REPORTS@SCM
AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

Survey on optimal isosystolic inequalities on the real projective plane

*Unai Lejarza Alonso
Universitat Autònoma de Barcelona
ulejarzalonsogjm@gmail.com
*Corresponding author

Resum *(CAT)*
Es revisen totes les desigualtats isosistòliques òptimes conegudes al pla projectiu real $\mathbb{RP}^2$, comparant-les amb el cas del 2-tor $\mathbb{T}^2$. Primer s'introdueixen nocions bàsiques de mètriques de Finsler. Després s'enuncien totes les desigualtats isosistòliques conegudes pel cas reversible i se'n dona la idea de prova. Finalment es tracten les desigualtats òptimes pel cas no-reversible. Actualment es coneixen totes les desigualtats òptimes per $\mathbb{T}^2$, tot i que no és així per $\mathbb{RP}^2$. S'hi presenten alguns petits progressos i arguments a favor de la desigualtat conjecturada en el cas encara obert.

Abstract *(ENG)*
All known optimal isosystolic inequalities on the real projective plane $\mathbb{RP}^2$ are surveyed, comparing them to the case of the 2-torus $\mathbb{T}^2$. First, basic notions on Finsler metrics are introduced. Then, all previously known isosystolic inequalities are stated and a sketch of proof is given in the reversible case. Finally, optimal inequalities in the non-reversible case are discussed. All optimal inequalities are currently known for $\mathbb{T}^2$, although this is not the case for $\mathbb{RP}^2$. Some recent minor advances for $\mathbb{RP}^2$ are presented, and some arguments are given in favour of the conjectured inequality in the remaining open case.

Keywords: *real projective plane, systole, isosystolic inequality, Riemannian metric, Finsler metric, Busemann–Hausdorff area, Holmes–Thompson area.*
MSC (2020): *Primary 53C60, 53C22. Secondary 52B10.*

Received: *June 30, 2024.*
Accepted: *September 4, 2024.*

1. Riemannian and Finsler metrics

Riemannian manifolds, introduced in the second half of the 19th century by Bernhard Riemann, are manifolds endowed with a scalar product on each tangent space. Usually, one works with an n-dimensional smooth manifold M and a Riemannian metric $g_x\colon T_xM \times T_xM \to \mathbb{R}$, denoting a scalar product that varies smoothly with $x \in M$. This scalar product gives rise to a norm on tangent vectors, by setting $\|v\|_x^g = \sqrt{g_x(v,v)}$, and to a length for curves $\gamma\colon [0,1] \to M$, by setting $\ell_g(\gamma) = \int_0^1 \|\gamma'(t)\|_{\gamma(t)}^g \,\mathrm{d}t$. The scalar product can alternatively be represented in a local chart by a collection of $n \times n$ positive definite and symmetric matrices $(g_{ij}(x))_{ij}$. That way, the canonical Riemannian measure $\mathrm{d}v_g$ of (M,g) in this local chart is given by the formula $\mathrm{d}v_g(x) = \sqrt{\det(g_{ij}(x))}\,\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n$.

Finsler manifolds are a generalisation of Riemannian manifolds, where each tangent space is endowed with a norm instead of with a scalar product. These metric structures were first considered in 1918 by Paul Finsler, although the term *Finsler manifold* was coined later by Élie Cartan, in 1934. Usually a norm $\|\cdot\|$ is a map from a vector space to $\mathbb{R}^+ = [0,\infty)$ that fulfils the following conditions: $\|v\| = 0$ only if $v = 0$, $\|\lambda v\| = |\lambda|\|v\|$ for $\lambda \in \mathbb{R}$ and $\|v + v'\| \leq \|v\| + \|v'\|$. In Finsler geometry, non-necessarily symmetric norms are considered more generally by replacing the second property by the condition $\|\lambda v\| = \lambda\|v\|$ for $\lambda \in \mathbb{R}^+$. The structure associated to a varying norm on each tangent space is called a Finsler metric and the norm at some point x is usually denoted by F_x. In analogy to the Riemannian case, one defines the length of a curve $\gamma\colon [0,1] \to M$ by $\ell_F(\gamma) = \int_0^1 F_{\gamma(t)}(\gamma'(t))\,\mathrm{d}t$. However, and in contrast to the Riemannian case, there is no unambiguously defined volume notion for Finsler metrics. Two of the most used ones are the Holmes–Thompson and the Busemann–Hausdorff volumes. The former is related to the standard symplectic form on T^*M, and, the latter, to the Hausdorff measure of a metric space in the symmetric case. From now on, only 2-dimensional manifolds will be considered. Fixing an auxiliary Riemannian metric g on M, the Holmes–Thompson and Busemann–Hausdorff areas are defined as

$$\begin{aligned} \mathrm{area}_{\mathrm{HT}}(M,F) &:= \frac{1}{\pi}\int_M |B_x^\circ|_g \,\mathrm{d}v_g, \text{ and} \\ \mathrm{area}_{\mathrm{BH}}(M,F) &:= \pi\int_M \frac{1}{|B_x|_g}\,\mathrm{d}v_g, \text{ respectively.} \end{aligned} \tag{1}$$

Here, $|B_x|_g$ denotes the Riemannian measure of the unit ball $B_x = \{v \in T_xM \mid F_x(v) \leq 1\}$, and B_x° its polar convex body with respect to g_x. Note that a Finsler metric F is uniquely defined specifying the unit spheres $U_x = \{v \in T_xM \mid F_x(v) = 1\}$ at each point $x \in M$.

