^3$ are not independent. Therefore the proportion of nonindependent pairs $p \sim n^3/n^4 \sim 1/n$ goes to zero as $n \to \infty$. Thus, by Theorem 4.13, the average value of the variables $X_{i,j}$,

$$\frac{\sum_{i<j} X_{i,j}}{n(n-1)/2} = \frac{\sum_{i<j} \#(\gamma_i \cap \gamma_j)}{n(n-1)/2} = \frac{\operatorname{area}(M, \mathcal{W})}{n(n-1)/2}$$

converges in probability to the expected value, which is equal to

$$\mathbb{E}(X_{i,j}) = \iint_{\Gamma \times \Gamma} \#(\gamma \cap \gamma')\, \mathrm{d}\mathbb{P}(\gamma)\, \mathrm{d}\mathbb{P}(\gamma') = \frac{2}{L^2} \pi \operatorname{area}(M)$$

by the Santaló+Blaschke formula (3.6). ■

This concludes the proof of Theorem 4.9. ■

Let us prove the following generalization of the weak law of large numbers.

Theorem 4.13 (Weak law of large numbers for identically distributed, mostly independent random variables). *Fix a real valued random variable X with finite expected absolute value $\mathbb{E}(|X|) < \infty$ and an integer $n > 0$. Then the average $\bar{X} = \frac{1}{n} \sum_i X_i$ of n random variables X_i, each with the same distribution as X, is near the expected value $\mathbb{E}(X)$ with probability arbitrarily close to 1 if the proportion of nonindependent pairs,*

$$p = \frac{\#\{(i,j) \mid X_i \text{ and } X_j \text{ are not independent}\}}{n^2},$$

is small; more precisely, if for all $\varepsilon, \delta > 0$, there exists $p_0 = p_0(X, \delta, \varepsilon) > 0$ such that if $p \le p_0$, then $\mathbb{P}(|\bar{X} - \mathbb{E}(X)| \ge \varepsilon) \le \delta$.

Remark 4.14. Note that we do not explicitly require n to be large, but this is generally necessary for p to be small, because each variable X_i is in general correlated with itself,[4] which implies that $p \ge n/n^2 = 1/n$. If these are the only correlations and n goes to infinity, then $p = 1/n \to 0$ and therefore $\bar{X}$ converges to $\mathbb{E}(X)$ in probability. In this way, we recover the usual weak law of large numbers.

Proof of Theorem 4.13. The proof is similar to the standard proof of the weak law of large numbers; see [42, Theorem 1.5.1] for instance. It proceeds by cases; only the first one requires attention to the nonindependent pairs.

[4] A random variable is independent of itself if and only if its probability distribution is concentrated in one value.

Case $\mathbb{E}(X^2) < \infty$ *and* $\mathbb{E}(X) = 0$. Fix $\varepsilon > 0$. We have to prove that the probability of deviation, $\mathbb{P}(|\bar{X}| > \varepsilon)$, gets arbitrarily low if p is sufficiently small. To apply Chebyshev's inequality, we compute

$$\mathbb{E}(\bar{X}^2) = \frac{1}{n^2}\sum_i \sum_j \mathbb{E}(X_i X_j) \leq p\,\mathbb{E}(X^2).$$

Here we have used the Cauchy–Schwarz inequality $\mathbb{E}(X_i X_j) \leq \mathbb{E}(X^2)$ and the fact that $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\,\mathbb{E}(X_j) = \mathbb{E}(X)^2 = 0$ if X_i and X_j are independent. Applying Chebyshev's inequality, we obtain

$$\mathbb{P}(|\bar{X}| \geq \varepsilon) \leq \frac{\mathbb{E}(\bar{X}^2)}{\varepsilon^2} \leq \frac{p\,\mathbb{E}(X^2)}{\varepsilon^2} \xrightarrow[p \to 0]{} 0,$$

which we had to prove.

Case $\mathbb{E}(X^2) < \infty$. This case follows from the previous one applied to the random variable $Y = X - \mathbb{E}(X)$, which satisfies $\mathbb{E}(Y^2) < \infty$ and $\mathbb{E}(Y) = 0$.

The general case $\mathbb{E}(|X|) < \infty$, which is not needed in this article, follows from a truncation argument as in the usual proof of the weak law of large numbers, given for instance in [42]. ∎

5. Minimal area of disks: from discrete to Finsler metrics

The goal of this section is to state a discrete version of the area lower bound on Finsler disks with minimizing interior geodesics and to show how to derive the area lower bound for Finsler metrics from its discrete version.

Let us recall the area lower bound for Finsler metrics we want to prove.

Theorem 5.1. *Let M be a self-reverse Finsler metric disk of radius r with minimizing interior geodesics. Then the Holmes–Thompson area of M satisfies*

$$\text{area}(M) \geq \frac{6}{\pi} r^2.$$

In order to state the discrete version of this result, we need to introduce the notion of simple discrete metric disks.

Definition 5.2. A topological disk D with a quasi wall system $\mathcal{W}$ is a *simple discrete metric disk* of radius r centered at an interior point $O \in D \setminus \mathcal{W}$ if the quasi wall system $\mathcal{W}$ is simple (see Definition 4.2), all the points of $D \setminus \mathcal{W}$ are at $d_{\mathcal{W}}$-distance at most r from O and all the points of $\partial D \setminus \mathcal{W}$ are at distance exactly r from O.

It is essential here to allow $\mathcal{W}$ to be a quasi wall system rather than a wall system. Indeed, all points of $\mathcal{W}$ located on ∂D necessarily have multiplicity 2.

The following result, which will be proved in the subsequent sections, can be seen as a discrete version of Theorem 5.1.

Theorem 5.3. *The discrete area of every simple discrete metric disk* $(D, \mathcal{W})$ *of radius* r *satisfies*

$$\operatorname{area}(D, \mathcal{W}) \geq \frac{3}{2} r^2.$$

Furthermore, equality is attained.

Assuming this discrete area lower bound, we can derive Theorem 5.1 as follows.

Proof of Theorem 5.1 *assuming Theorem* 5.3. Let M be a Finsler disk of radius r centered at O with minimizing interior geodesics. By Theorem 4.9, for every $\varepsilon > 0$, there exists a simple wall system $\mathcal{W}_M$, made up of n interior geodesics of M, satisfying the estimates (4.4) and (4.5). The simple wall system $\mathcal{W}_M$ decomposes M into convex polygonal cells. By definition, all the points in a cell are at the same distance from the center of M with respect to the discrete distance $d_{\mathcal{W}_M}$. Since M has minimizing interior geodesics, the geodesic rays of length r issuing from its center O form a geodesic foliation $\mathcal{F}$ of the punctured disk $M \setminus \{O\}$. The sides of the cells of M, which lie in the geodesics of $\mathcal{W}_M$, are transverse to the foliation $\mathcal{F}$, otherwise the origin O would lie in $\mathcal{W}$.

Consider a convex polygonal cell Δ of M not containing O. Choose an arbitrary interior point of Δ as its center. Denote by d the $d_{\mathcal{W}_M}$-distance from O to the interior of the cell Δ. The geodesic rays of the foliation intersecting Δ form a spray $\mathcal{F}_\Delta$, where each ray of $\mathcal{F}_\Delta$ intersects Δ along an interval with nonempty interior, except for the two extremal rays of the spray which intersect the convex polygonal cell Δ at two vertices; see Figure 1. Denote by β_Δ the broken line made up of two segments joining the center of Δ to these two extremal vertices. Note that every geodesic ray of the spray $\mathcal{F}_\Delta$ intersects the broken line β_Δ at a single point; see Figure 1. Since the rays of the spray are length-minimizing with respect to $d_{\mathcal{W}_M}$ (see Proposition 4.7), all the cells intersecting the spray between O and β_Δ are at $d_{\mathcal{W}_M}$-distance at most d from O, and all the cells intersecting the spray after β_Δ are at $d_{\mathcal{W}_M}$-distance at least d from O.

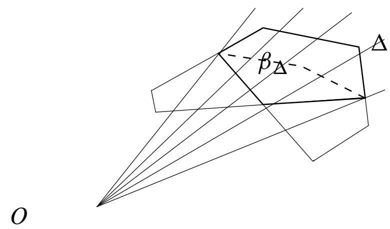

Fig. 1. Spray $\mathcal{F}_\Delta$ intersecting the convex polygonal cell Δ.

Denote by r_0 the integer part of $n(2r/L - \varepsilon)$. By (4.4), every boundary point $p \in \partial M \setminus \mathcal{W}_M$ is at $d_{\mathcal{W}_M}$-distance greater than r_0 from O, that is, $d_{\mathcal{W}_M}(O, p) > r_0$. A cell of M whose interior points are at $d_{\mathcal{W}_M}$-distance r_0 from O will be referred to as an *outermost cell*. The broken lines β_Δ, where Δ runs over all outermost cells of M, form a piecewise geodesic closed curve bounding a topological disk $D \subseteq M$ containing O. This curve can be smoothed to ensure that D is a smoothly bounded manifold. The restriction $\mathcal{W} = \mathcal{W}_M \cap D$ of $\mathcal{W}_M$ to D defines a simple quasi wall system on D. By construction, all the points of $D \setminus \mathcal{W}$ are at $d_{\mathcal{W}}$-distance at most r_0 from O and all the points of $\partial D \setminus \mathcal{W}$ are at distance exactly r_0 from O. Hence, $(D, \mathcal{W})$ is a simple discrete metric disk of radius r_0. By Theorem 5.3 and by definition of the discrete area (4.2), we have

$$\operatorname{area}(M, \mathcal{W}_M) \geq \operatorname{area}(D, \mathcal{W}) \geq \tfrac{3}{2} r_0^2.$$

Dividing by n^2, using (4.4) and (4.5), and letting ε go to zero, we obtain

$$\frac{\pi}{L^2} \operatorname{area}(M) \geq \frac{3}{2}\left(\frac{2}{L} r\right)^2.$$

Hence, $\operatorname{area}(M) \geq \frac{6}{\pi} r^2$. ■

Sections 6–9 are devoted to the proof of Theorem 5.3.

6. Quasi wall systems and interval families

In this section, we show how to encode a simple discrete disk as a 1-dimensional object.

We start by proving the following basic fact about simple discrete metrics.

Proposition 6.1. *Let D be a disk with a simple quasi wall system $\mathcal{W}$. Then*

$$d_{\mathcal{W}}(x, y) = \textit{number of walls of } \mathcal{W} \textit{ that separate } x \textit{ from } y \tag{6.1}$$

for any $x, y \in D \setminus \mathcal{W}$.

Note that if D is a Finsler disk with minimizing interior geodesics and $\mathcal{W}$ is geodesic, then this proposition follows from Proposition 4.7.

Proof of Proposition 6.1. It is clear that

$$d_{\mathcal{W}}(x, y) \geq \text{number of walls of } \mathcal{W} \text{ that separate } x \text{ from } y.$$

To prove the reverse inequality we will show the following.

Claim 6.2. *There exists a smooth path α from x to y that is in general position with respect to $\mathcal{W} \cup \partial D$ and crosses each wall of $\mathcal{W}$ at most once.*

Here, we say that a smooth curve α is *in general position* with respect to an immersed 1-submanifold N if it is regular, transverse to N and avoids the self-intersections of N. If α is piecewise smooth, we require in addition that none of its nonsmooth points lies in N.

The claim is a version of Levi's extension (or enlargement) lemma for pseudoline arrangements. This version concerns arrangements on a disk, rather than on the projective plane as in the more standard version of the lemma (found e.g. in [31, Thm. 5.1.1]).

