S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 5, стр. 177–188 (2008) УДК 514.15, 512.7
MSC 20F65

GEODESICS IN THE HEISENBERG GROUP: AN ELEMENTARY APPROACH

G. A. NOSKOV

Abstract. We derive in an elementary way the shape of geodesics of the left invariant Carnot-Caratheodory-Finsler metrics on the Heisenberg group. The only existing proof of this result was given by V. N. Berestovskii, using the Pontryagin maximum principle.

Keywords: Heisenberg group, geodesic, Isoperimetric Problem.

Introduction

The fundamental result of [1] states that every left-invariant length metric d [1] on a (connected) Lie group G is determined by a so-called Carnot-Caratheodory-Finsler metric (CCF-metric) μ_F. The last is defined by a pair (V_e, F), where V_e is a vector subspace of the Lie algebra L of G , generating L as a Lie algebra, and F is a norm on V_e. The subspace V_e generates a left-invariant distribution $\mathcal{D}$ of tangent subspaces on G, namely for $g \in G$ we set $V_g = dl_g(V_e)$, where l_g is the left translation by g. The metric μ_F assigns to each $g \in G$ a norm on V_g, whose pullback to V_e along the map dl_g is F. The length of an absolutely continuous path tangent to $\mathcal{D}$ is defined via the integral of the $F-$norm of the velocity vector of the path, and the distance $d(p,q)$ between $p, q \in G$ is equal to the infimum of the lengths of absolutely continuous paths which are tangent to $\mathcal{D}$. The content of [1] is that every left-invariant length metric d on G is of this form for a suitable μ_F.

We define a geodesic in a space with interior metric as a locally isometric mapping of a Euclidean straight line into the space. There are extremely few groups on which geodesics (and a fortiori shortest routes) can be more or less explicitly determined. And the Heisenberg group is among these exclusive groups due to another V. N.

Noskov G.A., Geodesics in the Heisenberg group: an elementary approach.

Received October, 9, 2006, published April, 8, 2008.

[1] This means that any two points p, q can be joined by a rectifiable path of length arbitrarily close to $d(p,q)$.

Berestovskii result [2]. In that paper the Heisenberg group with a 2-dimensional distribution is considered and the geodesics are described, using the celebrated "Pontryagin Maximum Principle"[9]. The aim of the present paper is to give a more elementary proof of this result, relying on some slight generalization of the Isoperimetric Problem for the shortest Minkowski length of the curves encircling a given Euclidean area [3].

Acknowledgement The work was supported by the SFB 701 of University Bielefeld. The author would like to thank H. Abels and V. N. Berestovskii for support and fruitful discussions. The help of Frau R. Engel in correcting the language of the paper was indispensable.

1. Heisenberg group and its horizontal distribution

The Heisenberg group H^3 over the reals. The group H^3 is the set $\mathbb{R}^3$ with multiplication

$$(x,y,z)(x',y',z') = (x+x', y+y', z+z' + \frac{1}{2}(xy' - yx')).$$

We identify the tangent space T_gH^3 at $g \in H^3$ with $\mathbb{R}^3$ via the left translation in $\mathbb{R}^3$.

Let $\partial_x, \partial_y, \partial_z$ be the coordinate vector fields on H^3. For $g = (x_0, y_0, z_0)$ the differential dl_g, acts as follows:

$$\partial_x \mapsto \partial_x - \frac{1}{2}y_0\partial_z, \partial_y \mapsto \partial_y + \frac{1}{2}x_0\partial_z, \partial_z \mapsto \partial_z.$$

The form dz on T_0H^3 can be extended by left translations to a left invariant differential form θ on H^3. Precisely, at the point $g = (x_0, y_0, z_0)$ we have

$$\theta_g = dz \circ dl_g^{-1},$$

from which it follows that

$$\theta_g\left(\partial_{xg}\right) = \frac{1}{2}y_0, \theta_g\left(\partial_{yg}\right) = -\frac{1}{2}x_0, \theta_g\left(\partial_{zg}\right) = 1,$$

where the subscript g denotes evaluation at g. We conclude that

$$\theta = \frac{1}{2}(ydx - xdy) + dz.$$

Define the field $\mathcal{P} = \{P_g\}$ of tangent planes at the points of H^3 :

$$P_g = \{v \in T_gH^3 : \theta_g v = 0\}$$
$$= \{X\partial_{xg} + Y\partial_{yg} + Z\partial_{zg} : \frac{1}{2}(yX - xY) + Z = 0\} \subset \mathbb{R}^3.$$

The field $\mathcal{P}$ is also called a (plane) distribution or polarization on H^3. A differentiable curve $t \mapsto g(t) = (x(t), y(t), z(t)), t \in [a,b]$ is horizontal if $\dot{g}(t)$ lies in the plane $P_{g(t)}$ for any t or equivalently $g\left(t\right)^{-1}\dot{g}(t) \in P_0$. In other words the horizontal curves are precisely the curves which are tangent to $\mathcal{P}$. In fact $P_g = dl_g(P_0)$,where $l_g$is a left translation by g.