Definition 1.1. A Finsler metric F on M is said to be reversible if $F_x(v) = F_x(-v)$ for all $(x,v) \in TM$. In other words, F is said to be reversible if all the unit balls are centrally symmetric.

Since a scalar product induces a symmetric norm on each tangent space, Riemannian metrics are a particular case of Finsler metrics. As sketched in [5, Proposition 3.5], the definitions in (1) are independent of the chosen auxiliary Riemannian g, and an easy consequence of the Blaschke–Santaló inequality is the following.

Proposition 1.2. *If F is a reversible Finsler metric on a manifold M, then* $\mathrm{area}_{\mathrm{BH}}(M,F) \geq \mathrm{area}_{\mathrm{HT}}(M,F)$ *and equality holds if and only if F comes from a Riemannian metric.*

2. Isosystolic inequalities

In either the Riemannian or Finsler case, there is a notion of length of curves, and for closed manifolds that are not simply connected one can define the following notion of systole.

Definition 2.1. The systole of a Finsler closed manifold (M, F) which is not simply connected is defined by

$$\operatorname{sys}(M, F) := \inf\{\ell_F(\gamma) \mid \gamma \text{ is a non-contractible loop in } M\}.$$

One expects that the area of a Finsler manifold for which all non-contractible loops have a length uniformly bounded from below cannot be made arbitrarily small. This is described by an inequality of the form

$$\operatorname{area}(M, F) \geq C \operatorname{sys}^2(M, F)$$

holding for some set of metrics F, where C is some positive constant. Such an inequality is called an *isosystolic inequality* and the constant might depend on the set of metrics considered. Usually one considers either Riemannian metrics, reversible Finsler metrics or all Finsler metrics. An isosystolic inequality is said to be optimal if the constant C cannot be improved. Finally, it is said that there is systolic freedom if such a positive constant does not exist.

The first optimal isosystolic inequality was found for the 2-torus in 1949 by Charles Loewner. As it is explained by his student Pao Ming Pu at the end of [6], Loewner found it during the lectures of a course on Riemannian geometry he was teaching at the time. He proved that for any Riemannian metric g on the 2-torus, $\operatorname{area}(\mathbb{T}^2, g) \geq \frac{\sqrt{3}}{2} \operatorname{sys}^2(\mathbb{T}^2, g)$, and that the constant $\frac{\sqrt{3}}{2}$ is optimal. Inspired by Loewner's method, Pu proved in [6] that for the real projective plane $\operatorname{area}(\mathbb{RP}^2, g) \geq \frac{2}{\pi} \operatorname{sys}^2(\mathbb{RP}^2, g)$ for any Riemannian metric g and that the constant $\frac{2}{\pi}$ is also optimal. For the case of Finsler metrics and the 2-torus, a complete summary of optimal isosystolic inequalities is done in [2]. This article gathers all known optimal constants, including the ones for Riemannian, reversible Finsler and not-necessarily reversible Finsler metrics for both Holmes–Thompson and Busemann–Hausdorff areas. There, $\mathbb{T}^2$ is identified with the quotient of the Euclidean plane $\mathbb{R}^2$ by the integer grid $\mathbb{Z}^2$. In that case, a metric on $\mathbb{T}^2$ is just a metric on $\mathbb{R}^2$ compatible with the quotient map, and non-contractible loops in $\mathbb{T}^2$ correspond to paths between points in $\mathbb{R}^2$ that differ by some $z \in \mathbb{Z}^2 \setminus \{(0,0)\}$. The strategy followed in the article is to reduce the general case to the case where the metric is flat, in the sense that the unit balls in $T_x\mathbb{T}^2$ are the same for all $x \in \mathbb{T}^2$. Then, the inequality is most of the times a consequence of previously known results in convex geometry. See [2] for all the details.

2.1 The real projective plane

Pu, in [6], followed an analogous procedure to what Loewner did with $\mathbb{T}^2$ but for $\mathbb{RP}^2$, so it might be interesting to explicit a parallelism between $\mathbb{RP}^2$ and $\mathbb{T}^2$. What is the universal covering map of $\mathbb{RP}^2$? How can non-contractible loops in $\mathbb{RP}^2$ be characterised? Is there an analogous notion of *flat* metric for $\mathbb{RP}^2$ that makes computations easier? To answer the first question, recall that $\mathbb{RP}^2$ can be defined as a quotient space identifying antipodal points on the 2-sphere $\mathbb{S}^2$, as is shown in Figure 1. The quotient map $\mathbb{S}^2 \to \mathbb{RP}^2 \cong \mathbb{S}^2/\{\pm \operatorname{Id}\}$ is the universal covering map over $\mathbb{RP}^2$ since $\mathbb{S}^2$ is simply connected, and plays an analogous role to the quotient map $\mathbb{R}^2 \to \mathbb{T}^2 \cong \mathbb{R}^2/\mathbb{Z}^2$.