We prove the claim by induction on the number of walls. Suppose the claim is valid for any quasi wall system $\mathcal{W}$ made up of n walls. Consider a simple quasi wall system $\mathcal{W}'$ obtained by adding an extra w' to $\mathcal{W}$. By inductive hypothesis, there is a smooth path α that satisfies all the conditions of the claim with respect to $\mathcal{W}$. By perturbing α, we ensure that it is transverse to w' as well. If α crosses w' at most once, then we are done. Otherwise, let x' and y' be the first and last points of α where α crosses w'. Note that they are generic points of w': they are neither on $\mathcal{W}$, nor on ∂D. Replace the segment of α from x' to y' by the segment $[x', y']$ of w', and let α' be the resulting curve. We claim that α' is a piecewise smooth curve, in general position with respect to $\mathcal{W}$, that crosses each wall of $\mathcal{W}$ at most once. This is because the segment $[x', y']$ that we inserted only crosses the walls of $\mathcal{W}$ that separate x' from y' (since it is part of a wall of the simple quasi wall system $\mathcal{W}'$), and these walls are necessarily crossed as well by the piece of α between x' and y' that we replaced.

The next step is to perturb the curve α' so that the segment $[x', y']$ is displaced sideways and away from w' and the resulting curve α'' is in general position with respect to $\mathcal{W} \cup \partial D$ and crosses $\mathcal{W}$ the same number of times as α' does, and in addition is transverse to w' and crosses w' at most once. Thus, α'' is in general position with respect to $\mathcal{W}' \cup \partial D$ and crosses each wall of $\mathcal{W}'$ at most once, but is nonsmooth at two points. To make it smooth, we modify it near these two points. ∎

Let $(D, \mathcal{W})$ be a simple discrete disk of radius r and center O (see Definition 5.2). Identify the boundary ∂D with the circle S^1, and identify the punctured disk $D \setminus \{O\}$ with the flat cylinder $\mathcal{C} = S^1 \times [0, +\infty)$. Under this identification, the point O of D corresponds to the point at infinity in the one-point compactification of the cylinder $\mathcal{C}$. Note that the universal cover of $\mathcal{C}$ is the half-plane $\mathcal{H} = \mathbb{R} \times [0, +\infty)$.

Definition 6.3. Given a simple arc α in the cylinder $M = D \setminus \{O\}$ (or in the half-plane $M = \mathbb{R} \times [0, +\infty)$) with endpoints on the boundary ∂M, denote by $\overline{\alpha}$ the segment of ∂D with the same endpoints, homotopic to α in M. The arc α *covers a point* p of ∂M if p lies in $\overline{\alpha}$. Similarly, the arc α *covers another arc* β if $\overline{\beta}$ lies in $\overline{\alpha}$. Two arcs α and α' are *adjacent* if the intervals $\overline{\alpha}$ and $\overline{\alpha'}$ are adjacent, meaning that they have exactly one point in common.

Definition 6.4. An arc in the flat cylinder $M = S^1 \times [0, +\infty)$ (or in the half-plane $M = \mathbb{R} \times [0, +\infty)$) with endpoints on the boundary ∂M is *standard* if it consists of a segment of slope 1 followed by a segment of slope -1; see Figure 2. A quasi wall system $\mathcal{W}$ is standard if its walls are standard arcs. For two boundary points $a, b \in S^1 = \partial M$ (or $a < b \in \mathbb{R}$ if M is the half-plane), we denote by $[a, b]$ the arc of S^1 that goes from a to b in the positive (i.e., counterclockwise) sense, and we denote by $\widehat{ab}$ the standard arc in M that is homotopic to $[a, b]$.

Let $(D, \mathcal{W})$ be a simple discrete disk of radius r centered at O. Denote by $\mathcal{I} = \mathcal{I}_{\mathcal{W}}$ the set of boundary intervals $\overline{\alpha}$ homotopic to the walls α of $\mathcal{W}$. The family $\mathcal{I}$ of intervals of S^1 contains all the information about $\mathcal{W}$ that is relevant to our problem of finding simple discrete disks of minimum area. For instance, two walls α, β of $\mathcal{W}$ meet on ∂D if and only if the intervals $\overline{\alpha}$, $\overline{\beta}$ have a common endpoint. That is,

$$\#I_{\partial D}(\alpha, \beta) = 1 \iff \#(\partial\overline{\alpha} \cap \partial\overline{\beta}) = 1. \tag{6.2}$$

Furthermore, assuming $\overline{\alpha}$ and $\overline{\beta}$ have no common endpoints, the arcs α and β cross in the interior of D if and only if the interval $\overline{\alpha}$ contains *exactly one* endpoint of $\overline{\beta}$. That is,

$$\#I_{\mathrm{Int}\, D}(\alpha, \beta) = 1 \iff \#(\overline{\alpha} \cap \partial\overline{\beta}) = 1. \tag{6.3}$$

One consequence of these formulas is that the discrete area of $(D, \mathcal{W})$ given by (4.2) may be computed from $\mathcal{I}$.

The following result characterizes the relation between the quasi wall system $\mathcal{W}$ and the interval family $\mathcal{I}$. Before stating this result, we need to introduce a definition. A point p of S^1 is *generic* with respect to a finite interval family $\mathcal{I}$ of S^1 if p is not an endpoint of any interval of $\mathcal{I}$. Alternatively, the endpoints of the intervals of $\mathcal{I}$ are the nongeneric points of S^1.

Proposition 6.5. *Let (D, W) be a simple discrete disk of radius r centered at O. The family $\mathcal{I} = \mathcal{I}_{\mathcal{W}}$ of intervals of S^1 has the following properties:*

(1) *no two intervals of $\mathcal{I}$ cover S^1;*

(2) *every generic point of S^1 is contained in exactly r intervals of $\mathcal{I}$;*

(3) *every nongeneric point of S^1 is an endpoint of exactly two, adjacent intervals of $\mathcal{I}$.*

Moreover, if a finite family $\mathcal{I}$ of intervals of S^1 satisfies conditions (1)–(3), then $\mathcal{I} = \mathcal{I}_{\mathcal{W}}$ for some quasi wall system $\mathcal{W}$ that makes D a simple discrete metric disk of radius r and center O. For instance, one may let $\mathcal{W}$ be the unique standard quasi wall system homotopic to $\mathcal{I}$ on $D \setminus \{O\}$.

Proof. (1) If two intervals $\overline{\alpha}, \overline{\beta} \in \mathcal{I}$ cover S^1, then the corresponding walls α, β of $\mathcal{W}$ would form a bigon containing the point O, which implies they cross twice, contradicting the hypothesis that $\mathcal{W}$ is simple.

(2) Consider a generic point $p \in S^1$. Since $\mathcal{W}$ is a simple quasi wall system on D, the distance between any pair of points of D is the number of walls that separate them (see Proposition 6.1). On the other hand, the walls that separate O from p are the walls that cover p. This yields the result.

(3) follows from the previous property: if $p \in S^1$ is the endpoint of some interval $\overline{\alpha} \in \mathcal{I}$, it must also be the startpoint of some other interval so that every generic point near p is contained in the same number r of intervals of $\mathcal{I}$. This means that p is the endpoint of two walls, and it cannot be the endpoint of more walls because $\mathcal{W}$ can only have simple self-intersections on ∂D since it is a quasi wall system (see Definition 4.2).

Now, let $\mathcal{I}$ be a finite family of intervals of S^1 satisfying conditions (1)–(3), and let $\mathcal{W}$ be the unique standard quasi wall system homotopic to $\mathcal{I}$ on $D \setminus \{O\}$. Clearly, $\mathcal{W}$ is a quasi wall system, and it is simple because it is made up of arcs that intersect each other at most once. Also, every point $p \in D \setminus \mathcal{W}$ is at distance at most r from O, and exactly r

if $p \in \partial D$. (A shortest path is the vertical ray from p to O.) This shows that $(D, \mathcal{W})$ is a simple discrete disk of radius r centered at O. ■

7. Inadmissible configurations in a minimal simple disk

In this section, we rule out some intersection patterns for an extremal quasi wall system on a disk.

Consider a quasi wall system $\mathcal{W}$ on D defining a simple discrete metric disk of radius r *with minimal discrete area*. By Proposition 6.5, we can assume that $\mathcal{W}$ is formed by standard arcs (see Definition 6.4).

Lemma 7.1. *No arc of $\mathcal{W}$ covers two* (*possibly adjacent*) *intersecting arcs of $\mathcal{W}$.*

Proof. Towards a contradiction, suppose that an arc γ of $\mathcal{W}$ covers two intersecting arcs $\alpha = \widehat{ac}$ and $\beta = \widehat{bd}$ of $\mathcal{W}$. Switching the roles of the two arcs if necessary, we may assume that the points a, b, c, d appear in that order in the interval $\overline{\gamma}$ (with possibly $b = c$); see Figure 2. Let $\mathcal{W}'$ be the collection of curves obtained from $\mathcal{W}$ by replacing α and β with the standard arcs $\alpha' = \widehat{ad}$ and $\beta' = \widehat{bc}$ (with no β' if $b = c$). See Figure 2. Note that, like $\mathcal{W}$, the immersed 1-submanifold $\mathcal{W}'$ is a quasi wall system on D. Moreover, we claim that $\mathcal{W}'$ also makes D a simple discrete metric disk of radius r centered at O. This is because none of the properties (1)–(3) of Proposition 6.5 is affected by the replacement. For instance, there is no arc δ of $\mathcal{W}'$ such that the intervals $\overline{\delta}$ and $\overline{\alpha'}$ cover the boundary ∂D, because in that case $\overline{\delta}$ and $\overline{\gamma}$ would also cover ∂D; however, the arcs $\overline{\delta}$ and $\overline{\gamma}$ are already present in $\mathcal{W}$, contradicting by Proposition 6.5 the fact that $\mathcal{W}$ is simple. Also, the fact that every generic point of ∂D is covered by exactly r arcs of the quasi wall system is clearly maintained, as also is the fact that each nongeneric boundary point is the common endpoint of two adjacent walls.

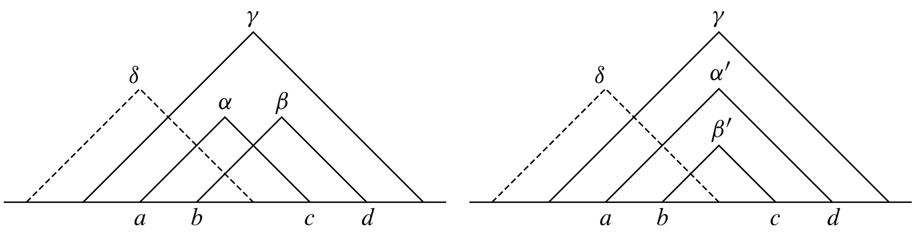

Fig. 2. Replacing two intersecting arcs covered by a third arc.