Absolute continuity. We have to work with a wider class of functions than just differentiable ones. A function $f : [a,b] \to \mathbb{R}$ is absolutely continuous if for any $\varepsilon > 0$, there is a $\delta > 0$ such that for any finite set of non-overlapping intervals (a_i, b_i), if $\sum_1^n |a_i - b_i| < \delta$ then $\sum_1^n |f(a_i) - f(b_i)| < \varepsilon$. An absolutely continuous function is continuous and of bounded variation. (A function $f : [a,b] \to \mathbb{R}$ is said to have a bounded variation if $\sup \sum_1^n |f(a_i) - f(a_{i-1})| < \infty$, were the supremum is taken over all finite sequences $a = a_0 < a_1 < ... < a_n = b$. This supremum is called the total variation of f over $[a, b]$.) A function f is of bounded variation iff f can be written as a difference two monotonically increasing functions [11]. In particular a function of bounded variation is differentiable almost everywhere. The derivative $D(F) = f(x) = F'(x)$ establishes a bijective correspondence:

$$D : \{absolutely\ continuous\ F : [a, b] \to \mathbb{R}, F(a) = 0\} \leftrightarrow \{integrable\ f : [a, b] \to \mathbb{R}\}.$$

The inverse is given by $I(f) = F(x) = \int_a^x f(t)dt$, op. cit.

We extend all these notions to vector functions in the obvious way and we extend the definition of the horizontal curve as follows. An absolutely continuous curve $g(t)$ on H^3 is said to be horizontal, if it is tangent to $\mathcal{P}$ almost everywhere.

Lemma 1. *An absolutely continuous curve $t \mapsto g(t) = (x(t), y(t), z(t)), t \in [a, b]$ is horizontal iff $\frac{1}{2}(y\dot{x} - x\dot{y}) + \dot{z} = 0$ almost everywhere.*

Proof. $g(t)$ is horizontal $\Leftrightarrow \theta(x, y, z) = \frac{1}{2}(ydx - xdy) + dz$ vanishes on $(\dot{x}(t), \dot{y}(t), \dot{z}(t))$ a. e. $\Leftrightarrow \frac{1}{2}(y\dot{x} - x\dot{y}) + \dot{z} = 0$ a. e.

□

The following lemma is a very special case of the well known theorem of Rashevskii-Chow, [10], [5].

Lemma 2. *(Connectivity) Every two points in $\mathbb{R}^3$ can be joined by a smooth horizontal path.*

Proof. By invariance it is enough to connect 0 to an arbitrary point (a, b, c). First we connect 0 to $(a, b, 0)$ by a path $h = (x, y) : [0, 1] \to \mathbb{R}^2$ with an area c, which means that $c = \frac{1}{2}\int_0^1(-y(s)\dot{x}(s) + x(s)\dot{y}(s))ds$, see Section 3. Then the path

$$\left(x(t), y(t), \frac{1}{2}\int_0^t (-y(s)\dot{x}(s) + x(s)\dot{y}(s))ds\right) \tag{1}$$

connects 0 to (a, b, c). It is horizontal since θ vanishes on the tangent vector

$$(\dot{x}(t), \dot{y}(t), \dot{z}(t)) = \left(\dot{x}(t), \dot{y}(t), \frac{1}{2}(-y(t)\dot{x}(t) + x(t)\dot{y}(t))\right).$$

□

We call the horizontal path constructed in this lemma as the lift of the plane path $h(t)$.

2. CCF-metrics on the Heisenberg group and their geodesics

A norm F on $\mathbb{R}^2$ gives rise to a metric on $\mathbb{R}^2$ and in this way we speak about the Minkowski geometry $(\mathbb{R}^2, F)$. To avoid confusion with the Euclidean length we will speak about the associated $F-$distance or the $F-$length. Of course the Minkowski geometry is completely determined by its unit disc $B_F = \{z \in \mathbb{R}^2 : F(z) \leq 1\}$.