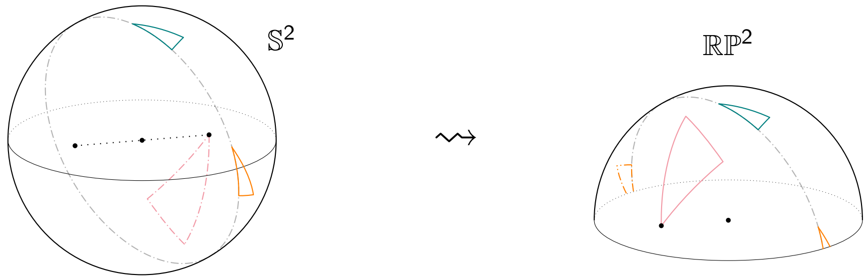


Figure 1: Universal covering map $\mathbb{S}^2 \to \mathbb{RP}^2$.

Alternatively, one could identify $\mathbb{RP}^2$ with a 2-disc $\mathbb{D}$ that has antipodal points on $\partial\mathbb{D}$ identified. When it comes to the characterisation of non-contractible loops, it can be shown that non-contractible loops in $\mathbb{RP}^2$ lift to paths in $\mathbb{S}^2$ joining antipodal points. See the illustration in Figure 2 for an intuitive idea and see, for instance, [5, Proposition 2.1] for a proof. More precisely, the condition of being non-contractible might be translated to the disc representation noting that a path in $\mathbb{S}^2$ from a point to its antipodal point must cross the horizon an odd number of times. As a subtlety, if the start and endpoints lie in the horizon, the open curve excluding these two points must cross the horizon an even number of times. Then, if starting and ending at points of the horizon counts as another cross, non-contractible loops in $\mathbb{RP}^2$ are characterised by crossing the horizon an odd number of times. Crossing the horizon is translated to jumping between opposite points of $\partial\mathbb{D}$, so non-contractible loops in $\mathbb{RP}^2$ are characterised by having an odd number of these jumps.

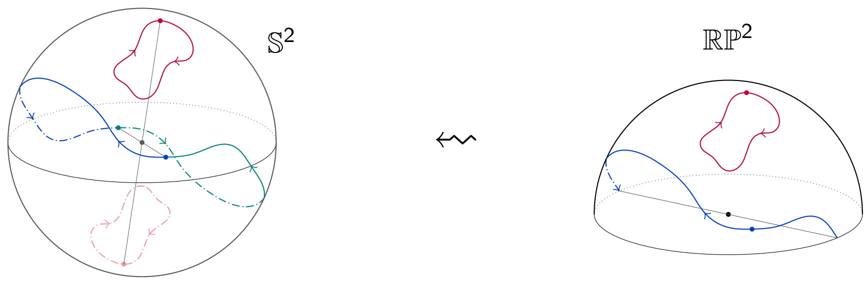


Figure 2: Correspondence between loops in $\mathbb{RP}^2$ and their lifts to $\mathbb{S}^2$.

Because translations are isometries of the Euclidean plane, a given convex body can be parallel transported from a point to another consistently to define a notion of flat Finsler metric on the 2-torus. Tangent vectors of $\mathbb{S}^2$ could also be parallel transported to another point. However, the transported vector will depend on how the parallel transport is performed. Thus, in order to get a well-defined notion of invariant metric on $\mathbb{S}^2$, one needs to assume the convex body to be rotationally invariant. In this special case, the metric is said to be a round metric on $\mathbb{RP}^2$, and can be alternatively defined as some multiple of the Riemannian metric obtained from the natural embedding of $\mathbb{S}^2$ in $\mathbb{R}^3$ as the unit Euclidean sphere. These metrics will play a similar role for $\mathbb{RP}^2$ compared to the role that flat Finsler metrics play on $\mathbb{T}^2$, although round metrics are much more restricted.

2.2 Previously known inequalities

As already mentioned, Pu proved in [6] that $\mathrm{area}(\mathbb{RP}^2, g) \geq \frac{2}{\pi}\,\mathrm{sys}^2(\mathbb{RP}^2, g)$ for any Riemannian metric and that equality holds if and only if g is isometric to a round metric on $\mathbb{RP}^2$. See [5, Section 4] for a proof that uses a more modern style, similarly to how the $\mathbb{T}^2$ case is treated in [2]. Note that in both cases equality holds for a flat or round metric, although for Pu's inequality all round metrics on $\mathbb{RP}^2$ are optimal while for Loewner's inequality only some flat metrics on $\mathbb{T}^2$ are optimal. In both cases, the procedure is to note that, by the uniformisation theorem, any metric is isometric to a conformal multiple of a flat or round one. Then, one observes that averaging the conformal factor gives a multiple of the flat or round metric, while it leaves the area invariant but increases the systole. Finally, the inequalities follow from the optimal flat or round metric cases. It can be computed that $\mathrm{area}(\mathbb{RP}^2, g) = \frac{2}{\pi}\,\mathrm{sys}^2(\mathbb{RP}^2, g)$ for any round metric on $\mathbb{RP}^2$ (see for instance [5, Section 4.2]). For the case of $\mathbb{T}^2$, before concluding, one must prove that the same isosystolic inequality holds also for any flat metric g. This is not as straightforward as for $\mathbb{RP}^2$, but it is equivalent to finding the Hermite constant γ_2, as is explained in [2].