Let us show that the area of $(D, \mathcal{W}')$ is less than the area of $(D, \mathcal{W})$ by comparing the number of self-intersections of the quasi wall systems $\mathcal{W}$ and $\mathcal{W}'$ according to the discrete area formula (4.2). First, observe that every pair of arcs of $\mathcal{W}$ different from α and β belongs to $\mathcal{W}'$. Therefore, these pairs of arcs give the same contribution to the discrete areas of $\mathcal{W}$ and $\mathcal{W}'$. Let $\delta = \widehat{pq}$ be an arc of $\mathcal{W}$ different from α and β. By considering cases depending on the location of the endpoints p and q with respect to the points a, b, c and d, we see that

$$\#I_{\mathrm{Int}\,D+\frac{1}{2}\partial D}(\delta,\alpha'\cup\beta') \le \#I_{\mathrm{Int}\,D+\frac{1}{2}\partial D}(\delta,\alpha\cup\beta).$$

In fact, equality holds unless p and q lie in the interiors of $[a,b]$ and $[c,d]$, in which case the inequality is strict. Finally, note that

$$\#I_{\mathrm{Int}\,D}(\alpha',\beta') = 0 \quad \text{and} \quad \#I_{\mathrm{Int}\,D}(\alpha,\beta) = 1.$$

We conclude that

$$\mathrm{area}(D,\mathcal{W}') \le \mathrm{area}(D,\mathcal{W}) - 1,$$

which contradicts the minimiality of the discrete area of $(D,\mathcal{W})$. ∎

Lemma 7.2. *No arc of $\mathcal{W}$ intersects two adjacent arcs of $\mathcal{W}$.*

Proof. Suppose that an arc γ of $\mathcal{W}$ intersects two adjacent arcs α and β of $\mathcal{W}$. We choose γ so that it is minimal with respect to the covering relation (i.e., no arc of $\mathcal{W}$ covered by γ intersects α and β). Denote by a,b,c,d,e the endpoints of the three arcs, in the order in which they are found on the interval $\overline{\alpha}\cup\overline{\beta}$. Thus, $\alpha=\widehat{ac}$, $\beta=\widehat{ce}$ and $\gamma=\widehat{bd}$, and no arc of $\mathcal{W}$ that covers c is covered by γ (other than γ itself); see Figure 3. Let c^- and c^+ be two points of ∂D close to c such that $[c^-,c^+]\cap\partial\mathcal{W}=\{c\}$. Let $\mathcal{W}'$ be the collection of curves obtained from $\mathcal{W}$ by replacing the three arcs α, β and γ with the four arcs $\alpha'=\widehat{ac^+}$, $\beta'=\widehat{c^-e}$, $\gamma^-=\widehat{bc^-}$ and $\gamma^+=\widehat{c^+d}$; see Figure 3.

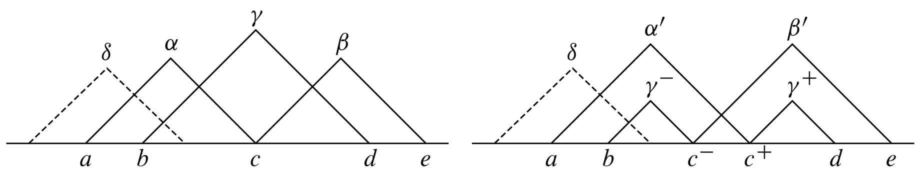

Fig. 3. Replacing a configuration of one arc intersecting two adjacent arcs.

Note that $\mathcal{W}'$ is a quasi wall system on the disk D. In fact, $\mathcal{W}'$ makes D a simple discrete disk of radius r centered at O. To see this we argue as in the proof of Lemma 7.1. By Proposition 6.5, it is enough to check that the family $\mathcal{I}=\mathcal{I}_{\mathcal{W}'}$ of boundary segments $\overline{\delta}$ corresponding to the walls δ of $\mathcal{W}'$ has properties (1)–(3) of Proposition 6.5. To check property (2) (that each generic point of ∂D is covered r times by the walls of $\mathcal{W}'$) note that both $\alpha\cup\beta\cup\gamma$ and $\alpha'\cup\beta'\cup\gamma^-\cup\gamma^+$ cover twice the generic points of $[b,d]$ and once the remaining generic points of $[a,e]$. Property (3) regarding nongeneric boundary points is also maintained, with the wall endpoint c replaced by the two points c^- and c^+. Finally, to check property (1), suppose δ and ε are two arcs of $\mathcal{W}'$ that cover the whole boundary ∂D. It is impossible for both δ and ε to be among the new arcs α', β' and $\gamma^\pm$ because that would mean that α and β already cover ∂D, contradicting the fact that $\mathcal{W}$ is simple. Similarly, the arcs δ and ε cannot both be among the unchanged arcs (those in $\mathcal{W} \cap \mathcal{W}'$) either, since otherwise $\mathcal{W}$ would not be simple. Therefore, δ is one of the unchanged arcs and ε is one of the new arcs α', β', $\gamma^{\pm}$. In the case $\varepsilon = \alpha'$, we see that δ and α' cannot cover ∂D since this would imply that δ and α already cover ∂D. This is because $\overline{\alpha'} \setminus \overline{\alpha}$ is contained in the interval $[c^-, c^+]$ which contains no endpoints of δ since $[c^-, c^+] \cap \mathcal{W} = \{c\}$. The case $\varepsilon = \beta'$ is analogous and the cases $\varepsilon = \gamma^{\pm}$ are easier to rule out since the arcs $\gamma^{\pm}$ are covered by γ. We conclude that property (1) is satisfied, thus $(D, \mathcal{W}')$ is a simple discrete metric disk of radius r.

Let us show that the area of $(D, \mathcal{W}')$ is less than the area of $(D, \mathcal{W})$. Again, we use the discrete area formula (4.2), which says

$$\operatorname{area}(D, \mathcal{W}) = \sum_{\{\delta,\varepsilon\}} \# I_{\operatorname{Int} D + \frac{1}{2}\partial D}(\delta, \varepsilon)$$

where the sum is over pairs $\{\delta, \varepsilon\}$ of different walls of $\mathcal{W}$. The pairs $\{\delta, \varepsilon\}$ of walls that are contained in $\mathcal{W} \cap \mathcal{W}'$ make the same contribution to $\operatorname{area}(D, \mathcal{W})$ and to $\operatorname{area}(D, \mathcal{W}')$. To evaluate the contribution of pairs $\{\delta, \varepsilon\}$ with $\delta \in W \cap \mathcal{W}'$ and $\varepsilon \notin \mathcal{W} \cap \mathcal{W}'$, we note that any arc $\delta = \widehat{pq}$ with no endpoints in $[c^-, c^+]$ satisfies

$$\# I_{\operatorname{Int} D + \frac{1}{2}\partial D}(\delta, \alpha' \cup \beta' \cup \gamma^+ \cup \gamma^-) = \# I_{\operatorname{Int} D + \frac{1}{2}\partial D}(\delta, \alpha \cup \beta \cup \gamma)$$

unless $p \in [b, c]$ and $q \in [c, d]$. This is seen by considering case by case the possible locations of p and q with respect to a, b, c, d, e. The equality holds for all arcs $\delta = \widehat{pq} \in \mathcal{W} \cap \mathcal{W}'$, because the exceptional case $p \in [b, c]$ and $q \in [c, d]$ is excluded by how γ was chosen: the arc $\gamma = \widehat{cd}$ covers no other arc $\delta = \widehat{pq}$ of $\mathcal{W}$ that in turn covers c. Finally, to compute the contribution of the pairs $\{\delta, \varepsilon\}$ where none of the two arcs δ and ε is in $\mathcal{W} \cap \mathcal{W}'$, we note that

$$\# I_{\operatorname{Int} D + \frac{1}{2}\partial D}(\alpha' \cup \beta' \cup \gamma^- \cup \gamma^+) = 2$$

while

$$\# I_{\operatorname{Int} D + \frac{1}{2}\partial D}(\alpha \cup \beta \cup \gamma) = 5/2.$$

We conclude that $\operatorname{area}(D, \mathcal{W}') = \operatorname{area}(D, \mathcal{W}) - 1/2$, contradicting the minimality of $\mathcal{W}$. ∎

8. Pairs of adjacent arcs

In this section we show that the sequences of adjacent arcs in an extremal quasi wall system on a disk have a periodic structure.

Consider a quasi wall system $\mathcal{W}$ on the disk D, made up of standard arcs, defining a simple discrete metric disk of radius r centered at O with minimal discrete area as in Section 7. Recall that the upper half-plane $\mathcal{H} = \mathbb{R} \times [0, +\infty)$ is the universal cover of the cylinder $\mathcal{C} = S^1 \times [0, +\infty) = D \setminus \{O\}$. We identify its boundary $\partial\mathcal{H}$ with the real line $\mathbb{R}$. Let $\mathcal{W}_{\mathcal{H}}$ be the quasi wall system on $\mathcal{H}$ formed by all the lifts of the arcs of $\mathcal{W}$.

Since D is a disk of radius r, it follows that every generic point of $\partial\mathcal{H}$ is covered by exactly r arcs of $\mathcal{W}_{\mathcal{H}}$. To ensure this uniform coverage, each endpoint of an arc must be the startpoint of another arc, and thus each arc of $\mathcal{W}_{\mathcal{H}}$ belongs to a bi-infinite sequence of consecutive arcs, called a "strand" of $\mathcal{W}_{\mathcal{H}}$.

Definition 8.1. A *strand* of $\mathcal{W}_{\mathcal{H}}$ is a bi-infinite sequence $(\alpha_i)_{i\in\mathbb{Z}}$ of consecutive arcs of $\mathcal{W}_{\mathcal{H}}$ of the form

$$\alpha_i = \widehat{a_i a_{i+1}}.$$

The points a_i where the strand $(\alpha_i)_i$ meets the boundary $\partial\mathcal{H}$ are called the *stops* of the strand. The *width* of an arc α_i is the number $a_{i+1} - a_i$.

Since each strand of arcs covers the generic points of $\partial\mathcal{H}$ once, it follows that the quasi wall system $\mathcal{W}_{\mathcal{H}}$ is composed of exactly r strands.

The following result describes how each strand intersects a pair of adjacent arcs of $\mathcal{W}_{\mathcal{H}}$.

Lemma 8.2. *Let $\alpha_0 = \widehat{a_0 a_1}$ and $\alpha_1 = \widehat{a_1 a_2}$ be two adjacent arcs of $\mathcal{W}_{\mathcal{H}}$. Then every strand of $\mathcal{W}_{\mathcal{H}}$ has exactly one arc with endpoints on the boundary interval $I = [a_0, a_2)$. This arc is covered by α_0 or by α_1.*

Proof. The strand that contains the arcs α_0 and α_1 clearly satisfies the conclusion. Thus let $(\beta_i)_{i\in\mathbb{Z}}$ be any other strand of $\mathcal{W}_{\mathcal{H}}$, numbered so that the arc β_0 covers the point a_1. This strand has a stop in I, otherwise β_0 would cover the two adjacent arcs α_0 and α_1, in contradiction with Lemma 7.1. Also, the strand $(\beta_i)_i$ cannot have stops in both intervals $[a_0, a_1)$ and $[a_1, a_2)$, as otherwise the arc β_0 would intersect the two adjacent arcs α_0 and α_1, contrary to Lemma 7.2. Thus the strand $(\beta_i)_i$ has stops in exactly one of the intervals $[a_0, a_1)$ and $[a_1, a_2)$, say, the second one; see Figure 4. Furthermore, it cannot

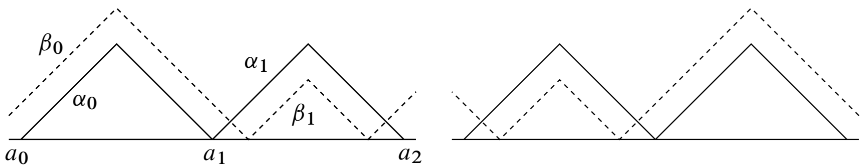

Fig. 4. Leaping over every other arc.

have just one stop in this interval, since otherwise the two adjacent arcs β_0, β_1 that share this stop would intersect α_1, contradicting Lemma 7.2. Also, it cannot have three stops in the interval, as otherwise the adjacent arcs β_1 and β_2 would be covered by α_1, in contradiction with Lemma 7.1. We conclude that the strand $(\beta_i)_i$ has exactly two stops (and therefore one arc) in the interval $[a_0, a_2)$, and both of these stops are covered by one of the arcs α_0 or α_1; see Figure 4. ■

Let n be the number of walls of the quasi wall system $\mathcal{W}$ on the disk D. From now on, changing the parameterization of the boundary circle $S^1 = \partial D$, we assume that S^1 is a circle of length n, thus $S^1 = \mathbb{R}/n\mathbb{Z}$, and that the endpoints of the walls of $\mathcal{W}$ are located at half-integer points. (This implies that the distance between two adjacent integer points is equal to 1.) Therefore, on the universal cover of the cylinder $\mathcal{C} = D \setminus \{O\}$, which is the upper half-plane $\mathcal{H}$, we have $\partial \mathcal{W}_{\mathcal{H}} = \mathbb{Z} + \frac{1}{2} \subseteq \mathbb{R} = \partial \mathcal{H}$.