Consider a norm F on the plane $P_0 \subset H^3$. The connectivity lemma allows to introduce the distance function on H^3 similarly as in Riemannian geometry. Namely, we push forward F onto each plane P_g via the differential of the left translation by g and define the $F-$length of an absolutely continuous path $g(t) = (x(t), y(t), z(t)), t \in [a, b]$ by integrating the velocity vector of $g(t)$. It is easy to see that this length coincides with the $F-$length of the (x, y)- projection:

$$l_F(g) = \int\limits_g F(dx, dy) = \int\limits_a^b F(\dot{x}(t), \dot{y}(t))dt.$$

Next we define a Carnot-Caratheodory-Finsler distance (CCF-distance) $d_F(g, h)$ for $g, h \in H^3$as the infimum of the lengths of piecewise smooth horizontal paths joining g to h. One can prove more: d_F is a complete length distance on H^3 and the topology defined by d_F is the Euclidean topology [1]. In this case, due to the results of S. E. Cohn-Vossen [6] any two points in (H^3, d_F) can be joined by a length minimizing path α. Moreover, when parameterized by arc length, α is an absolutely continuous horizontal path. We call a minimizing path a geodesic (although this does not agree with the usual terminology of Riemannian geometry where geodesics are locally minimizing paths).

It is proven in [2] that a path $g(t)$ in H^3, beginning at the unit $e \in H^3$, is geodesic if and only if it satisfies Pontryagin's Maximum Principle for a certain time-optimal control problem. Moreover, the projections $(x(t), y(t))$ onto the Minkowski plane $z = 0$ with norm F come in two kinds - they are either 1) geodesics of Minkowski geometry $(\mathbb{R}^2, F)$, or 2) parts of periodic trajectories, parameterized by arclength, over isoperimetric paths; and $z(t)$ equals the oriented area (on the Euclidean plane in Cartesian coordinates x and y) that is swept out by the moving vector $(x(t), y(t)), 0 < t < T$. If the unit sphere ∂F is strictly convex, then the geodesics of the first kind are one-parameter subgroups in H^3 with unit tangent vector $\dot{g}(t) \in L_0$. If ∂F is not strictly convex, then there are other geodesics (of first kind).

3. Geodesics and the isoperimetric problem

Area. The coning of a path $h(t), a \leq t \leq b$ in $\mathbb{R}^2$ is the closed path obtained by traversing first from 0 to $h(a)$ along a line segment, then traversing h and then returning to the origin along a line segment. By the (oriented) area of a path $h(t), a \leq t \leq b$ we mean the signed area of the coning of h. We denote it by $area(h)$.

Lemma 3. *For any horizontal path $g(t) = (x(t), y(t), z(t)), a \leq t \leq b$ in H^3 we have $z(b) - z(a) = area(h)$, where $h(t) = (x(t), y(t))$ is the horizontal projection of $g(t)$.*

Proof. Introduce the one-form $\omega = \frac{1}{2}(-ydx + xdy)$. It satisfies $d\omega = dx \wedge dy$ and its restriction to any line L through the origin vanishes, i.e. $\omega_L = 0$. According to Stokes' theorem, the area $area(c)$ enclosed by a closed planar path c is $\int_c \omega$. Let c be the coning of h. Because of the vanishing of $\omega_L$we have $\int_h \omega = \int_c \omega$. Finally, by Lemma 1

$$z(b) - z(a) = \int\limits_a^b \dot{z}dt = \frac{1}{2}\int\limits_h (-ydx + xdy) = \int\limits_c \omega = area(h).$$

□

Geodesics in H^3 and the isoperimetric problem in $\mathbb{R}^2$. By the left invariance it is enough do describe geodesics connecting 0 to some point $q \in H^3$.

Lemma 4. *A horizontal path $g : [0,T] \to H^3, g(t) = (x(t), y(t), z(t))$ is the shortest horizontal path from 0 to $g(T) = (a,b,c)$ if and only if its horizontal projection $h(t) = (x(t), y(t))$ has minimal F-length among all the paths in $\mathbb{R}^2$ joining 0 to (a,b) with area c.*

Proof. Let $h(t) = (x(t), y(t)), t \in [0,T]$ be a path of minimal $F-$length among the paths in $\mathbb{R}^2$ joining 0 to (a,b) and with area c. Consider the horizontal path

$$g(t) = \left(x(t), y(t), \frac{1}{2}\int_0^t (-y(s)\dot{x}(s) + x(s)\dot{y}(s))ds \right),$$

see (1). It joins 0 to (a,b,c) and its length is equal to

$$\int_g F(dx, dy) = \int_0^T F(\dot{x}, \dot{y})\, dt.$$

The path $g(t)$ is the shortest horizontal path, joining 0 to (a,b,c) because any other such path $g_1(t)$ should have a horizontal projection $h_1(t)$ of the same area as $g(t)$ has, by (3) and thus the $F-$length of $h_1(t)$ is at least as large as that of $h(t)$ by the minimality assumption. Conversely, suppose that $g(t)$ is the shortest horizontal path, joining 0 to (a,b,c) but there is a plane path $h_1(t)$ with area c which is strictly shorter than $h(t)$. Then the lift $g_1(t)$ of $h_1(t)$ joins 0 to (a,b,c) and its length is strictly smaller than that of $g(t)$, -a contradiction.