Ivanov proved in [4] that $\mathrm{area}_{\mathrm{HT}}(\mathbb{RP}^2, F) \geq \frac{2}{\pi}\,\mathrm{sys}^2(\mathbb{RP}^2, F)$ also holds for reversible Finsler metrics. The idea of the proof is first to consider a non-contractible loop γ_0 on $\mathbb{RP}^2$ such that $\ell_F(\gamma_0) = \mathrm{sys}(\mathbb{RP}^2, F)$, which can be done by compactness arguments. Such loops are usually called *systolic loops*. As is shown in Figure 2, the union of the two lifts of γ_0 divides the 2-sphere in two 2-discs. Considering the pullback metric φ on one of the discs $\mathbb{D}$, the inequality is reduced to finding an inequality between $\mathrm{area}_{\mathrm{HT}}(\mathbb{RP}^2, F) = \mathrm{area}_{\mathrm{HT}}(\mathbb{D}, \varphi)$ and the length of $\partial\mathbb{D}$. Introducing *cyclic maps* $f = (f_1, \dots, f_n)$, Ivanov proves that

$$\mathrm{area}_{\mathrm{HT}}(\mathbb{D}, \varphi) \geq \frac{1}{2\pi} \int_{\partial\mathbb{D}} \sum_{i=1}^{n} f_i \cdot \mathrm{d}f_{i+1}. \tag{2}$$

Finally, Ivanov notes that for cyclically ordered and equidistant points $\{p_i\}_{i=1}^{n} \subseteq \partial\mathbb{D}$, the choice $f_i(x) = d_\varphi(p_i, x)$ leads to a *cyclic map*. See [4, Section 3] for the definition, properties and examples of *cyclic maps*. Under the assumption of a reversible metric, $\int_{\partial\mathbb{D}} f_i \cdot \mathrm{d}f_{i+1}$ is easy to compute using an arc-length parametrisation of $\partial\mathbb{D}$. In fact, it amounts to computing the signed area of the curve shown in Figure 3a. The signed area of each rectangle is $\frac{4\,\mathrm{sys}^2(\mathbb{RP}^2, F)}{n}\left(1 - \frac{2}{n}\right)$, which leads to

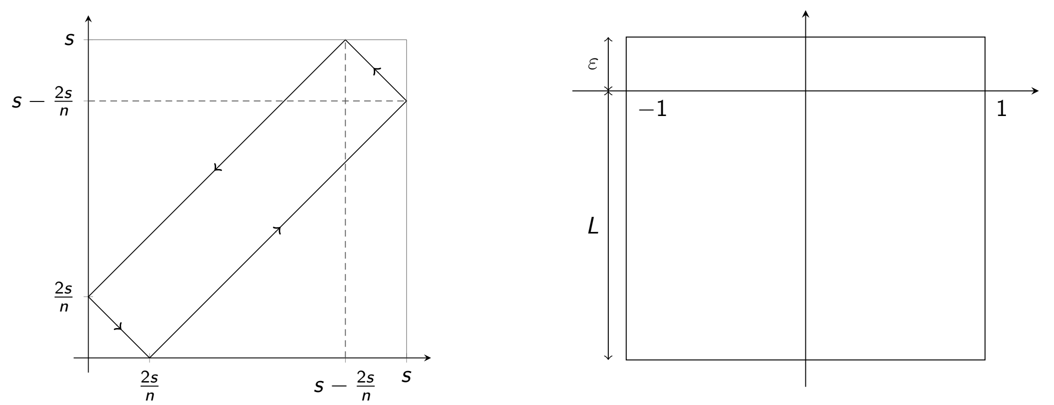
$$\mathrm{area}_{\mathrm{HT}}(\mathbb{RP}^2, F) = \mathrm{area}_{\mathrm{HT}}(\mathbb{D}, \varphi) \geq \frac{2}{\pi}\,\mathrm{sys}^2(\mathbb{RP}^2, F)\left(1 - \frac{2}{n}\right). \tag{3}$$

The proof is concluded noting that n can be chosen arbitrarily large. Ivanov's result and Proposition 1.2 imply that $\mathrm{area}_{\mathrm{BH}}(\mathbb{RP}^2, F) \geq \frac{2}{\pi}\,\mathrm{sys}^2(\mathbb{RP}^2, F)$ for any reversible Finsler metric. Note that the inequality is optimal in both cases, because equality holds for any round metric on $\mathbb{RP}^2$, which is Riemannian.