Note that the quasi wall system $\mathcal{W}_{\mathcal{H}}$ is periodic of period n (where n is the number of walls of $\mathcal{W}$) in the sense that it is invariant by the horizontal translation of length n. However, the following result implies that $\mathcal{W}_{\mathcal{H}}$ is also periodic with period $2r$, where r is the number of strands of $\mathcal{W}_{\mathcal{H}}$ (see Definition 8.1).

Lemma 8.3. *The sum of the widths of two adjacent arcs α_0, α_1 of $\mathcal{W}_{\mathcal{H}}$ is equal to $2r$.*

Proof. Consider two adjacent arcs $\alpha_0 = \widehat{a_0 a_1}$ and $\alpha_1 = \widehat{a_1 a_2}$ as in Lemma 8.2. According to that lemma, each of the r strands of $\mathcal{W}_{\mathcal{H}}$ has exactly two stops in the interval $[a_0, a_2)$. Therefore there are $2r$ half-integers in that interval. It follows that $a_2 - a_0 = 2r$. ∎

Denote by $\mathcal{S}_{[t,t+2r)} = [t, t+2r) \times [0, +\infty)$ a strip of width $2r$ of the half-plane $\mathcal{H}$. The following result describes the arcs of the quasi wall system $\mathcal{W}_{\mathcal{H}}$ that are contained in such a strip.

Proposition 8.4. (1) *Each strip $\mathcal{S}_{[t,t+2r)}$ contains exactly one arc of each strand* (*and each of these arcs determines its strand completely*).

(2) *The r arcs contained in a strip $\mathcal{S}_{[t,t+2r)}$ do not intersect each other.*

(3) *Any pair of strands intersects each other exactly twice in the strip $\mathcal{S}_{[t,t+2r)}$.*

Proof. (1) Consider a strand $(\alpha_i)_{i\in\mathbb{Z}}$, with $\alpha_i = \widehat{a_i a_{i+1}}$. According to Lemma 8.3, we have $a_{i+2} = a_i + 2r$ for all i. This implies that the strip $\mathcal{S}_{[t,t+2r)}$ contains exactly two stops and thus exactly one arc of the strand $(\alpha_i)_i$. The same equation implies that two consecutive stops determine the strand.

(2) Consider a second strand $(\beta_j)_{j\in\mathbb{Z}}$, with $\beta_j = \widehat{b_j b_{j+1}}$. Assuming that two arcs α_0 and β_0 of $\mathcal{W}_{\mathcal{H}}$ intersect, we want to show that they are not contained in a strip $\mathcal{S}_{[t,t+2r)}$. We may assume without loss of generality that $a_0 < b_0$, therefore $b_0 \in (a_0, a_1)$. Since the strand $(\beta_j)_j$ has a stop in the interval $[a_0, a_1)$, by Lemma 8.2 it cannot have a stop in $[a_1, a_2)$. It follows that $b_1 > a_2 = a_0 + 2r$, hence the arcs $\alpha_0 = \widehat{a_0, a_1}$ and $\beta_0 = \widehat{b_0, b_1}$ are not contained in a strip of width $2r$.

(3) Consider two strands $(\alpha_i)_{i\in\mathbb{Z}}$ and $(\beta_j)_{j\in\mathbb{Z}}$ as above. Since $a_{i+2} = a_i + 2$ as shown in (1), the strand $(\alpha_i)_i$ is invariant by the horizontal translation by $2r$. The same holds with $(\beta_j)_j$. We want to show that they cross exactly twice in a strip $\mathcal{S}_{[t,t+2r)}$. By invariance under the horizontal translation by $2r$, we may choose t arbitrarily. For instance, we can choose $t = a_0$. By Lemma 8.2, the strand $(\beta_j)_j$ has stops in exactly one of the intervals (a_0, a_1) and (a_1, a_2). Thus, it intersects (twice) exactly one of the arcs $\alpha_0 = \widehat{a_0 a_1}$, $\alpha_1 = \widehat{a_1 a_2}$. ∎

We also note the following.

Lemma 8.5. *In the quasi wall system $\mathcal{W}_{\mathcal{H}}$, there is an arc of width* 1.

Proof. Let $\alpha_0 = \widehat{a_0 a_1}$ be an arc that is minimal with respect to covering (i.e., α_0 does not cover any arc of $\mathcal{W}_{\mathcal{H}}$). We want to show that $a_1 - a_0 = 1$. By Lemma 8.2, each strand other than the one generated by α_0 has two stops in the interval (a_0, a_2), both contained either in (a_0, a_1) or in (a_1, a_2). Thus, if the interval (a_0, a_1) has any stop, it has in fact two stops of a strand, and therefore there is an arc of $\mathcal{W}_{\mathcal{H}}$ covered by α_0. However, this possibility is excluded by the minimality of α_0. Therefore, the interval (a_0, a_1) has no stops and hence its endpoints a_0 and a_1 are consecutive half-integers. ∎

9. Proof of the discrete area lower bound

We can now proceed to the proof of the discrete area lower bound for simple discrete metric disks (see Theorem 5.3), making use of the previous notations and constructions. Namely, let us prove the following.

Theorem 9.1. *The discrete area of every simple discrete metric disk of radius r is at least $\frac{3}{2}r^2$.*

Proof. Let $(D, \mathcal{W}')$ be a simple discrete metric disk of radius r and center O that has minimal area. Recall that the punctured disk $D \setminus \{O\}$ is identified with the flat cylinder $\mathcal{C} = S^1 \times [0, +\infty)$. As shown in Section 6, $\mathcal{W}'$ is homotopic in $\mathcal{C}$ to a quasi wall system $\mathcal{W}$ made up of standard arcs such that $(D, \mathcal{W})$ is also a discrete disk of radius r centered at O and has the same area as $(D, \mathcal{W}')$. Thus we must show that $\operatorname{area}(D, \mathcal{W}) \geq \frac{3}{2}r^2$. Also, we may assume that the lift of $\mathcal{W}$ to the universal cover $\mathcal{H} = \mathbb{R} \times [0, +\infty)$ is a quasi wall system $\mathcal{W}_{\mathcal{H}}$ such that $\partial \mathcal{W}_{\mathcal{H}} = \mathbb{Z} + \frac{1}{2} \subseteq \mathbb{R} = \partial \mathcal{H}$ as in Section 8.

Let $t \in \mathbb{R}$ be a generic number. By Proposition 8.4, the weighted number of self-intersections of the quasi wall system $\mathcal{W}_{\mathcal{H}}$ that lie in the strip $\mathcal{S}_{[t,t+2r)}$ is

$$\# I_{\operatorname{Int} \mathcal{H} + \frac{1}{2} \partial \mathcal{H}}(\mathcal{W}_{\mathcal{H}}|_{\mathcal{S}_{[t,t+2r)}}) = 2\frac{r(r-1)}{2} + \frac{1}{2}2r = r^2. \tag{9.1}$$

The first term counts the crossings between the different strands: each pair of strands crosses twice, and the crossings are located in the interior of the half-plane $\mathcal{H}$. The second term counts, with weight $\frac{1}{2}$, the intersections that lie in the boundary $\partial \mathcal{H}$; these are the intersections between adjacent arcs that belong to the same strand. Thus, the discrete area of the disk $(D, \mathcal{W})$ is

$$\begin{aligned}\operatorname{area}(D, \mathcal{W}) &= \# I_{\operatorname{Int} \mathcal{H} + \frac{1}{2} \partial \mathcal{H}}(\mathcal{W}_{\mathcal{H}}|_{\mathcal{S}_{[t,t+n)}}) = \frac{n}{2r} \# I_{\operatorname{Int} \mathcal{H} + \frac{1}{2} \partial \mathcal{H}}(\mathcal{W}_{\mathcal{H}}|_{\mathcal{S}_{[t,t+2r)}}) \\ &= \frac{n}{2r} r^2,\end{aligned}$$

where n is the number of walls of $\mathcal{W}$.

To finish we will show that $n \geq 3r$. Let $(\alpha_i = \widehat{a_i a_{i+1}})_{i \in \mathbb{Z}}$ be a strand of $\mathcal{W}_{\mathcal{H}}$ such that $a_0 - a_{-1} = 1$. Such a strand exists by Lemma 8.5. Moreover, we may assume that $a_0 = \frac{1}{2}$ and $a_{-1} = -\frac{1}{2}$. The interval (a_0, a_1) has width $2r - 1$ (by Lemma 8.3) and contains $2r - 2$ half-integers.

Each of these half-integers is either the startpoint or the endpoint of one of the $r-1$ arcs that are covered by α_0 (see Proposition 8.4). Let b_0 be the rightmost of the $r-1$ startpoints. Note that

$$b_0 \geq a_0 + (r-1). \tag{9.2}$$

This point b_0 is a stop of a strand $(\beta_j = \widehat{b_j b_{j+1}})_{j\in\mathbb{Z}}$. The arc β_0 is covered by α_0 and the arc $\beta_1 = \widehat{b_1 b_2}$ intersects the arc α_0. The arcs α_0 and β_1 cannot extend over a whole fundamental domain $\mathcal{S}_{[t,t+n)}$ of the universal cover, by property (1) of Proposition 6.5. Therefore, $n > b_2 - a_0$. On the other hand, by Lemma 8.3 and inequality (9.2), we have

$$b_2 = b_0 + 2r \geq a_0 + 3r - 1.$$

We conclude that $n > 3r - 1$, or equivalently $n \geq 3r$, as we had to prove. ■

10. Simple discrete metric disks of minimal area

In this section, we analyze the equality case of Theorem 9.1.