□

The lemma reduces the problem of geodesics to the following

Isoperimetric problem (IP). *Given point p in the Minkowski plane $(\mathbb{R}^2, F)$ and a number A find a path from 0 to p of a minimal F-length whose coning has the area A.*

"The isoperimetric problem has been a source of mathematical ideas and techniques since its formulation in classical antiquity, and it is still alive and well in its ability to both capture and nourish the mathematical imagination"[4].

4. Reformulation of the isoperimetric problem

In this section we divide IP into three cases and formulate the solution, see Theorem 6. Throughout this section by an F-plane we mean the complex plane $\mathbb{C} = \mathbb{R}^2$ endowed with a norm F. The unit disc is $B_F = \{z : F(z) \leq 1\}$.

Case $p = 0$. In this case the existence and uniqueness of the solution of the Isoperimetric Problem was given by Busemann [3]. It can be described as follows.

Consider the dual disc $B_F^\circ = \{z' : z \cdot z' \leq 1, z \in B_F\}$ (dot denotes the standard scalar product on $\mathbb{C} = \mathbb{R}^2$). Its boundary path ∂B_F°, rotated by $\pi/2$ (that is the path $I_F = i\partial B_F^\circ$) is called an isoperimetrix. It is convenient to call an isoperimetrix also any path obtained from I_F by dilation and translation and to call by an isoperimetric path any subpath of an isoperimetrix. (Recall that a dilation centered at the point p is the transformation, fixing p and stretching the distances by some constant factor.)

Busemann has proved that an isoperimetrix $I = I_F$, oriented counter-clockwise is the unique shortest closed path encircling the area $A = area\,(B_F^\circ)$. Hence (by dilation and translation) the isoperimetrix gives the solution of the IP in the case when A is arbitrary and $p = 0$.

Case $p \neq 0, B_F$ is strictly convex. The following "closing"trick works. The case $A = 0$ is clear, so we may assume $A > 0$. It is well known that B_F is strictly convex if and only if its dual path I° is of class C^1, see f.e. [7]. Start with any isoperimetrix I passing through $0, p$ and oriented counter-clockwise. The line segment $[0, p]$ divides I into two isoperimetric subpaths I_1, I_2. Let I_1 be the one for which 0 is the starting point and p is the end point. If it happens that $area\,(I_1) = A$, then the isoperimetric path I_1 is the solution of IP for the data (p, A). Indeed, if not and I' is a shorter solution, then the concatenation closed path $I'I_2$ is shorter than $I_1I_2 = I$ and has the same area $A + area\,(I_2) = area(I)$. This contradicts Busemann's result [3], which asserts that I has minimal $F-$length among the closed paths with the area $area(I)$.

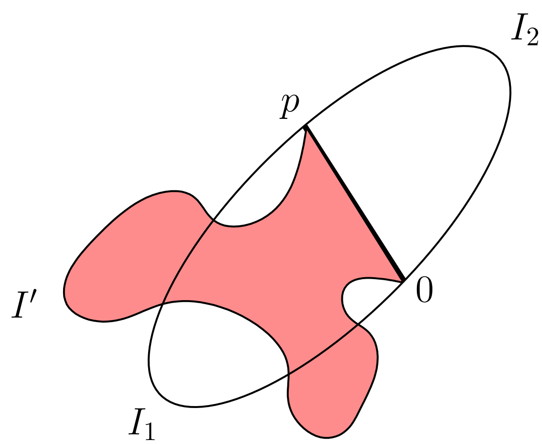

FIGURE 1. Closing trick.

Suppose now, that A is different from $area\,(I_1)$. Let us deform I continuously dilating and translating, but keeping the points $0, p$ on I. Since I is of class C^1 the area varies continuously and takes all possible positive values hence, under suitable deformation, $area\,(I_1)$ can be made equal to A and we get the solution by the previous paragraph.

Case $p \neq 0, B_F$ is not strictly convex. In this case the isoperimetrix is not of class C^1, that is it has "corner"points. This implies that not every number A can be realized as the area of an isoperimetric path. Indeed, the isoperimetric path, say P, is obtained by expanding isoperimetrix I, but if there is a corner point on I it remains a corner with the same angle under any dilation, see Figure 3. Thus the area of an isoperimetric path can not be made arbitrarily small. A concrete example of this phenomena is given by a metric $F = l_1$ on $\mathbb{R}^2$. Its dual is the l_∞-metric, so the isoperimetrix is a square with sides parallel to the axes x, y. Take, say $p = (1, 1)$ then there is only one isoperimetric path from 0 to p, it traverses first the $x-$axis and then the y-axis. In particular, the area of such a path is equal $1/2$ and can not be made less than that by any dilation and translation of I.