Round metrics do not seem to be relevant for Ivanov's result. Nevertheless, they play an important role in the case of $\mathbb{T}^2$. A *stable norm* on $T_x\mathbb{T}^2$, introduced in [3], is defined as $\|z\|_x = \lim_{k\to\infty} \frac{d(x, x+kz)}{k}$ for $z \in \mathbb{Z}^2$. This norm depends on the original Finsler metric on $\mathbb{T}^2 \cong \mathbb{R}^2/\mathbb{Z}^2$, and it can be shown to be independent of x. This means that the *stable metric* is flat, and it turns out that $\mathrm{area}_{\mathrm{HT}}(\mathbb{T}^2, F) \geq \mathrm{area}_{\mathrm{HT}}(\mathbb{T}^2, \|\cdot\|)$ and $\mathrm{sys}(\mathbb{T}^2, F) = \mathrm{sys}(\mathbb{T}^2, \|\cdot\|)$. Moreover, as is proven in [2], $\mathrm{area}_{\mathrm{BH}}(\mathbb{T}^2, F) \geq \mathrm{area}_{\mathrm{BH}}(\mathbb{T}^2, \|\cdot\|)$ also for reversible metrics. Thus, all these optimal isosystolic inequalities reduce to their respective flat cases. Following what is explained in [2], Minkowski's first theorem implies that $\mathrm{area}_{\mathrm{BH}}(\mathbb{T}^2, F) \geq \frac{\pi}{4}\,\mathrm{sys}^2(\mathbb{T}^2, F)$ for reversible and flat metrics, being optimal for the supremum norm. Due to a theorem by Mahler, the areas of a symmetric convex ball and its dual are related by $|B_x| \cdot |B_x^\circ| \geq 8$, being also optimal for the supremum norm. This implies that $\text{area}_{\text{HT}}(\mathbb{T}^2, F) \geq \frac{2}{\pi} \text{sys}^2(\mathbb{T}^2, F)$ is optimal for flat and reversible metrics. By the properties of the *stable norm* one deduces that the previous optimal inequalities for flat and reversible metrics are also valid for any reversible metrics. As a final comment, the optimal isosystolic inequalities for $\text{area}_{\text{HT}}(\mathbb{T}^2, F)$ and $\text{area}_{\text{BH}}(\mathbb{T}^2, F)$ are different for the reversible case, in contrast with the case of $\mathbb{RP}^2$. This is because, among Finsler metrics, optimal metrics F_0 for $\mathbb{T}^2$ are not Riemannian, and satisfy $\text{area}_{\text{BH}}(\mathbb{T}^2, F_0) > \text{area}_{\text{HT}}(\mathbb{T}^2, F_0)$ by Proposition 1.2, while optimal metrics for $\mathbb{RP}^2$ are the round ones, which are Riemannian.

3. Systolic freedom for Busemann–Hausdorff area

Minkowski's theorem prevents symmetric convex bodies $K \subseteq \mathbb{R}^2$ such that $\text{int}(K) \cap \mathbb{Z}^2 = \{(0,0)\}$ from having a Lebesgue measure $|K| > 4$, as is explained in [2, Section 3]. The condition $\text{int}(K) \cap \mathbb{Z}^2 = \{(0,0)\}$ ensures that the flat metric F with unit ball K fulfils $\text{sys}(\mathbb{T}^2, F) \geq 1$. This key fact implies the optimal inequality for the Busemann–Hausdorff area and reversible metrics. However, for non-symmetric convex bodies the theorem no longer applies. In fact, as is proven in [2, Section 3.2], there exists a family of flat metrics F_ε such that $\text{sys}(\mathbb{T}^2, F_\varepsilon) = 1$ and $|K_\varepsilon| = \frac{(1+\varepsilon)^2}{2\varepsilon}$ for the corresponding unit ball K_ε. By definition of the Busemann–Hausdorff area, letting $\varepsilon \to 0$ allows one to have $\text{area}_{\text{BH}}(\mathbb{T}^2, F_\varepsilon)$ arbitrarily small, proving systolic freedom.



(a) Here s denotes $\text{sys}(\mathbb{RP}^2, F)$.

(b) The upper direction points to the pole.

Figure 3: In the left, curves in $\mathbb{R}^2$ whose signed areas give the result of the individual integrals in (2). In the right, unit balls along the meridians of the hemisphere.

For the case of $\mathbb{RP}^2$, an analogous procedure would be to look for arbitrarily large unit balls that do not lead to an arbitrarily small value for the systole. This is proven to be possible in [5, Section 6], which leads to the conclusion that systolic freedom also holds in the non-reversible case for area_{BH}. The idea behind the construction in [5] is to build a metric in a hemisphere of $\mathbb{S}^2$ such that the equator contains a systolic loop of some fixed length. In order to have a small value for $\text{area}_{\text{BH}}(\mathbb{RP}^2, F)$, one needs to have large unit balls in great part of the hemisphere of $\mathbb{S}^2$. However, these large unit balls (which lead to short distances) must be such that a systolic loop still lies inside the equator. This is done with unit balls of arbitrarily large size L

in one direction and arbitrarily small size ε in the opposite direction, as is shown in Figure 3b. These balls are allowed to be arbitrarily large and they prevent curves that go towards the pole from being too short. Note that they are convex sets containing the origin, so they correspond to some non-reversible Finsler metric. The final step is to make such a metric on a hemisphere of $\mathbb{S}^2$ well-defined and compatible with a metric on $\mathbb{RP}^2$. First of all, one needs to have a well-defined unit ball at the pole: it cannot depend on the meridian that approaches the point. This can be achieved changing smoothly the unit ball in Figure 3b to a rotationally invariant one around the pole. Besides, a metric F on $\mathbb{S}^2$ is compatible with a metric on $\mathbb{RP}^2$ if $F_x(u_1) = F_{-x}(u_2)$, where u_1 and u_2 are the different lifts of some $v \in T_{[x]}\mathbb{RP}^2$. Geometrically, assume that one observes $\mathbb{S}^2$ from the point such that x and $-x$ are the closest and furthest points of the equator, respectively. As is illustrated in Figure 2, from this point of view, u_1 and u_2 are half turn rotations of one another. A change in point of view so that $-x$ is now in front and still with the pole above corresponds to a horizontal flip of the view of $T_{-x}\mathbb{S}^2$. In conclusion, the unit balls of antipodal points in the equator must be vertically flipped when seen the same way as in Figure 3b. Thus, it is enough to change smoothly the unit balls in the equator to vertically symmetric ones in order to have a compatible metric.