Proposition 10.1. *For every positive integer r, there is a simple discrete metric disk of radius r and area $\frac{3}{2}r^2$. It is unique up to isotopy of the disk with the center fixed.*

Proof. Recall the proof of Theorem 9.1. Let $\mathcal{W}'$ be a simple quasi wall system such that $(D, \mathcal{W}')$ is a simple discrete metric disk of radius r with minimal discrete area. Consider the simple quasi wall system $\mathcal{W}$ homotopic to $\mathcal{W}'$ made up of standard arcs. To attain the lower bound on area$(D, \mathcal{W})$ and so on area$(D, \mathcal{W}')$, we must have $n = 3r$, therefore inequality (9.2) must be an equality. This implies that, for the $r-1$ arcs covered by α_0, the $r-1$ startpoints must precede the $r-1$ endpoints in the interval (a_0, a_1). In consequence, these $r-1$ arcs together with the arc α_0 form a chain with respect to the covering relation; see Figure 5. This implies that the r arcs are completely determined, and by Proposition 8.4, so are the quasi wall systems $\mathcal{W}_{\mathcal{H}}$ and $\mathcal{W}$, which are made up of standard arcs. Thus, the quasi wall system $\mathcal{W}_{\mathcal{H}}$ contains all arcs of the form $\widehat{kr-s, kr+s}$ with k integer and $s \in (0, r)$ half-integer; see Figure 5. Similarly, the quasi wall system $\mathcal{W}$ is

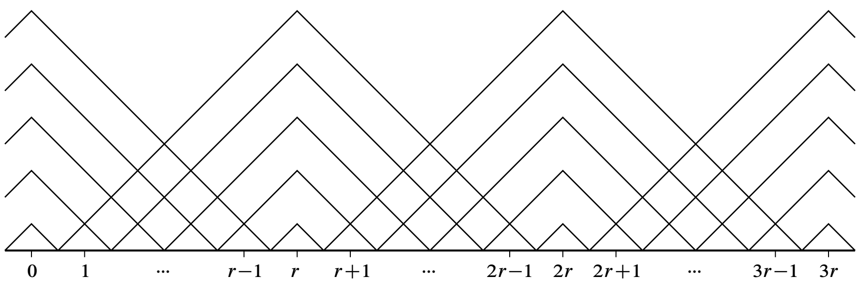

Fig. 5. The lift $W_{\mathcal{H}}$ corresponding to an area-minimizing simple discrete disk $(D, \mathcal{W})$ of radius $r = 5$, where the quasi wall system $\mathcal{W}$ consists of standard arcs.

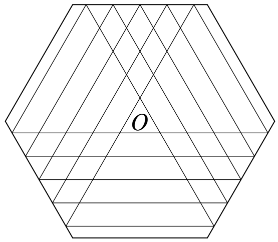

Fig. 6. An area-minimizing simple discrete disk $(D, \mathcal{W})$ of radius $r = 5$ where the topological disk D is a hexagon and the quasi wall system $\mathcal{W}$ consists of straight lines.

obtained from $\mathcal{W}_{\mathcal{H}}$ by taking the quotient of $\mathcal{H}$ under the horizontal translation by $3r$; see Figure 6. This proves the uniqueness of the simple discrete metric disk of minimal area, but only up to homotopy of the quasi wall system. The uniqueness up to isotopy of the disk follows from the next result.

Lemma 10.2. *Let $\mathcal{W}$ and $\mathcal{W}'$ be simple quasi wall systems on the disk D, homotopic in $D \setminus \{0\}$ and forming no triangle in $D \setminus \{O\}$. Then there is an isotopy of D which fixes O and carries $\mathcal{W}$ to $\mathcal{W}'$.*

Proof. We proceed by induction on the number n of walls. The case $n = 1$ is trivial. In general, we argue as follows.

Let γ be a wall of $\mathcal{W}$ that covers no other wall of $\mathcal{W}$ (see Definition 6.3). The curve γ divides the disk D into two topological closed disks A and B which intersect along γ, with $O \in A$. The part of $\mathcal{W}$ that lies in B consists of $k \geq 0$ arcs going from γ to $\partial B \setminus \gamma$. These arcs are pairwise disjoint, otherwise they would form a triangle in $B \subseteq D \setminus \{O\}$. The part of $\mathcal{W}$ that lies in A, excluding γ, is a quasi wall system on A with $n - 1$ walls.

Let γ' be the wall of $\mathcal{W}'$ homotopic to γ in $D \setminus \{O\}$. We apply to $\mathcal{W}'$ a first isotopy of $D \setminus \{O\}$ to ensure that $\gamma' = \gamma$. The wall γ' does not cover any other wall β' of $\mathcal{W}'$, as otherwise the wall β of $\mathcal{W}$ homotopic to β' would cross γ twice. Similarly to $\mathcal{W}$, the part of $\mathcal{W}'$ lying in B consists of k pairwise disjoint arcs going from γ to $\partial B \setminus \gamma$. Thus, by applying a second isotopy, we may ensure that $\mathcal{W}' \cap B = \mathcal{W} \cap B$. Finally, we get $(\mathcal{W}' \setminus \gamma') \cap A = (\mathcal{W} \setminus \gamma) \cap A$ by applying an isotopy of the disk A fixing O, whose existence is guaranteed by the inductive hypothesis. ∎

Now, the walls of $\mathcal{W}$ do not bound any triangle in $D \setminus \{O\}$ (where each side lies in a wall); see Figures 6 and 5. Since two arcs of $\mathcal{W}$ intersect each other if and only if the same holds for the corresponding homotopic arcs of $\mathcal{W}'$, we deduce that the walls of $\mathcal{W}'$ do not form any triangle in $D \setminus \{O\}$ either. The uniqueness of the simple discrete metric disk of minimal area up to isotopy of the disk fixing O follows from Lemma 10.2. ∎

Remark 10.3. The isotopy between $\mathcal{W}$ and $\mathcal{W}'$ can also be derived from [27], where it is proved that two wall systems on a closed surface which are homotopic to each other and are both in minimally crossing position (i.e., they attain the minimum number of self-intersections possible in their homotopy class) can be obtained from each other by isotopies and triangle flip moves (called "type III moves" in [27]). Strictly speaking, we first need to adapt this result to quasi wall systems on surfaces with boundary. Since $\mathcal{W}$ and $\mathcal{W}'$ do not form any triangle in $D \setminus \{0\}$, we conclude that they are isotopic in D.

11. Construction of almost minimizing Finsler disks

In this section, we construct a Finsler disk of radius r with minimizing interior geodesics whose area is arbitrarily close to the lower bound $\frac{6}{\pi}r^2$ given by Theorem 1.2.

Let us first go over Busemann's construction of projective metrics in relation with Hilbert's fourth problem. We refer to [2, 18, 37, 38, 41] and references therein for an account on the subject.

The space Γ of oriented lines in $\mathbb{R}^2$ can be identified with $S^1 \times \mathbb{R}$. Under this identification, an oriented line γ is represented by a pair $(e^{i\theta}, p)$ where $e^{i\theta}$ is the direction of the oriented line γ and $p = \langle \overrightarrow{OH} \times e^{i\theta}, \overrightarrow{e_z} \rangle$ is the signed distance from the origin O to γ. Here, H is a point of γ, the vector $\overrightarrow{e_z}$ is the third vector in the canonical basis of $\mathbb{R}^3$, thus it is a unit vector orthogonal to $\mathbb{R}^2$, and "$\times$" is the vector product in $\mathbb{R}^3$.

Definition 11.1. Let μ be a (nonnegative) Borel measure on Γ. Consider the following conditions:

(1) the measure is invariant under the involution of Γ reversing the orientation of lines;

(2) the measure of every compact subset of Γ is finite;

(3) the set of all oriented lines passing through any given point of $\mathbb{R}^2$ has measure zero;

(4) the set of all oriented lines passing through any given line segment in $\mathbb{R}^2$ has positive measure.

A Borel measure μ satisfying (1)–(3) induces a length function

$$\mathrm{length}_\mu(\alpha) = \frac{1}{4} \int_{\gamma \in \Gamma} \#(\gamma \cap \alpha)\, \mathrm{d}\mu(\gamma)$$

defined for any curve α in the plane $\mathbb{R}^2$. For this kind of length function, straight segments are shortest paths, therefore the pseudo-distance associated to this length function is

$$d_\mu(x, y) = \frac{1}{4} \int_{\gamma \in \Gamma} \#(\gamma \cap [x, y])\, \mathrm{d}\mu(\gamma) = \frac{1}{4}\mu(\Gamma_{[x,y]}),$$

where Γ_A denotes the set of lines $\gamma \in \Gamma$ that intersect a subset or point A contained in the plane $\mathbb{R}^2$. The pseudo-distance d_μ is *projective*, which means that $d(x, z) = d(x, y) + d(y, z)$ for all $x, y, z \in \mathbb{R}^2$ with $y \in [x, z]$, and in fact every continuous projective distance is obtained from a unique measure μ; see [2]. If μ also satisfies (4) then d_μ is a projective distance (and vice versa).

For example, the product measure λ, given by $\mathrm{d}\lambda = \mathrm{d}\theta\, \mathrm{d}p$, yields the Euclidean distance $d_\lambda(x, y) = |y - x|$.

The projective distance induced by a Borel measure satisfying (1)–(4) is not Finsler in general. Borel measures inducing a Finsler metric can be characterized as follows (see [4] for a presentation of this result due to Pogorelov [38], and [7] for a generalization).

Theorem 11.2. *Let μ be a Borel measure on Γ satisfying* (1)–(4)*. The distance d_μ is Finsler if and only if μ is a positive smooth measure. In this case, the smooth measure on Γ induced by the symplectic form associated to the Finsler metric* (*see* (3.8)) *coincides with μ.*

Here, a measure μ on Γ is *(positive) smooth* if it admits a (positive) smooth function h as density, that is, $\mathrm{d}\mu = h\,\mathrm{d}\lambda$.

Remark 11.3. The geodesics of a plane with a projective Finsler metric d_μ are the straight lines parametrized by μ-length. Therefore, a plane with a projective Finsler metric has minimizing geodesics.

We may define the area of a Borel set D in the plane with a measure μ on Γ satisfying (1)–(3) by the Santaló+Blaschke formula (3.6),

$$\mathrm{area}_\mu(D) = \frac{1}{8\pi}\int_{\gamma_0\in\Gamma}\int_{\gamma_1\in\Gamma} \#(\gamma_0\cap\gamma_1\cap D)\,\mathrm{d}\mu(\gamma_1)\,\mathrm{d}\mu(\gamma_0). \tag{11.1}$$

In other terms, the area measure is the normalized pushforward measure

$$\mathrm{area}_\mu = \frac{1}{8\pi} i_*(\mu\times\mu), \tag{11.2}$$

where $i : \Gamma\times\Gamma\setminus\Delta_\Gamma \to \mathbb{R}P^2$ maps each ordered pair of different lines to its intersection point in the projective plane $\mathbb{R}P^2 \supseteq \mathbb{R}^2$. (Note that the diagonal Δ_Γ has measure zero because μ has no atoms.) This area function coincides with Holmes–Thompson area if the metric is Finsler (see (3.6)).