Thus, for non strictly convex B_F and general (p, A) there is no solution of the IP, which is a subpath of the isoperimetrix. We will show below in this section that if this is the case then the solution of IP is even simpler: the isoperimetric path is geodesic in the $F-$metric! The shape of such a geodesic can be easily retrieved from the Minkowski unit circle ∂B_F.

The tangent cone $T_z C$ of the convex set C at the point z is defined by

$$T_z C = closure\{w : \exists \lambda > 0; z + \lambda w \in C\}.$$

(Sometimes, if $z \in \partial C$, we write $T_z(\partial C)$ instead of $T_z C$). If $z \in \partial C$ then $T_z(\partial C)$ is either a halfplane (in case of smooth point z) or an acute convex cone which can be written in a form $T_z C = \mathbb{R}_+ a + \mathbb{R}_+ b$ for some non collinear a, b. We choose the basis (a, b), to be the unit vectors (in the Euclidean sense) and oriented counter-clockwise. We call $\mathbb{R}_+ a, \mathbb{R}_+ b$ the extreme rays of the tangent cone. We say that a segment σ is tangential to a compact convex set C if the supporting line l for C parallel to σ contains a point z in ∂C at which ∂C is differentiable. In particular in this case l is tangent to ∂C.

The triangle Δ_p. Let I be an isoperimetrix and suppose that the line segment $[0, p], p \in \mathbb{R}^2$ is a chord of I. The chord divides I into two subpaths I_1, I_2 and we may assume that I_1 is the one such that the concatenation of I_1 and $[0, p]^{-1}$ is continuous, closed and positively oriented curve. If $[0, p]$ is tangential to I, then we define $\Delta_p = [0, p]$. If $[0, p]$ is not tangential to I then the supporting line l for I_1 parallel to $[0, p]$ intersects I_1 in a unique point z and I is not differentiable at z. The point z divides I_1 into two subpaths I_0, I_p so that I_0 joins 0 to z and I_p joins z to p. The tangent cone $T_z I$ is acute and we can write $T_z I = \mathbb{R}_+ e_0 + \mathbb{R}_+ e_p$ where the extreme ray $\mathbb{R}_+ e_0$ is tangent to the path I_0^{-1} at z and the extreme ray $\mathbb{R}_+ e_p$ is tangent to the path I_p at z. Consider the translated cone $v + T_z I$ such that the halfline $v + \mathbb{R}_+ e_0$ contains 0 and the halfline $v + \mathbb{R}_+ e_p$ contains p. Define Δ_p to be the triangle with the vertices $v, 0, p$. Also we define the triangle Δ'_p similar to Δ_p with the vertex z and the opposite side containing $[0, p]$.

The function $\mu(p)$. For $p \in \mathbb{R}^2$ let $\mu(p) = \mu_F(p)$ be the infimum of the positive areas swept out by the subpaths of isoperimetrices joining 0 to p. The function $\mu(p)$ is continuous in p and $\mu(\lambda p) = \lambda^2 \mu(p)$ for any real λ.

Lemma 5. $\mu(p) = area(\Delta_p)$.

Proof. We follow the notation from the definition of Δ_p. The convex hull $ch(I_1)$ of I_1 contains Δ_p and is contained in Δ'_p. It follows that $area(\Delta_p) \leq area(I_1) \leq area(\Delta'_p)$. It is sufficient to prove that $area(\Delta'_p)$ tends to $area(\Delta_p)$ when the diameter of I tends to infinity. This in turn, by similarity of the triangles Δ'_p and Δ_p, is equivalent to say that the length of the side of Δ'_p, containing $[0, p]$, tends to the (Euclidean) length $|p|$ of the side $[0, p]$ of Δ_p. The above length is $length(l_p \cap (z + T_z I))$, where l_p is the line that contains $[0, p]$. Let c_ε be the chord of I of length $\varepsilon > 0$, close to z and parallel to $[0, p]$ and let l_ε be the line through this chord. Since I has one sided derivatives at z it follows that

$$length(l_\varepsilon \cap (z + T_z I)) = length(c_\varepsilon) + o(\varepsilon). \tag{2}$$

Now dilate I by $|p|/\varepsilon$ with center z. Then the chord c_ε will be dilated to the chord c_p of the isoperimetrix $\frac{|p|}{\varepsilon} I$, congruent to $[0, p]$, and the segment $l_\varepsilon \cap (z + T_z I)$ will