Note that both smoothing procedures can be done without changing the Lebesgue measure of the unit balls and that the angular integration region is $\left(0, \frac{\pi}{2}\right) \times (0, 2\pi)$, which has an area of π^2. Then, for this metric, one gets from (1) that $\mathrm{area}_{\mathrm{BH}}(\mathbb{RP}^2, F) = \frac{\pi}{2(\varepsilon+L)} \cdot \pi^2$, which can be made arbitrarily small for $L \to \infty$. When it comes to the systole, recall that a lift of a non-contractible loop γ must jump between opposite points of the equator an odd number of times. Considering only a part of γ if necessary, one can assume that γ joins opposite points of the equator without any other jump in between. Note that the unit balls of Figure 3b are a non-symmetric version of the supremum norm $\|(u_1, u_2)\| = \max\{|u_1|, |u_2|\}$. For this non-symmetric version it can be computed that $\|(u_1, u_2)\| = \max\left\{|u_1|, \frac{u_2}{\varepsilon}, -\frac{u_2}{L}\right\}$. See [5, Proposition 6.2] for the details. If $\gamma = (\gamma_1, \gamma_2)$ does not enter in the smoothen zone around the pole,

$$\ell_F(\gamma) = \int_0^1 F_{\gamma(t)}(\gamma_1'(t), \gamma_2'(t))\,\mathrm{d}t \geq \left|\int_0^1 \gamma_1'(t)\,\mathrm{d}t\right| = |\gamma_1(1) - \gamma_1(0)| \geq \pi.$$

Note that equality holds if $\gamma_2'(t) = 0$ and γ_1 increases or decreases monotonically between azimuthal coordinates that differ exactly in π. If γ enters the smoothen zone around the pole, the first part of γ must join the initial point with the zone. By what has been mentioned above, the length of vectors pointing to the pole is proportional to $\frac{1}{\varepsilon}$. Then, a small enough choice of ε would imply that $\ell_F(\gamma) > \pi$ also, and therefore $\mathrm{sys}(\mathbb{RP}^2, F) = \pi$. In the end, $\mathrm{area}_{\mathrm{BH}}(\mathbb{RP}^2, F) = \frac{\pi}{2(\varepsilon+L)}\,\mathrm{sys}^2(\mathbb{RP}^2, F) < \frac{\pi}{2L}\,\mathrm{sys}^2(\mathbb{RP}^2, F)$ for any value of $L > 0$. In particular, since L can be chosen arbitrarily large, there is systolic freedom for $\mathbb{RP}^2$ and the Busemann–Hausdorff area. See [5, Section 6] for more details.

4. Optimal inequalities for non-reversible metrics

Álvarez Paiva, Balacheff and Tzanev proved in [1, Theorem IV] that $\mathrm{area}_{\mathrm{HT}}(\mathbb{T}^2, F) \geq \frac{3}{2\pi}\,\mathrm{sys}^2(\mathbb{T}^2, F)$ for flat metrics and that equality holds when the unit ball is the triangle with vertices $(1, 0)$, $(0, 1)$ and $(-1, -1)$. Finally, by the properties of the *stable norm*, one deduces that

$$\mathrm{area}_{\mathrm{HT}}(\mathbb{T}^2, F) \geq \mathrm{area}_{\mathrm{HT}}(\mathbb{T}^2, \|\cdot\|) \geq \frac{3}{2\pi}\,\mathrm{sys}^2(\mathbb{T}^2, \|\cdot\|) = \frac{3}{2\pi}\,\mathrm{sys}^2(\mathbb{T}^2, F)$$

also for any Finsler metric.

Finding the optimal isosystolic inequality for the more general Finsler case for area_{HT} and $\mathbb{RP}^2$ is still an open problem. Existence of an optimal inequality can be proven by symmetrising the metric. Indeed, considering the symmetric metric $\tilde{F}_x(u) = F_x(u) + F_x(-u)$, it can be proven in dimension 2 that $|\tilde{B}_x^\circ| \leq 6|B_x^\circ|$ (see [7, Theorem 1]). If $\gamma \subseteq \mathbb{RP}^2$ is a systolic loop for $\tilde{F}$, the inverted loop $-\gamma$ is also non-contractible, and then $\text{sys}(\mathbb{RP}^2, \tilde{F}) = \ell_{\tilde{F}}(\gamma) = \ell_F(\gamma) + \ell_F(-\gamma) \geq 2\,\text{sys}(\mathbb{RP}^2, F)$. Joining these inequalities with the optimal inequality for reversible metrics,

$$\text{area}_{\text{HT}}(\mathbb{RP}^2, F) \geq \frac{1}{6}\,\text{area}_{\text{HT}}(\mathbb{RP}^2, \tilde{F}) \geq \frac{1}{6} \cdot \frac{2}{\pi}\,\text{sys}^2(\mathbb{RP}^2, \tilde{F}) \geq \frac{2}{3} \cdot \frac{2}{\pi}\,\text{sys}^2(\mathbb{RP}^2, F).$$

Note that this implies that the constant $\frac{2}{\pi}$ can be improved, at most, by a factor of $\frac{2}{3}$ for non-reversible metrics. However, [7, Theorem 2] states that $\text{area}_{\text{HT}}(\mathbb{RP}^2, F) = \frac{1}{6}\,\text{area}_{\text{HT}}(\mathbb{RP}^2, \tilde{F})$ if and only if almost all unit balls are triangles. The fact that the optimal metric for the reversible case is a round one, far from having symmetrised triangular unit balls, suggests that $\frac{4}{3\pi}$ is not optimal.