11.1. Construction of a non-Finsler extremal disk

Let us construct a non-Finsler projective pseudo-metric disk satisfying the equality case in Theorem 1.2. Consider the three pairs of one-parameter families $L_k^\pm$ of oriented lines in $\mathbb{R}^2$ defined as

$$\begin{aligned} L_k^+ &: \mathbb{R}_+ \to \Gamma = S^1\times\mathbb{R}, & L_k^- &: \mathbb{R}_+ \to \Gamma = S^1\times\mathbb{R},\\ t &\mapsto (\mathrm{e}^{\mathrm{i}\frac{2k\pi}{3}}, t), & t &\mapsto (\mathrm{e}^{\mathrm{i}(\frac{2k\pi}{3}+\pi)}, -t), \end{aligned}$$

where $k\in\{0,1,2\}$; see Figure 7. Note that the lines $L_k^+(t)$ and $L_k^-(t)$ only differ in their orientation. We will sometimes denote these families of lines by L_k when the orientation does not matter. Consider the (nonsmooth) Borel measure on Γ

$$\mu_{\mathrm{ext}} = \nu_0 + \nu_1 + \nu_2,$$

where

$$\nu_k = \tfrac{1}{2}[(L_k^+)_*(\mathcal{L}) + (L_k^-)_*(\mathcal{L})]$$

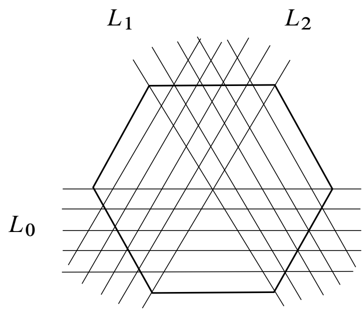

Fig. 7. Extremal pseudo-metric disk.

is the average of the pushforwards to Γ of the Lebesgue measure $\mathcal{L}$ on $\mathbb{R}_+$. Let D_k be the line passing through O orthogonal to L_k. Let $\bar{D}_k \subseteq D_k$ be the ray from O that intersects orthogonally every line $L_k(t)$. Denote by π_k the orthogonal projection of $\mathbb{R}^2$ to D_k. By construction, the d_{ν_k}-pseudo-distance between two points $x, y \in \mathbb{R}^2$ is equal to one quarter of the Euclidean length of the projection of $[x, y]$ to D_k lying in $\bar{D}_k$. That is,

$$d_{\nu_k}(x, y) = \tfrac{1}{4} \operatorname{length}(\pi_k([x, y]) \cap \bar{D}_k) \leq \tfrac{1}{4}|x - y|$$

for all $x, y \in \mathbb{R}^2$. Furthermore,

$$d_{\mu_{\mathrm{ext}}}(x, y) = \sum_{k=0,1,2} d_{\nu_k}(x, y) = \frac{1}{4} \sum_{k=0,1,2} \operatorname{length}(\pi_k([x, y]) \cap \bar{D}_k).$$

Observe also that the measure μ_{ext} satisfies (1)–(3), but not (4). Thus, $d_{\mu_{\mathrm{ext}}}$ is a projective pseudo-distance on $\mathbb{R}^2$.

The disk $D_{\mu_{\mathrm{ext}}}(r)$ of radius r for the pseudo-distance $d_{\mu_{\mathrm{ext}}}$ with center the origin O of $\mathbb{R}^2$ is the minimal regular hexagon containing the Euclidean disk of radius $4r$, whose vertices are $\frac{2\sqrt{3}}{3} 4r e^{ik\frac{\pi}{3}}$ for $k \in \{0, \dots, 5\}$; see Figure 7. A direct computation using (11.1) shows that its area is $\frac{6}{\pi} r^2$. Thus, the disk $D_{\mu_{\mathrm{ext}}}(r)$ is a non-Finsler projective pseudo-metric disk satisfying the equality case in Theorem 1.2. One can think of it as an extremal (degenerate) metric for the problem considered. Observe also that $D_{\mu_{\mathrm{ext}}}(r)$ is not rotationally symmetric.

Remark 11.4. By identifying all pairs of points at zero pseudo-distance, the pseudo-metric disk $D_{\mu_{\mathrm{ext}}}(r)$ identifies with the closed ball $D(r)$ of radius r centered at the tip of a cone composed of three copies of a quadrant of the ℓ^1-plane glued together. It follows from a direct computation that the Holmes–Thompson area of the disk $D(r)$ is equal to $\frac{6}{\pi} r^2$. Defined in this way, the metric on $D(r)$ is non-Finsler (e.g., it has a singularity at the origin and the tangent norms are neither smooth nor strongly convex) but can still be thought of as an extremal (degenerate) metric. Note that the (pseudo-)metrics on $D_{\mu_{\mathrm{ext}}}(r)$ and $D(r)$ can be viewed as continuous versions of the extremal simple discrete disk (see Section 10).

11.2. Construction of a Finsler nearly extremal disk

In the rest of this section, we explain how to modify the pseudo-metric $d_{\mu_{\text{ext}}}$ so as to obtain a projective Finsler disk of radius r whose area is arbitrarily close to $\frac{6}{\pi}r^2$. First, the projective pseudo-metric $d_{\mu_{\text{ext}}}$ can be approximated by a projective metric by simply adding to μ a multiple $\varepsilon\lambda$ of the uniform measure λ (given by $\mathrm{d}\lambda = \mathrm{d}\theta\,\mathrm{d}p$) so that (4) is also satisfied; this changes $d_{\mu_{\text{ext}}}$ by adding ε times the Euclidean distance. This projective metric is not Finsler, but in turn it can be approximated by a Finsler metric; see [38]. More generally, every projective distance d_μ, where μ is a Borel measure satisfying (1)–(4), can be approximated by a projective Finsler distance on every compact subset of $\mathbb{R}^2$. Thus, by Theorem 11.2, there exists a sequence μ_n of positive smooth measures on Γ such that the corresponding sequence of Finsler distances d_{μ_n} uniformly converges to $d_{\mu_{\text{ext}}}$ on every compact set of $\mathbb{R}^2$. This approximation result is obtained by a convolution argument on the distance function d_μ. Although it is possible that the measures μ_n weakly converge to μ_{ext}, this issue is not addressed in [38]. This leads us to consider a slightly different approach. Instead of regularizing the distance function, we smooth out the measure μ_{ext} and show that the corresponding projective Finsler distance converges to $d_{\mu_{\text{ext}}}$. This alternative approach to the regularization of a projective distance provides a weak convergence of measure by construction, which allows us to estimate areas as well as distances.

We proceed as follows. First, we truncate the measure μ_{ext} by setting a bound for the absolute value of the p coordinate of the lines $\gamma \in \Gamma$. In this way, we obtain a probability measure μ_0 on Γ, without changing the corresponding distance function in a neighborhood of the origin. Similarly, we truncate the uniform measure λ to a probability measure λ_0. This enables us to use standard theorems on weak convergence of probability measures.

Let us now describe the convolution process. For $\varepsilon > 0$, let h_ε be a smooth nonnegative function on $\Gamma = \mathbb{R}/2\pi\mathbb{Z}\times\mathbb{R}$, with support in $(-\varepsilon,\varepsilon)\times(-\varepsilon,\varepsilon)$, such that $\int_\Gamma h_\varepsilon(\theta,p)\,\mathrm{d}\theta\,\mathrm{d}p = 1$. For each $\varepsilon > 0$, consider the positive smooth measure μ_ε on Γ with density $h_\varepsilon * \mu_0$, that is,

$$\mathrm{d}\mu_\varepsilon = (h_\varepsilon * \mu_0)\,\mathrm{d}\lambda,$$

where $h_\varepsilon * \mu_0$ is the smooth function on Γ defined by the convolution

$$h_\varepsilon * \mu_0(\gamma) = \int_\Gamma h_\varepsilon(\gamma - \gamma')\,\mathrm{d}\mu_0(\gamma')$$

and λ is the standard product measure on $\Gamma = \mathbb{R}/2\pi\mathbb{Z} \times \mathbb{R}$, given by $\mathrm{d}\lambda = \mathrm{d}\theta\,\mathrm{d}p$. By [16, §1.4.3], the smooth measure μ_ε weakly converges to μ_0 as $\varepsilon \to 0$. Define also the measure

$$\mu_\varepsilon^+ = (1-\varepsilon)\mu_\varepsilon + \varepsilon\lambda_0,$$

which also converges to μ_0 as $\varepsilon \to 0$. By Theorem 11.2, the distance $d_{\mu_\varepsilon^+}$ induced by μ_ε^+ is a projective Finsler distance on a neighborhood of the origin in $\mathbb{R}^2$.

To approximate distances and areas, we have the following tools.

Lemma 11.5. *Let μ and μ_n be probability measures on Γ satisfying conditions* (1)–(3) *of Definition* 11.1. *If μ_n weakly converges to μ, then the distance d_{μ_n} converges uniformly to d_μ on every compact subset of $\mathbb{R}^2$.*

Proof. Note first that the distance between two points $x, y \in \mathbb{R}^2$ is

$$d_\mu(x, y) = \mu(\Gamma_{[x,y]}),$$

where $\Gamma_{[x,y]}$ denotes the set of lines that intersect the segment $[x, y]$. Thus, for a specific pair of points x, y, the weak convergence $\mu_n \to \mu$ implies that $d_{\mu_n}(x, y) \to d_\mu(x, y)$ by the Portmanteau theorem [13, Theorem 2.1], since $\Gamma_{[x,y]}$ is a continuity set for μ. That is, its boundary

$$\partial\Gamma_{[x,y]} = \Gamma_x \cup \Gamma_y$$

(where Γ_z is the set of lines that contain a point z) has measure $\mu(\partial\Gamma_{[x,y]}) = 0$ since $\mu(\Gamma_z) = 0$ for each point $z \in \mathbb{R}^2$ by condition (3) on μ.

To show that this convergence holds uniformly for x, y in any given compact set $K \subseteq \mathbb{R}^2$, let $\mathcal{A}$ be the family of sets $\Gamma_{[x,y]}$ for $x, y \in K$. According to [14, Theorems 2 and 3], to show uniform convergence $\mu_n(A) \to \mu(A)$ for all sets $A \in \mathcal{A}$, it is sufficient to show that $\mu(B_\delta(\partial A)) \to 0$ uniformly as $\delta \to 0$, where $B_\delta(S)$ denotes the δ-neighborhood of a set $S \subseteq \Gamma$ (say, with respect to the supremum distance in terms of the coordinates θ, p). Now,

$$B_\delta(\partial\Gamma_{[x,y]}) = B_\delta(\Gamma_x) \cup B_\delta(\Gamma_y),$$

therefore it suffices to show that $\mu(B_\delta(\Gamma_x)) \to 0$ uniformly for all $x \in K$ as $\delta \to 0$. Suppose that this is not the case. Then there are sequences $\delta_m \to 0$ and $x_m \to x \in K$ such that $\mu(B_{\delta_m}(\Gamma_{x_m}))$ does not tend to zero. However, we also have

$$B_{\delta_m}(\Gamma_{x_m}) \subseteq B_{\delta'_m}(\Gamma_x)$$

for some sequence $\delta'_m \to 0$ (namely, $\delta'_m = \delta_m + |x_m - x|$), which yields $B_{\delta'_m}(\Gamma_x) \to \Gamma_x$ and therefore

$$\mu(B_{\delta_m}(\Gamma_{x_m})) \le \mu(B_{\delta'_m}(\Gamma_x)) \to \mu(\Gamma_x) = 0.$$

This contradiction finishes the proof. ∎

Lemma 11.6. *Let μ and μ_n be probability measures on Γ satisfying conditions* (1)–(3) *of Definition* 11.1. *If μ_n weakly converges to μ, then* $\text{area}_{\mu_n}(K)$ *converges to* $\text{area}_\mu(K)$ *for every compact set $K \subseteq \mathbb{R}^2$ such that $\mu(\partial K) = 0$.*

Proof. We will use some properties of weak convergence of probability measures; see [13]. Since μ_n weakly converges to μ, it follows from [13, Example 3.2] that the product measure $\mu_n \times \mu_n$ converges weakly to $\mu \times \mu$ on $\Gamma \times \Gamma$. Restricting to the set $\Gamma \times \Gamma \setminus \Delta_\Gamma$, the measures $\mu \times \mu$ and $\mu_n \times \mu_n$ are still probability measures since the diagonal $\Delta_\Gamma \subseteq \Gamma \times \Gamma$ has zero measure because μ and μ_n have no atoms. Moreover, since the diagonal Δ_Γ is a closed set, the product measure $\mu_n \times \mu_n$ weakly converges to $\mu \times \mu$ on $\Gamma \times \Gamma \setminus \Delta_\Gamma$ by condition (iv) of the Portmanteau theorem [13, Theorem 2.1].