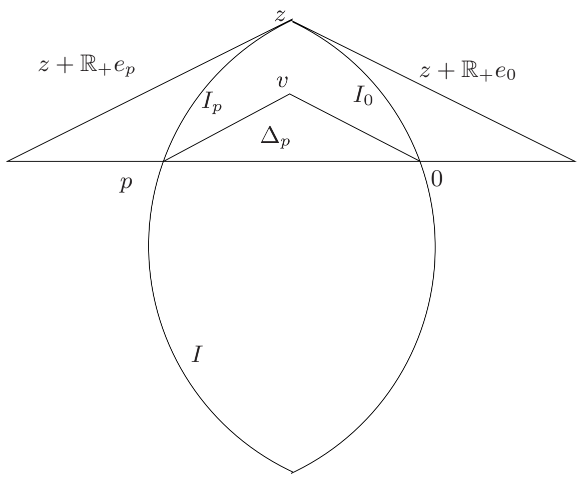

FIGURE 2. The triangle Δ_p.

be transformed to the segment $l'_p \cap (z + T_z I)$, where l'_p is the line through c_p. It follows from (2) that

$$length\left(l'_p \cap (z + T_z I)\right) - |p| = o\left(\varepsilon\right) \frac{|p|}{\varepsilon}$$

which tends to zero with $\varepsilon \to 0$. The result for the chord $[0, p]$ follows from this equality since the chord is congruent to c_p.

□

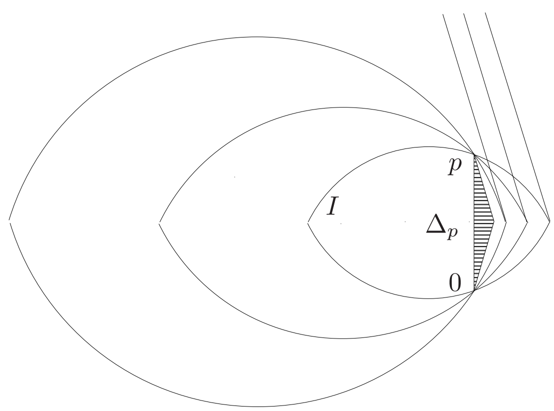

FIGURE 3. The triangle Δ_p as a limit of dilated curved triangles formed by $[0, p]$ and subpaths of isoperimetices.

We can now formulate the main technical result of the paper:

Theorem 6. *For any Minkowski plane* $\left(\mathbb{R}^2, F\right)$ *the solution of the Isoperimetric Problem is given in terms of the function* μ *as follows: 1)If* $A > \mu(p)$ *then the solution exists, is unique and is a subpath of an isoperimetrix, 2) If* $A \leq \mu(p)$ *then* *there exists the solution path, geodesic relative to the metric F and moreover such a path can be chosen as a concatenation of two line segments.*

In the case 2) of the Theorem 6 the solution is not unique in general.

5. Proof of Theorem 6

Case $\boldsymbol{A > \mu(\mathbf{p})}$. The proof is easy and makes use of the "closing"trick as for the strictly convex case. Indeed, there is a piece α of the isoperimetrix I connecting 0 to p with an area A. Let β be the remaining piece so that the concatenation $\alpha\beta$ constitute I, traversed in the counter clockwise direction. Suppose α is not a solution, then there is a path α' shorter than α and with the same area. Then $\alpha'\beta$ has the same area as I but shorter contradicting the Busemann solution I for closed paths. The same argument works for the uniqueness statement.

Case $\boldsymbol{A \leq \mu(\mathbf{p})}$. This is the main case and it will occupy the rest of the paper. We aim to show that there is a geodesic from 0 to p with an area A, which thus have to be the solution of IP. Clearly, any other solution then is the geodesic, too. Moreover we will show that this geodesic can be chosen as a concatenation of two line segments.

Digression about geodesics in Minkowski geometry. Let $(\mathbb{R}^2, F)$ be the Minkowski geometry with a unit disc $B = B_F$. We define a continuous path $\gamma : [t_0, t_1] \to \mathbb{R}^2$ to be an $F-$geodesic (or $B-$geodesic) if it is an isometric embedding . Call a closed convex cone in $\mathbb{R}^2$ (with apex 0) a geodesic cone if its intersection with the unit ball B_F is a triangle. The terminology is explained by the fact that an absolutely continuous path $\alpha(t)$ is geodesic iff there exists a geodesic cone C such that the velocity vector $\alpha'(t)$ belongs to C for almost every t. More generally, a path α from 0 to p is geodesic iff each directed chord $[a, b]$ of α is in a direction contained in the unique face of the unit ball containing $b - a$ in its relative interior, [8], Section 3, Prop. 3.