Conjecture 4.1. *The optimal isosystolic inequality for Finsler metrics and Holmes–Thompson area is* $\text{area}_{\text{HT}}(\mathbb{RP}^2, F) \geq \frac{2}{\pi}\,\text{sys}^2(\mathbb{RP}^2, F)$.

The author has tried to attack the non-reversible case and Holmes–Thompson area with little success. Consider the family of metrics used in the proof of systolic freedom in the previous section. For simplicity, consider the metric before the smoothing, which can be done in an irrelevant arbitrarily small region. Imposing that the systole is still attained along the equator amounts to imposing that $\frac{1}{\varepsilon} + \frac{1}{L} \geq 2$. Indeed, as before, if γ does not touch the pole, $\ell_F(\gamma) \geq \pi$. And if it touches it, it must go up and then back down, having a length $\ell_F(\gamma) \geq \frac{\pi}{2}\left(\frac{1}{\varepsilon} + \frac{1}{L}\right) \geq \pi$. The dual convex body of the unit balls of Figure 3b can be computed to be the convex hull of the points $(\pm 1, 0)$, $\left(0, \frac{1}{\varepsilon}\right)$ and $\left(0, -\frac{1}{L}\right)$. This convex kite has Lebesgue measure $\frac{1}{\varepsilon} + \frac{1}{L}$, and similarly to the Busemann–Hausdorff case, by (1),

$$\text{area}_{\text{HT}}(\mathbb{RP}^2, F) = \frac{\frac{1}{\varepsilon} + \frac{1}{L}}{\pi} \cdot \pi^2 = \frac{1}{\pi}\left(\frac{1}{\varepsilon} + \frac{1}{L}\right)\text{sys}^2(\mathbb{RP}^2, F).$$

In conclusion, $\text{area}_{\text{HT}}(\mathbb{RP}^2, F) \geq \frac{2}{\pi}\,\text{sys}^2(\mathbb{RP}^2, F)$ if $\frac{1}{\varepsilon} + \frac{1}{L} \geq 2$, which prevents the existence of shortcuts through the pole. The smoothening process would just lead to results arbitrarily close to the above inequality, agreeing with Conjecture 4.1.

Any unit ball can be drawn inside a rectangle and containing a triangle that touches three of the furthest points from the origin. This might leave shortest lengths invariant and it might be interesting to perform a similar test for triangle-shaped unit balls. For example, consider triangles with vertices $(1, 0)$, $(-\delta, \varepsilon)$ and $(-\delta, -L)$. In this case, the dual triangle has vertices $\left(-\frac{1}{\delta}, 0\right)$, $\left(1, \frac{1+\delta}{\varepsilon}\right)$ and $\left(1, -\frac{1+\delta}{L}\right)$, and Lebesgue measure $\frac{(1+\delta)^2}{2\delta}\left(\frac{1}{\varepsilon} + \frac{1}{L}\right)$. The norm is not so easy to compute but one could expect that imposing that the systole is attained around the equator would imply the same (or worse) inequality. It would be a surprise if there existed values for ε, L and δ that prove Conjecture 4.1 wrong. The author's search of examples that prove the conjecture wrong has been unfruitful and looking for ways to prove it might be more sensible.

A minor advance in this direction has been achieved in [5, Theorem 5.13], giving a slight generalisation of Ivanov's result for reversible metrics. It states that the inequality is also true for metrics such that the distance between any two points of a systolic loop γ_0 is attained through γ_0. In other words, one needs to have no shortcuts between points of γ_0 that deviate from γ_0. In this case, if γ_0 connects x to y (and not the other way around), the definition of systole ensures that there are no shortcuts from x to y. However, in the non-reversible case, there might be shortcuts from y to x. Ivanov's assumption is to have a reversible metric, which implies that there are no such shortcuts. The assumption in [5, Theorem 5.13] is weaker but still ensures that there are no such shortcuts. The proof is essentially the same that the one for Ivanov's theorem although Figure 3a gets slightly modified. For instance, the curve is no longer contained in the square $[0, s]^2$, and the short straight lines become unknown but bounded. The corresponding curve is shown in [5, Figure 6], and the inequality (3) is modified to

$$\operatorname{area}_{\mathrm{HT}}(\mathbb{RP}^2, F) = \operatorname{area}_{\mathrm{HT}}(\mathbb{D}, \varphi) > \frac{2}{\pi} \operatorname{sys}^2(\mathbb{RP}^2, F) \left(\frac{n-1}{n} - 2 \cdot \frac{n-1}{n^2} \right).$$

Luckily, for arbitrarily large n the inequality becomes $\operatorname{area}_{\mathrm{HT}}(\mathbb{RP}^2, F) \geq \frac{2}{\pi} \operatorname{sys}^2(\mathbb{RP}^2, F)$. A sufficient condition to avoid shortcuts is that the systolic curve γ_0 has the same forward and backward length. In particular, this holds if $F_{\gamma_0(t)}(\gamma_0'(t)) = F_{\gamma_0(t)}(-\gamma_0'(t))$ for all t. In other words, reversibility of the metric along a systolic curve is enough. Some ideas to attack the general case would be to try to modify the metric around a systolic curve to a case under which the theorem holds. This might be easier than to modify the metric at all points, although the attempts done by the author lead to inconclusive scenarios. For instance, making the unit balls symmetric along a systolic curve by enlarging them, $\operatorname{area}_{\mathrm{HT}}$ decreases but shortcuts might appear. Instead, if the balls are symmetrised by stretching them, the systole must increase, but so does the area. The only way the author has tried to define a kind of an overall averaged norm on $\mathbb{S}^2$ is considering