Furthermore, since the function $i : \Gamma \times \Gamma \setminus \Delta_\Gamma \to \mathbb{R}P^2$ is continuous, the pushforward measure $i_*(\mu_n \times \mu_n)$ weakly converges to $i_*(\mu \times \mu)$ by the definition of weak convergence; see [13, p. 14]. Therefore, the area measure $\text{area}_{\mu_n} = \frac{1}{8\pi} i_*(\mu_n \times \mu_n)$ weakly converges to area_μ, with both area measures considered as probability measures on the projective plane $\mathbb{R}P^2$; see (11.2). Finally, to show that $\text{area}_{\mu_n}(K) \to \text{area}_\mu(K)$, we must check, according to part (v) of the Portmanteau theorem, that K is a continuity set of area_μ, which by definition means that $\mu(\partial_{\mathbb{R}P^2} K) = 0$. This follows from the facts that K is compact and $\mu(\partial_{\mathbb{R}^2} K) = 0$. ■

Consider the disk $D_{\mu_\varepsilon^+}(r)$ centered at O of radius r for the distance $d_{\mu_\varepsilon^+}$. The numbers $r > 0$ and $\varepsilon > 0$ are small enough so that the truncations of μ_{ext} and λ have no effect on the disk $D_{\mu_\varepsilon^+}(r)$. The number r is fixed while ε goes to 0.

Proposition 11.7. *The disk $D_{\mu_\varepsilon^+}(r)$ is a projective Finsler disk with minimizing interior geodesics, whose area converges to $\frac{6}{\pi} r^2$ as $\varepsilon \to 0$. Therefore, the area lower bound in Theorem 1.2 is sharp.*

Proof. The fact that $d_{\mu_\varepsilon^+}$ is a projective Finsler metric follows from Theorem 11.2, and the fact that its geodesics are minimizing was stated in Remark 11.3.

To compute the area of the disk $D_{\mu_\varepsilon^+}(r)$ we proceed as follows. By uniform convergence of the metrics (Lemma 11.5), for every $\delta > 0$ and every $\varepsilon > 0$ small enough, we have

$$D_{\mu_0}(r - \delta) \subseteq D_{\mu_\varepsilon^+}(r) \subseteq D_{\mu_0}(r + \delta).$$

Therefore,

$$\text{area}_{\mu_\varepsilon^+}(D_{\mu_0}(r - \delta)) \leq \text{area}_{\mu_\varepsilon^+}(D_{\mu_\varepsilon^+}(r)) \leq \text{area}_{\mu_\varepsilon^+}(D_{\mu_0}(r + \delta)).$$

Since the sets $D_{\mu_0}(r \pm \delta)$ are compact and have boundary of μ_0-measure zero, Lemma 11.6 shows that

$$\text{area}_{\mu_\varepsilon^+}(D_{\mu_0}(r \pm \delta)) \to \text{area}_{\mu_0}(D_{\mu_0}(r \pm \delta)) \quad \text{as } \varepsilon \to 0.$$

Since this holds for each $\delta > 0$, we conclude that

$$\text{area}_{\mu_\varepsilon^+}(D_{\mu_\varepsilon^+}(r)) \to \text{area}_{\mu_0}(D_{\mu_0}(r)) \quad \text{as } \varepsilon \to 0.$$ ■

12. Appendix: Differentiability of distance-realizing paths on Finsler surfaces with boundary

Consider a smooth manifold M with smooth boundary endowed with a Finsler metric F. Recall that a distance-realizing curve is a curve $\alpha : I \to M$ defined on an interval $I \subseteq \mathbb{R}$ such that

$$d_F(\alpha(t), \alpha(t')) = t' - t$$

for all $t < t'$.

If the manifold has an empty boundary (or, more generally, a convex boundary), then its distance-realizing curves satisfy a differential equation, and it is therefore clear that they are smooth. However, if the boundary is not convex, then the distance-realizing curves are not C^2 in general, and they are not even determined by their initial velocity vector. This happens, for instance, on the Euclidean plane minus an open disk.

In the case of Riemannian manifolds with boundary, it was claimed in [3, 43] that distance-realizing curves are C^1. This result can also be recovered from [35] by gluing together two copies of a Riemannian manifold M along their boundaries. The Riemannian metric obtained on the resulting double manifold N is α-Hölder continuous for any $\alpha \in (0, 1]$; see [35, Example 3.3]. By [35], the geodesics on N are C^1 (and even $C^{1,\frac{\alpha}{2-\alpha}}$), from which we can deduce that the distance-realizing curves on M are also C^1. This argument does not hold for Finsler metrics. Indeed, the double of a Finsler metric is not even a continuous Finsler metric in general.

Here, by adapting the argument of [3], we prove that the same result holds for Finsler surfaces.

Theorem 12.1. *On a Finsler surface M with boundary, every distance-realizing curve $\alpha : I \to M$ is C^1. Furthermore, the velocity vectors $\alpha'(t)$ have unit norm.*

Let us introduce some technical definitions. We assume without loss of generality that the surface M is the closed upper half-space of $\mathbb{R}^2$.

Definition 12.2. Let $\alpha : I \to M$ be a continuous curve, where $I \subseteq \mathbb{R}$ is an interval. Fix $t_0 \in I$ and denote $x_0 = \alpha(t_0)$. An *arrival velocity* of α at t_0 is a vector $v \in T_{x_0}M$ that is an accumulation point of the set of vectors

$$V^- = \left\{ \frac{\alpha(t) - \alpha(t_0)}{t - t_0} \;\middle|\; t < t_0 \right\}$$

as $t \to t_0$. Similarly, a *departure velocity* of α at t_0 is a vector $v \in T_{x_0}M$ that is an accumulation point of the set of vectors

$$V^+ = \left\{ \frac{\alpha(t) - \alpha(t_0)}{t - t_0} \;\middle|\; t > t_0 \right\}$$

as $t \to t_0$. Note that if α is left (resp. right) differentiable at t_0, then α has exactly one arrival (resp. departure) velocity at t_0.

We begin by proving a weak differentiability result.

Lemma 12.3. *Let (M, F) be a Finsler manifold with boundary and let $\alpha : I \to M$ be a distance-realizing curve. Fix $t_0 \in I$ and denote $x_0 = \alpha(t_0)$. Then*

(1) *The curve α has at least one arrival velocity and one departure velocity at t_0 (unless $t_0 = \min I$ or $t_0 = \max I$, respectively).*
(2) *Every arrival or departure velocity v has norm $F_{x_0}(v) = 1$.*
(3) *If the curve α is differentiable on one side at an interior point t_0 of I, then α is differentiable at t_0.*

Proof. By continuity of the Finsler metric at x_0, we can bound F_x below and above by two multiples of the norm $F_{x_0} = |\cdot|$ for every x close enough to x_0. That is,

$$\lambda^- |v| \leq F_x(v) \leq \lambda^+ |v|$$

for every $v \in \mathbb{R}^n$, which in turn implies that

$$\lambda^- |x - x_0| \leq d_F(x_0, x) \leq \lambda^+ |x - x_0|.$$

This implies that the sets $V^{\pm}$ are bounded when $t \to t_0$, which implies the first claim. In fact, as $x \to x_0$, the optimal coefficients $\lambda^{\pm}$ converge to 1, which implies the second claim.

To prove the last claim, we assume that the curve α is left differentiable at an interior point t_0 of I. (The argument is similar if α is right differentiable at t_0.) Let v^- be the arrival tangent vector. Let us prove that α is right differentiable at t_0 and has departure tangent vector $v^+ = v^-$. For contradiction, assume that the set of vectors V^+ has an accumulation point $v^+ \neq v^-$ as $t \to t_0$. As already noticed in the second claim, we have $|v^-| = |v^+| = 1$. Since the norm $F_{x_0} = |\cdot|$ is strictly convex, we also have $|v^- + v^+| < 2$. Let $\tau_m \to 0$ be a decreasing sequence of positive numbers such that

$$y_m^+ = \alpha(t_0 + \tau_m) = \alpha(t_0) + \tau_m v^+ + o(\tau_m).$$

Since α is left differentiable at t_0, we also have

$$y_m^- = \alpha(t_0 - \tau_m) = \alpha(t_0) - \tau_m v^- + o(\tau_m).$$

Thus,

$$d_F(y_m^+, y_m^-) \leq \lambda^+ |y_m^+ - y_m^-| = \lambda^+ \tau_m |v^+ + v^- + o(1)|.$$

For m large enough, we can take λ^+ arbitrarily close to 1. It follows from the inequality $|v^+ + v^-| < 2$ that

$$d_F(y_m^-, y_m^+) < 2\tau_m,$$

contradicting the assumption that α is a distance-realizing curve. ∎

Before proceeding to the proof of Theorem 12.1, we extend the Finsler metric F to a surface $M^+ \supseteq M$ with empty boundary; see Remark 2.2. As for any Finsler surface with empty boundary, every point of M^+ has a normal neighborhood, that is, an open neighborhood U such that for any $x, y \in U$, there is a unique geodesic from x to y contained in U and this geodesic is the unique distance-realizing arc from x to y in M^+; see [9, p. 160]. Note that if this geodesic is contained in M, then it is also the unique distance-realizing arc from x to y in M.

Proof of Theorem 12.1. We assume first that the metric is self-reverse.

Let $\alpha : I \to M$ be a distance-realizing curve. Let $t_0 \in I$ and let $x_0 = \alpha(t_0)$. If $x_0 = \alpha(t_0)$ lies in the interior of M then the arc α coincides with a geodesic in a neighborhood of t_0, where it is C^1 (and we are done). Thus, we can assume that x_0 lies in ∂M.

Again, we assume without loss of generality by working in a small enough neighborhood of x_0 that M is a closed half-space of $M^+ = \mathbb{R}^2$ and that every geodesic arc is a unique distance-realizing arc.

Suppose that the arc α is not right differentiable at some $t_0 \in I$. (The argument is similar if α is not left differentiable at t_0.) The arc α has two departure velocities v and w. Let K_v and K_w be two convex cones based at x_0 that contain the points $x_0 + v$ and $x_0 + w$ in their interior and only meet at x_0. Take a unit vector $u \in T_{x_0}M$ not tangent to ∂M that points into the interior of M and separates K_v from K_w, and denote by γ_u the geodesic with initial velocity $\gamma_u'(t_0) = u$. This geodesic does not visit K_v or K_w in some interval (t_0, t_2). On the other hand, the arc $\alpha(t)$ visits the cones K_v and K_w infinitely many times in any interval (t_0, τ) with $\tau > t_0$. Therefore, it must cross the geodesic γ_u at some time $t_1 \in (t_0, t_2)$. Since γ_u is the unique distance-realizing path between any of its points, the arc α coincides with γ_u in $[t_0, t_1]$. Thus α does not visit K_v and K_w in (t_0, t_1). This contradiction proves that α is right differentiable at t_0. It follows from Lemma 12.3 that α is differentiable at every interior point $t_0 \in I$.