Duality theory. Suppose now that D is a compact convex set in $\mathbb{R}^2$ that contains the origin as an interior point. The support function of D is

$$s_D(x) = \sup\{xy |\ y \in D\}.$$

The radial function $\rho_D(x), x \neq 0$ is defined to be the positive number such that $\rho_D(x)x \in \partial D$. The dual D° of D is defined by

$$D^\circ := \{x \in \mathbb{R}^n : xy \leq 1 \text{ for all } y \in D\} = \{x \in \mathbb{R}^n : s_D(x) \leq 1\}.$$

The operation $D \mapsto D^\circ$ is an involution on the set of convex bodies, containing the origin in their interiors. We need the following important fact: the support and radial functions of D and D° respectively are multiplicatively inverse to each other, [12], Thm. 2. 2. 13.

Duality between flat segments and nonsmooth points. Recall the definition of the dual cone C° of the cone C in $\mathbb{R}^n$

$$C^\circ = \{v : v \cdot c \leq 0 \quad \forall c \in C\}.$$

Lemma 7. *Let D be the unit ball of the norm F on $\mathbb{R}^2$ and let D° be its dual. Let $T = T_z(D^\circ)$ be the tangent cone to D° at a nonsmooth point $z \in \partial D^\circ$. Then the cone T° is geodesic for the metric F.*

Proof. T can be uniquely written in the form $T = \mathbb{R}_+a+\mathbb{R}_+b$ such that a, b are unit (Euclidean) vectors, $a \neq -b$ and the basis $\{a, b\}$ is positively oriented. It is easy to see that $T^\circ = \mathbb{R}_+(-ia) + \mathbb{R}_+ib$. We shall prove that T° is $F-$geodesic, that is $T^\circ \cap D$ is a triangle. There are unique positive α, β, such that $-\alpha ia, \beta ib \in \partial D$. The radial function ρ_D equals 1 at $-\alpha ia, \beta ib$, hence the support function s_{D° equals 1 at these points too. We conclude that $-\alpha ia \cdot z = \beta ib \cdot z = 1$ (the lines $z+\mathbb{R}a, z+\mathbb{R}b$ are supporting for D°). It immediately follows that for any convex combination $u \in [-\alpha ia, \beta ib]$ we have $u \cdot z = 1$, that is $u \in \partial D$. Thus the segment $[-\alpha ia, \beta ib]$ entirely lies in ∂D, i. e. $T^\circ \cap D$ is the triangle with the vertices $0, -\alpha ia, \beta ib$.

□

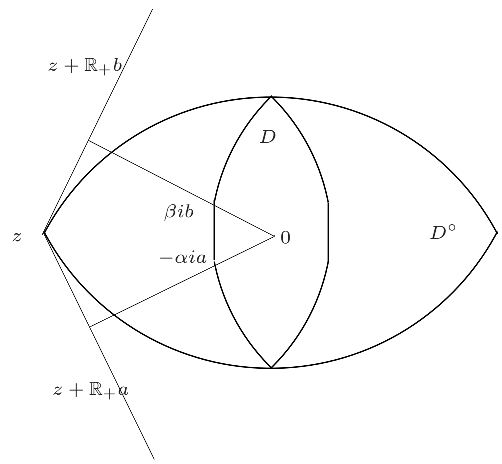

FIGURE 4. The point $z \in \partial D^\circ$ is dual to the segment $[-\alpha ia, \beta ib] \subseteq \partial D$.

Now we return to the unit disc $B = B_F$ from Theorem 6.

Lemma 8. *Let $T = T_z(iB^\circ)$ be the tangent cone to iB° at the point $z \in i\partial B^\circ$. Then the (two-side infinite) path α obtained by traversing the boundary $\partial T_z(iB^\circ)$ (with unit speed and without backtracking) is $F-$geodesic.*

Proof. Clearly, it is enough to prove that any finite length subpath of α is $F-$geodesic. The case of a smooth point z is clear so we may assume z to be singular. Write T in the form $T = \mathbb{R}_+a+\mathbb{R}_+b$ such that a, b are unit (Euclidean) vectors, $a \neq -b$ and the basis $\{a, b\}$ is positively oriented. It is easy to see that $T^\circ = \mathbb{R}_+(-ia) + \mathbb{R}_+ib$, i.e., the extreme rays of T° are obtained by rotating a, b by $-\pi/2, \pi/2$ respectively. By lemma 7 the cone T° is $iB-$geodesic. The rotation by $\pi/2$ shows that the cone $iT^\circ = \mathbb{R}_+a + \mathbb{R}_+(-b)$ is $B-$geodesic. This implies in particular that the path, starting from origin and traversing first distance T linearly in $-b$-direction , then traversing distance T linearly in a-direction is $F-$geodesic. But this path is congruent to a subpath P_T of α. Since the union $\cup_{T>0}P_T$ coincides with α, hence the last path is also $F-$geodesic.