$$\tilde{F}_x(v) = \int_{\mathrm{SO}(3)} F_{\sigma(x)}((T_x\sigma)v)\, \mathrm{d}\mu(\sigma),$$

where μ is the unique left-invariant Haar measure on SO(3) such that $\mu(\mathrm{SO}(3)) = 1$. Intuitively, the unit norm has been averaged over all directions around a point and over all points, so that $\tilde{F}$ corresponds to a round metric on $\mathbb{RP}^2$. It can be proved that $\operatorname{sys}(\mathbb{RP}^2, \tilde{F}) \geq \operatorname{sys}(\mathbb{RP}^2, F)$, because any curve joining antipodal points under the action of $\sigma \in \mathrm{SO}(3)$ has the same property. However, $\operatorname{area}_{\mathrm{HT}}(\mathbb{RP}^2, F) \geq \operatorname{area}_{\mathrm{HT}}(\mathbb{RP}^2, \tilde{F})$ can be false in some cases. For instance, considering the unit balls in Figure 3b, the average norm for the tangent vector $(1, 0)$ in all directions should be

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} F_x(\cos t, \sin t)\, \mathrm{d}t = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \max\left\{ |\cos t|, \frac{\sin t}{\varepsilon}, -\frac{\sin t}{L} \right\} \mathrm{d}t = \frac{\sqrt{1+\varepsilon^2}}{\pi\varepsilon} + \frac{\sqrt{1+L^2}}{\pi L}.$$

Then, the unit sphere is given by all vectors lying on the Euclidean circle with radius $r = \frac{\pi}{\frac{\sqrt{1+\varepsilon^2}}{\varepsilon} + \frac{\sqrt{1+L^2}}{L}}$. For the case of $\varepsilon = L = 1$, recall that $|B_x^\circ| = \frac{1}{\varepsilon} + \frac{1}{L} = 2$, and for the averaged metric,

$$|\tilde{B}_x^\circ| = \frac{\pi}{r^2} = \frac{1}{\pi} \left(\frac{\sqrt{1+\varepsilon^2}}{\varepsilon} + \frac{\sqrt{1+L^2}}{L} \right)^2 = \frac{(\sqrt{2}+\sqrt{2})^2}{\pi} = \frac{8}{\pi} > |B_x^\circ|.$$

This shows that the averaging procedure fails to have good properties even for the supremum norm. As was suggested by F. Balacheff, another approach could be to consider a contact structure on the unitary tangent bundle $S^*\mathbb{RP}^2$. With contact forms there is a theorem similar to the uniformisation theorem that says that the initial contact form and a fixed round one are contactomorphic. One might be able to average over the group of diffeomorphisms of $S^*\mathbb{RP}^2$ that leaves the round contact form invariant. This is similar to the fact that the action of SO(3) leaves a round metric on $\mathbb{S}^2$ invariant. It turns out that $S^*\mathbb{RP}^2$ is isomorphic to the Lens space $L(4, 1)$. However, the systole seems to be more difficult to deal with.

A final idea to believe that Conjecture 4.1 is true is the following. Consider an attempt of minimising the Holmes–Thompson area only around a systolic loop with a fixed length. In order to decrease the value of area_{HT} one must increase the Lebesgue measure of the unit balls. However, this process could be intuitively done until the metric is symmetric along the systolic loop because otherwise the systole might decrease. In conclusion, it seems sensible that the metric that minimises area_{HT} is symmetric along a systolic loop, and the generalisation of Ivanov's theorem would apply in this case.

Acknowledgements

This work is based on an undergraduate thesis, which was partially supported by the Spanish Ministry "Ministerio de Educación y Formación Profesional" grant "Beca de Colaboración". Thanks to the SCM for giving me the chance of writing this work after being the thesis awarded the "Premi Emmy Noether 2024" by them. Finally, I want to thank my supervisor Florent Balacheff for his excellent guidance throughout the whole process and because of giving me the opportunity to discover what research is like in mathematics.

References

[1] J.C. Álvarez Paiva, F. Balacheff, K. Tzanev, Isosystolic inequalities for optical hypersurfaces, *Adv. Math.* **301** (2016), 934–972.

[2] F. Balacheff, T. Gil Moreno de Mora, Isosystolic inequalities on two-dimensional Finsler tori, *EMS Surv. Math. Sci.* (2024), Published online first.

[3] D. Burago, S. Ivanov, On asymptotic volume of Finsler tori, minimal surfaces in normed spaces, and symplectic filling volume, *Ann. of Math. (2)* **156(3)** (2002), 891–914.

[4] S.V. Ivanov, Filling minimality of Finslerian 2-discs, *Proc. Steklov Inst. Math.* **273(1)** (2011), 176–190.

[5] U. Lejarza Alonso, Optimal isosystolic inequalities on real projective planes, *Grad. J. Math.* **8(2)** (2023), 46–72.

[6] P.M. Pu, Some inequalities in certain nonorientable Riemannian manifolds, *Pacific J. Math.* **2(1)** (1952), 55–71.

[7] C.A. Rogers, G.C. Shephard, The difference body of a convex body, *Arch. Math. (Basel)* **8** (1957), 220–233.