Suppose α is not C^1 on the right at t_0. (The argument is similar in case it is not C^1 on the left.) The vector $v = \alpha'(t_0)$ points inside M or is tangent to the boundary of M. Since the velocities $\alpha'(t)$ are unit vectors and the curve α is not C^1 on the right at t_0, its derivative α' has an accumulation point $w \neq v$ when t goes to t_0 from the right. Let u be a unit vector spanning a line that separates v from w. Consider three disjoint neighborhoods U, V, W of u, v, w such that for any u', v', w' in U, V, W respectively, the line spanned by u' separates v' from w'. Let K_V be the union of the rays contained in M starting at x_0 with direction $v' \in V$, and let R be any of these rays. Note that u is transverse to all these rays. Working in a small enough neighborhood of x_0, we can assume that the family Γ of geodesics that visit R with velocity u foliates the cone K_V, and that their tangent vectors do not deviate too much from u and thus lie in U. Since the velocity of α at t_0 lies in the open set V, the arc α restricted to some nontrivial interval $[t_0, t_3)$ lies in K_V. Now, since w is an accumulation point for α' when t goes to t_0 from the right, there exists $t_2 \in (t_0, t_3)$ such that $w' = \alpha'(t_2)$ lies in W. Let $x_2 = \alpha(t_2)$, and let v' be the direction from x_0 to x_2. Let γ be the geodesic of Γ passing through x_2, and let $u' \in U$ be its velocity at x_2. The vector $w' = \alpha'(t_2)$ points strictly inside the region of M bounded by γ containing x_0, since the vector v' points outside, and the line generated by the vector u' separates v' from w'.

Therefore, the arc α starting at x_0 must cross γ a first time at $t_1 \in (t_0, t_2)$ before crossing it again at t_2. Since γ is the unique distance-realizing path between $\alpha(t_1)$ and $\alpha(t_2)$, the arc α coincides with γ in $[t_1, t_2]$, which contradicts the fact that α' is transverse to γ at t_2 (or t_1). This finishes the proof of Theorem 12.1 for self-reverse metrics.

In the case of directed metrics we adapt the argument as follows. Apart from the foliation Γ, we need a second foliation Γ^- of K_V by geodesics transverse to the ray R with initial velocity $-u$. Then we proceed as in the proof above and after choosing the point x_1 in K_V, we let γ and γ^- be the two geodesics of Γ and Γ^- passing through x_1. We keep only the part of each geodesic before it reaches x_1 and discard the rest. These two half-geodesics bound a region of K_V containing x_0. The curve α points strictly inside this region at x_1. Therefore, it must cross either γ or γ^- a first time before reaching x_1. We derive a contradiction as in the previous proof. ■

Acknowledgments. The first author thanks the Laboratoire d'analyse et de mathématiques appliquées at the Université Gustave Eiffel/Université Paris Est Créteil and the Groupe Troyanov at the École Polytechnique Fédérale de Lausanne for hosting him as a postdoc while this work was done. The second author would like to thank the Fields Institute and the Department of Mathematics at the University of Toronto, where part of this work was accomplished, for their hospitality. The authors thank the referees for their comments, which helped improve the exposition.

Funding. Partially supported by the Bézout Labex (ANR-10-LABX-58) and the ANR project Min-Max (ANR-19-CE40-0014). This funding information was added in the final, issue version of the paper on March 19, 2024. By the publisher's fault, it was omitted in the online first version. We apologise to the authors and readers for the inconvenience.

References

[1] Abraham, R., Marsden, J. E.: Foundations of mechanics. Benjamin/Cummings, Advanced Book Program, Reading, MA (1978) Zbl 0393.70001 MR 515141

[2] Alexander, R.: Planes for which the lines are the shortest paths between points. Illinois J. Math. **22**, 177–190 (1978) Zbl 0379.50002 MR 490820

[3] Alexander, R., Alexander, S.: Geodesics in Riemannian manifolds-with-boundary. Indiana Univ. Math. J. **30**, 481–488 (1981) Zbl 0469.53039 MR 620261

[4] Álvarez Paiva, J. C.: Symplectic geometry and Hilbert's fourth problem. J. Differential Geom. **69**, 353–378 (2005) Zbl 1088.53047 MR 2169868

[5] Álvarez Paiva, J. C.: Some problems on Finsler geometry. In: Handbook of differential geometry. Vol. II, Elsevier/North-Holland, Amsterdam, 1–33 (2006) Zbl 1147.53059 MR 2194667

[6] Álvarez Paiva, J. C., Berck, G.: What is wrong with the Hausdorff measure in Finsler spaces. Adv. Math. **204**, 647–663 (2006) Zbl 1099.53018 MR 2249627

[7] Álvarez Paiva, J.-C., Berck, G.: Finsler surfaces with prescribed geodesics. arXiv:1002.0243 (2010)

[8] Arnol'd, V. I.: Mathematical methods of classical mechanics. Grad. Texts in Math. 60, Springer, New York (1989) Zbl 0692.70003 MR 1345386

[9] Bao, D., Chern, S.-S., Shen, Z.: An introduction to Riemann–Finsler geometry. Grad. Texts in Math. 200, Springer, New York (2000) Zbl 0954.53001 MR 1747675

[10] Berger, M.: Some relations between volume, injectivity radius, and convexity radius in Riemannian manifolds. In: Differential geometry and relativity, Math. Phys. Appl. Math. 3, Reidel, Dordrecht, 33–42 (1976) Zbl 0342.53038 MR 0448253

[11] Berger, M.: Volume et rayon d'injectivité dans les variétés riemanniennes de dimension 3. Osaka Math. J. **14**, 191–200 (1977) Zbl 0353.53028 MR 467595

[12] Berger, M.: Une borne inférieure pour le volume d'une variété riemannienne en fonction du rayon d'injectivité. Ann. Inst. Fourier (Grenoble) **30**, 259–265 (1980) Zbl 0421.53028 MR 597027

[13] Billingsley, P.: Convergence of probability measures. 2nd ed., Wiley, New York (1999) Zbl 0944.60003 MR 1700749

[14] Billingsley, P., Topsøe, F.: Uniformity in weak convergence. Z. Wahrsch. Verw. Gebiete **7**, 1–16 (1967) Zbl 0147.15701 MR 209428

[15] Blaschke, W.: Zur Variationsrechnung. Abh. Math. Sem. Univ. Hamburg **11**, 359–366 (1935) Zbl 0056.32902 MR 3069662

[16] Bogachev, V. I.: Weak convergence of measures. Math. Surveys Monogr. 234, Amer. Math. Soc., Providence, RI (2018) Zbl 1412.60003 MR 3837546

[17] Burago, D., Burago, Y., Ivanov, S.: A course in metric geometry. Grad. Stud. Math. 33, Amer. Math. Soc., Providence, RI (2001) Zbl 0981.51016 MR 1835418

[18] Busemann, H.: Problem IV: Desarguesian spaces. In: Mathematical developments arising from Hilbert problems (De Kalb, IL, 1974), Proc. Sympos. Pure Math. 28, Amer. Math. Soc., Providence, RI, 131–141 (1976) Zbl 0352.50010 MR 0430935

[19] Chambers, G. R., Croke, C., Liokumovich, Y., Wen, H.: Area of convex disks. Proc. Amer. Math. Soc. (to appear); arXiv:1701.06594

[20] Cossarini, M.: Discrete surfaces with length and area and minimal fillings of the circle. Ph.D. thesis, IMPA (2009)

[21] Cossarini, M.: Finsler geodesics: a square-free approach. In preparation (2020)

[22] Croke, C. B.: Some isoperimetric inequalities and eigenvalue estimates. Ann. Sci. École Norm. Sup. (4) **13**, 419–435 (1980) Zbl 0465.53032 MR 608287

[23] Croke, C. B.: On the volume of metric balls. Proc. Amer. Math. Soc. **88**, 660–664 (1983) Zbl 0521.53044 MR 702295

[24] Croke, C. B.: Curvature free volume estimates. Invent. Math. **76**, 515–521 (1984) Zbl 0552.53016 MR 746540

[25] Croke, C. B.: Area of small disks. Bull. London Math. Soc. **41**, 701–708 (2009) Zbl 1177.53041 MR 2521366

[26] Croke, C. B., Katz, M.: Universal volume bounds in Riemannian manifolds. In: Surveys in differential geometry, Vol. VIII (Boston, MA, 2002), Surv. Differ. Geom. 8, Int. Press, Somerville, MA, 109–137 (2003) Zbl 1051.53026 MR 2039987

[27] de Graaf, M., Schrijver, A.: Making curves minimally crossing by Reidemeister moves. J. Combin. Theory Ser. B **70**, 134–156 (1997) Zbl 0888.57001 MR 1441263

[28] Dieudonné, J.: Treatise on analysis. Vol. III. Pure Appl. Math. 10-III, Academic Press, New York (1972) Zbl 0268.58001 MR 0350769

[29] Durán, C. E.: A volume comparison theorem for Finsler manifolds. Proc. Amer. Math. Soc. **126**, 3079–3082 (1998) Zbl 0936.53046 MR 1473664

[30] Federer, H.: Geometric measure theory. Grundlehren Math. Wiss. 153, Springer, New York (1969) Zbl 0176.00801 MR 0257325

[31] Felsner, S., Goodman, J. E.: Pseudoline arrangements. In: Handbook of discrete and computational geometry, 3rd ed., 125–157 (2017)

[32] Gromov, M.: Metric structures for Riemannian and non-Riemannian spaces. Modern Birkhäuser Classics, Birkhäuser Boston, Boston, MA (2007) Zbl 1113.53001 MR 2307192

[33] Ivanov, S. V.: On two-dimensional minimal fillings. Algebra i Analiz **13**, 26–38 (2001) (in Russian); English transl.: St. Petersburg Math. J. **13**, 17–25 (2002) Zbl 0995.58010 MR 1819361

[34] Ivanov, S. V.: Filling minimality of Finslerian 2-discs. Trudy Mat. Inst. Steklova **273**, 192–206 (2011) (in Russian); English transl.: Proc. Steklov Inst. Math. **273**, 176–190 (2011) Zbl 1241.53062 MR 2893545

[35] Lytchak, A., Yaman, A.: On Hölder continuous Riemannian and Finsler metrics. Trans. Amer. Math. Soc. **358**, 2917–2926 (2006) Zbl 1091.53048 MR 2216252

[36] Mennucci, A. C. G.: Geodesics in asymmetric metric spaces. Anal. Geom. Metr. Spaces **2**, 115–153 (2014) Zbl 1310.53039 MR 3210893

[37] Papadopoulos, A.: Hilbert's fourth problem. In: Handbook of Hilbert geometry, IRMA Lect. Math. Theor. Phys. 22, Eur. Math. Soc., Zürich, 391–431 (2014) MR 3329889

[38] Pogorelov, A. V.: Hilbert's fourth problem. Scripta Ser. Math., V. H. Winston, Washington, DC, and Wiley, New York (1979) MR 550440

[39] Santaló, L. A.: Measure of sets of geodesics in a Riemannian space and applications to integral formulas in elliptic and hyperbolic spaces. Summa Brasil. Math. **3**, 1–11 (1952) Zbl 0049.23804 MR 56318

[40] Santaló, L. A.: Integral geometry and geometric probability. Encyclopedia Math. Appl. 1, Addison-Wesley, Reading, MA (1976) Zbl 0342.53049 MR 0433364

[41] Szabó, Z. I.: Hilbert's fourth problem. I. Adv. Math. **59**, 185–301 (1986) Zbl 0608.53062 MR 835025

[42] Tao, T.: Poincaré's legacies, pages from year two of a mathematical blog. Part II. Amer. Math. Soc., Providence, RI (2009) Zbl 1175.00010 MR 2541289

[43] Wolter, F.: Interior metric, shortest paths and loops in riemannian mainfolds with not necessarily smooth boundary. Diplomarbeit, Technische Universität Berlin (1979)