□

Finalizing the proof of the theorem. Suppose $A \leq \mu(p)$. The triangle Δ_p has $[0, p]$ as a side and two other sides being parallel to the extreme rays of the cone $T_z(iB^\circ)$. By the previous lemma the path obtained by traversing the boundary

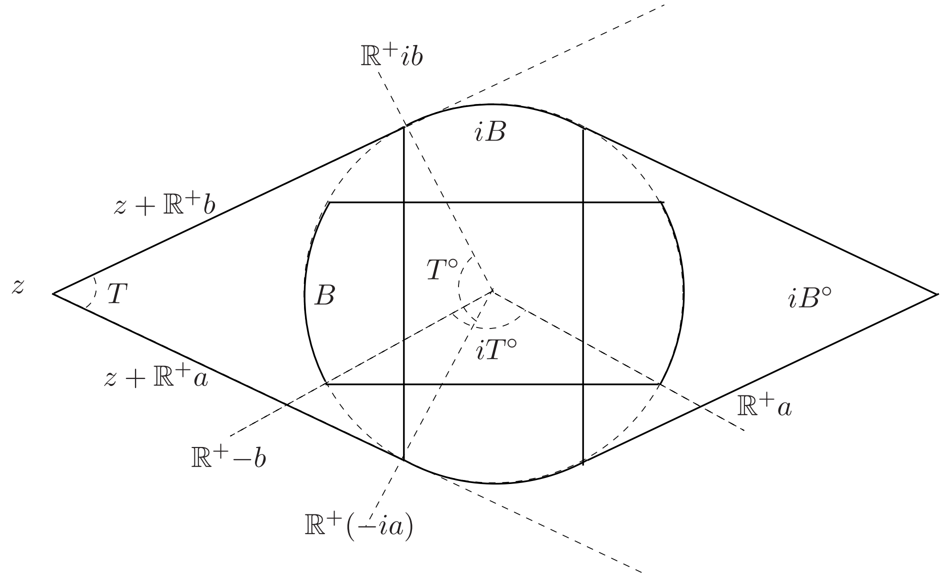

FIGURE 5. The boundary of the cone $T = T_z B^\circ$ is F-geodesic.

$\partial T_z\,(iB^\circ)$ without backtracking is $F-$geodesic. In particular, the concatenation α of the sides of Δ_p different from $[0,p]$ is an $F-$geodesic. This path solves the Isoperimetric Problem for the data $p, A = \mu\,(p)$. If $A < \mu\,(p)$, we change from Δ_p to a suitable subtriangle with area A and whose sides constitute an F-geodesic.

REFERENCES

[1] V. N. Berestovskiĭ, *Homogeneous manifolds with an intrinsic metric. I*, Sibirsk. Mat. Zh., **29**:6 (1988), 17–29. translation in Siberian Math. J. 29 (1988), no. 6, 887–897 (1989)

[2] V. N. Berestovskii, *Geodesics of nonholonomic left-invariant intrinsic metrics on the Heisenberg group and isoperimetric curves on the Minkowski plane*, *Siberian Math. J.*, **35**:1 (1994), 3–11, 1994. translation in Siberian Math. J. 35 (1994), no. 1, 1–8

[3] Herbert Busemann, *The isoperimetric problem in the Minkowski plane*, Amer. J. Math., **69** (1947), 863–871.

[4] Isaac Chavel, *Isoperimetric inequalities*, volume 145 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2001. Differential geometric and analytic perspectives.

[5] Wei-Liang Chow, *Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann., **117** (1939), 98–105.

[6] S. È. Con-Vossen, *Nekotorye voprosy differentsialnoi geometrii v tselom*, Edited by N. V. Efimov. Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1959.

[7] Joseph Diestel, *Geometry of Banach spaces—selected topics*, Lecture Notes in Mathematics, **485**, Springer-Verlag, Berlin, 1975.

[8] H. Martini, K. J. Swanepoel, G. Weiss, *The geometry of Minkowski spaces - a survey. Part I*, Expo. Math. 19 (2001), 97–142. (Errata: Expo. Math. 19 (2001), p. 364.)

[9] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *Matematicheskaya teoriya optimalnykh protsessov*, "Nauka", Moscow, fourth edition, 1983.

[10] P.K. Rashevskii, *About connecting two points of complete nonholonomic space by admissible curve* Uch. Zapiski ped. inst. Libknexta, **3** (1938), 83–94.

[11] Walter Rudin, *Real and complex analysis*, McGraw-Hill Book Co., New York, third edition, 1987.

[12] A. C. Thompson, *Minkowski geometry*, volume 63 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1996.

Gennady A. Noskov
Sobolev Institute of Mathematics,
Pevtsova st., 13,
644099, Omsk, Russia
Fakultät für Mathematik,
Universtät Bielefeld,
33501, Bielefeld, Germany
E-mail address: g.noskov@googlemail.com